Zurich Lectures in Advanced Mathematics

Edited by

Mathematics in Zurich has a long and distinguished tradition, in which the writing of lecture notes volumes and research monographs play a prominent part. The Zurich Lectures in Advanced Mathematics series aims to make some of these publications better known to a wider audience. The series has three main constituents: lecture notes on advanced topics given by internationally renowned experts, graduate text books designed for the joint graduate program in Mathematics of the ETH and the University of Zurich, as well as contributions from researchers in residence at the mathematics research institute, FIM-ETH. Moderately priced, concise and lively in style, the volumes of this series will appeal to researchers and students alike, who seek an informed introduction to important areas of current research.

Previously published in this series:

Yakov B. Pesin, *Lectures on partial hyperbolicity and stable ergodicity*
Sun-Yung Alice Chang, *Non-linear Elliptic Equations in Conformal Geometry*
Sergei B. Kuksin, *Randomly forced nonlinear PDEs and statistical hydrodynamics in 2 space dimensions*
Pavel Etingof, *Calogero-Moser systems and representation theory*

Published with the support of the Huber-Kudlich-Stiftung, Zürich

Guus Balkema
Paul Embrechts

High Risk Scenarios and Extremes

A geometric approach

European Mathematical Society

A. A. Balkema
Department of Mathematics
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
Netherlands
guus@science.uva.nl

P. Embrechts
Department of Mathematics
ETH Zurich
8092 Zurich
Switzerland
embrechts@math.ethz.ch

The cover shows part of the edge and of the convex hull of a realization of the Gauss-exponential point process. This point process may be used to model extremes in, for instance, a bivariate Gaussian or hyperbolic distribution. The underlying theory is treated in Chapter III.

2000 Mathematics Subject Classification 60G70, 60F99, 91B30, 91B70, 62G32, 60G55

ISBN 978-3-03719-035-7

The Swiss National Library lists this publication in The Swiss Book, the Swiss national bibliography, and the detailed bibliographic data are available on the Internet at http://www.helveticat.ch.

Contact address:

European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum FLI C4
CH-8092 Zürich
Switzerland

Phone: +41 (0)44 632 34 36
Email: info@ems-ph.org
Homepage: www.ems-ph.org

Typeset using the authors' TEX files: I. Zimmermann, Freiburg
Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Printed in Germany
9 8 7 6 5 4 3 2 1

Annemarie and my daughters have looked on my labour with a mixture of indulgence and respect. Thank you for your patience.

Guus

For Gerda, Krispijn, Eline and Frederik. Thank you ever so much for the wonderful love and support over the many years.

Paul

Foreword

These lecture notes describe a way of looking at extremes in a multivariate setting. We shall introduce a continuous one-parameter family of multivariate generalized Pareto distributions that describe the asymptotic behaviour of exceedances over linear thresholds. The one-dimensional theory has proved to be important in insurance, finance and risk management. It has also been applied in quality control and meteorology. The multivariate limit theory presented here is developed with similar applications in mind. Apart from looking at the asymptotics of the conditional distributions given the exceedance over a linear threshold – the so-called high risk scenarios – one may look at the behaviour of the sample cloud in the given direction. The theory then presents a geometric description of the multivariate extremes in terms of limiting Poisson point processes.

Our terminology distinguishes between extreme value theory and the limit theory for coordinatewise maxima. Not all extreme values are coordinatewise extremes! In the univariate theory there is a simple relation between the asymptotics of extremes and of exceedances. One of the aims of this book is to elucidate the relation between maxima and exceedances in the multivariate setting. Both exceedances over linear and elliptic thresholds will be treated. A complete classification of the limit laws is given, and in certain instances a full description of the domains of attraction. Our approach will be geometrical. Symmetry will play an important role.

The charm of the limit theory for coordinatewise maxima is its close relationship with multivariate distribution functions. The univariate marginals allow a quick check to see whether a multivariate limit is feasible and what its marginals will look like. Linear and even non-linear monotone transformations of the coordinates are easily accommodated in the theory. Multivariate distribution functions provide a simple characterization of the max-stable limit distributions and of their domains of attraction. Weak convergence to the max-stable distribution function has almost magical consequences. In the case of greatest practical interest, positive vectors with heavy tailed marginal distribution functions, it entails convergence of the normalized sample clouds and their convex hulls.

Distribution functions are absent in our approach. They are so closely linked to coordinatewise maxima that they do not accommodate any other interpretation of extremes. Moreover, distribution functions obscure an issue which is of paramount importance in the analysis of samples, the convergence of the normalized sample cloud to a limiting Poisson point process. Probability measures and their densities on $\mathbb{R}^d$ provide an alternative approach which is fruitful both in developing the theory and in handling applications. The theory presented here may be regarded as a useful complement to the multivariate theory of coordinatewise maxima.

These notes contain the text of the handouts, substantially revised, for a Nachdiplom course on point processes and extremes given at the ETH Zurich in the spring semester of 2005, with the twenty sections of the book roughly corresponding to weekly two-hour lectures.

Acknowledgements. Thanks to Matthias Degen, Andrea Höing and Silja Kinnebrock for taking care of the figures, to Marcel Visser for the figures on the AEX, and to Hicham Zmarrou for the figures on the DAX. We thank Johanna Nešlehová for her assistance with technical problems. We also thank her for her close reading of the extremal sections of the manuscript and her valuable comments. A special word of thanks to Nick Bingham for his encouraging words, his extensive commentary on an earlier version of the text, and his advice on matters of style and punctuation. The following persons helped in the important final stages of proofreading: Daniel Alai, Matthias Degen, Dominik Lambrigger, Natalia Lysenko, Parthanil Roy and Johanna Ziegel. Dietmar Salamon helped us to understand why discontinuities in the normalization are unavoidable in certain dimensions. We would also like to thank Erwin Bolthausen and Thomas Kappeler, who as editors of the series gave us useful input early on in the project. Thomas Hintermann, Manfred Karbe and Irene Zimmermann did an excellent job transforming the MS into a book. Guus Balkema would like to thank the Forschungsinstitut für Mathematik (FIM) of the ETH Zurich for financial support, and the Department of Mathematics of the ETH Zurich for its hospitality. He would also like to express his gratitude to the Korteweg–de Vries Instituut of the University of Amsterdam for the pleasant working conditions and the liberal use of their facilities.

Contents

Starred sections may be skipped on a first reading

Introduction

Browsing quickly through the almost 400 pages that follow, it will become immediately clear that this book seems to have been written by mathematicians for mathematicians. And yet, the title has the catchy "High Risk Scenarios" in it. Is this once again a cheap way of introducing finance related words in a book title so as to sell more copies? The obvious answer from our, the authors' point of view, must be no. This rather long introduction will present our case of defense: though the book is indeed written by mathematicians for a mathematically inclined readership, at the same time it grew out of a deeper concern that *quantitative risk management* (QRM) is facing problems where new mathematical theory is increasingly called for. It will be difficult to force the final product you are holding in your hands into some specific corner or school. From a mathematical point of view, techniques and results from such diverse fields as stochastics (probability and statistics), analysis, geometry and algebra appear side by side with concepts from modern mathematical finance and insurance, especially through the language of portfolio theory. At the same time, risk is such a broad concept that it is very much hoped that our work will eventually have applications well beyond the financial industry to areas such as reliability engineering, biostatistics, environmental modelling, to name just a few.

The key ingredients in most of the theory we present relate to the concepts of risk, extremes, loss modelling and scenarios. These concepts are to be situated within a complex random environment where we typically interpret complexity as high-dimensional. The theory we present is essentially a one-period theory, as so often encountered in QRM. Dynamic models, where time as a parameter is explicitly present, are not really to be found in the pages that follow. This does not mean that such a link cannot be made; we put ourselves however in the situation where a risk manager is judging the riskiness of a complex system over a given, fixed time horizon. Under various assumptions of the random factors that influence the performance of the system the risk manager has to judge today how the system will perform by the end of the given period. At this point, this no doubt sounds somewhat vague, but later in this introduction we give some more precise examples where we feel that the theory as presented may eventually find natural applications.

A first question we would like to address is

"Why we two?"

There are several reasons, some of which we briefly like to mention, especially as they reflect not only our collaboration but also the way QRM as a field of research and applications is developing. Both being born in towns slightly below or above sea level, Amsterdam and Antwerp, risk was always a natural aspect of our lives. For the second author this became very explicit as his date of birth, February 3, 1953, was only two days after the disastrous flooding in Holland. In the night of January 31 to February 1, 1953, several 100 km of dykes along the Dutch coast were breached in a severe storm. The resulting flooding killed 1836 people, 72 000 people needed to be evacuated, nearly 50 000 houses and farms and over 200 000 ha of land were flooded. A local newspaper, *De Yssel- en Lekstreek*, on February 6, 1953 ran a headline "Springtij en orkaan veroorzaken nationale ramp. Nederland in grote watersnood"[1]. The words of the Dutch writer Marsman from 1938 came back to mind: "En in alle gewesten, wordt de stem van het water, met zijn eeuwige rampen, gevreesd en gehoord."[2] As a consequence, the Delta Project came into being with a clear aim to build up a long-lasting coastal protection through an elaborate system of dykes and sluices. Though these defense systems could never guarantee 100% safety for the population at risk, a safety margin of 1 in 10 000 years for the so-called Randstad (the larger area of land around Amsterdam and Rotterdam) was agreed upon. Given these safety requirements, dyke heights were calculated, e.g. 5.14 m above NAP (Normaal Amsterdams Peil). A combination of environmental, socioeconomic, engineering and statistical considerations led to the final decision taken for the dyke and sluice constructions. For the Dutch population, the words of Andries Vierlingh from the book *Tractaet van Dyckagie* (1578) "De meeste salicheyt hangt aen de hooghte van eenen dyck"[3] summarized the feeling of the day. From a stochastic modelling point of view, the methodology entering the solution of problems encountered in the *Delta Project* is very much related to the analysis of extremes. Several research projects related to the modelling of extremal events emerged, examples of which include our PhD theses Balkema [1973] and Embrechts [1979]. Indirectly, events and discussions involving risk and extremes have brought us together over many years.

By now, the stochastic modelling of extremes, commonly referred to as *Extreme Value Theory* (EVT), has become a most important field of research, with numerous key contributors all over the world. Excellent textbooks on the subject of EVT exist or are currently being written. Moreover, a specialized journal solely devoted to the stochastic theory of extremes is available (*Extremes*). Whereas the first author (Balkema) continued working on fundamental results in the realm of heavy tailed phenomena, the second author (Embrechts) became involved more in areas related to finance, banking, insurance and risk management. Banking and finance have their own tales of extremes. So much so that Alan Greenspan in a presentation to the Joint Central Bank Research Conference in Washington D.C. in 1995 stated (see Greenspan [1996]):

> "From the point of view of the risk manager, inappropriate use of the normal distribution can lead to an understatement of risk, which must be

[1] "Spring tide and hurricane cause a national disaster. The Netherlands in severe water peril."
[2] "And in every direction, one hears and fears the voice of the water with its eternal perils."
[3] "Most of the happiness depends on the height of a dyke."

> balanced against the significant advantage of simplification. From the central bank's corner, the consequences are even more serious because we often need to concentrate on the left tail of the distribution in formulating lender-of-last resort policies. Improving the characterization of the distribution of extreme values is of paramount concern."

Also telling is the following statement taken from *Business Week* in September 1998, in the wake of the LTCM hedge fund crisis:

> "Extreme, synchronized rises and falls in financial markets occur infrequently but they do occur. The problem with the models is that they did not assign a high enough chance of occurrence to the scenario in which many things go wrong at the same time – the 'perfect storm' scenario."

Around the late nineties, we started discussions on issues in QRM for which further methodological work was needed. One aim was to develop tools which could be used to model markets under extreme stress scenarios. The more mathematical consequence of these discussions you are holding in your hands.

It soon became clear to us that the combination of extremes and high dimensions, in the context of scenario testing, would become increasingly important. So let us turn to the question

"Why in the first part of the title High Risk Scenarios and Extremes?"

The above mentioned Delta Project and QRM have some obvious methodological similarities. Indeed protecting the coastal region of a country from sea surges through a system of dykes and sluices can be compared with protecting the financial system (or bank customers, insurance policy holders) from adverse market movements through the setting of a sufficiently high level of regulatory risk capital or reserve. In the case of banking, this is done through the guidelines of the Basel Committee on Banking Supervision. For the insurance industry, a combination of international guidelines currently under discussion around Solvency 2 and numerous so-called local statutory guidelines have been set up. The concept of dyke height in the Delta Project translates into the notion of risk measure, in particular into the widely used notion of *Value-at-Risk* (VaR). For instance, for a given portfolio, a 99% 10-day VaR of one million euro means that the probability of incurring a portfolio loss of one million euro or more by the end of a two-week (10 trading days) period is 1%. The 10 000 year return period in the dyke case is to be compared with the 99% confidence level in the VaR case. Both sea surges and market movements are complicated functions of numerous interdependent random variables and stochastic processes. Equally important are the differences. The prime one is the fact that the construction of a dyke concerns the modelling of natural (physical, environmental) processes, whereas finance (banking) is very much about the modelling of social phenomena. Natural events may enter as triggering events for extreme market movements but are seldom a key modelling ingredient. An example where a natural event caused more than just a stir for the bank involved was the Kobe earthquake and its implications for the downfall of Barings Bank; see Boyle & Boyle [2001]. For the life insurance industry, stress events with major consequences are pandemics for instance. Also relevant are considerations concerning longevity and of course market movements, especially related to interest rates. Moving to the non-life insurance and reinsurance industry, we encounter increasingly the relevance of the modelling of extreme natural phenomena like storms, floods and earthquakes. In between we have for instance acts of terrorism like the September 11 attack. The "perfect storm scenario" where many things go wrong at the same time is exemplified through the stock market decline after the New Economy hype, followed by a longer period of low interest rates which caused considerable problems for the European life insurance industry. This period of economic stress was further confounded by increasing energy prices and accounting scandals.

In order to highlight more precisely the reasons behind writing these lectures, we will restrict our attention below to the case of banking. Through the Basel guidelines, very specific QRM needs face that branch of the financial industry. For a broad discussion of concepts, techniques and tools from QRM, see McNeil, Frey & Embrechts [2005] and the references therein. Besides the regulatory side of banking supervision, we will also refer to the example of portfolio theory. Here relevant references are for instance Korn [1997] and Fernholz [2002]. Under the *Basel guidelines* (www.bis.org/bcbs) for market risk banks calculate VaR; this involves a holding period of 10 days at the 99% confidence level for regulatory (risk) capital purposes and 1 day 95% VaR for setting the bank's internal trading limits. Banks and regulators are well aware of the limitations of the models and data used so that, besides the inclusion of a so-called multiplier in the capital charge formula, banks complement their VaR reporting with so-called stress scenarios. These may include larger jumps in key market factors like interest rates, volatility, exchange rates, etc. The resulting question is of the "what if"-type. *What* happens to the bank's market position *if* such an extreme move occurs. Other stress scenarios may include running the bank's trading book through some important historical events like the 1987 crash, the 1998 LTCM case or September 11. Reduced to their simplest, but still relevant form, the above stress scenarios can be formalized as follows. Suppose that the market (to be interpreted within the CAPM-framework, say; see Cochrane [2001]) moves strongly against the holder of a particular portfolio. Given that information, what can be said about risk measurement features of that portfolio. Another relevant question in the same vein is as follows. Suppose that a given (smaller) portfolio moves against the holder's interest and breaches a given risk management (VaR) limit. How can one correct some (say as few as possible) positions in that portfolio so that the limit is not breached anymore. For us, motivating publications dealing with this type of problem are for instance Lüthi & Studer [1997], Studer [1997] and Studer & Lüthi [1997].

The *high-dimensionality* within our theory is related to the number of assets in the portfolio under consideration. Of course, in many applications in finance, dimension reduction techniques can be used in order to reduce the number of assets to an effective dimensionality which often is much lower and indeed more tractable. The decision to be made by the risk manager is to what extent important information may have been lost in that process. But even after a successful dimension reduction, an effective dimensionality between five and ten, say, still poses considerable problems for the application of standard EVT techniques. By the nature of the problem extreme observations are rare. The curse of dimensionality very quickly further complicates the issue.

In recent years, several researchers have come up with high-dimensional (market) models which aim at a stochastic description of macro-economic phenomena. When we restrict ourselves to the continuous case, the multivariate normal distribution sticks out as the benchmark model par excellence. Besides the computational advantages for the calculation of various relevant QRM quantities such as risk measures and capital allocation weights, it also serves as an input to the construction of more elaborate models. For instance the widely used *Student t* model can be obtained as a random mixture of multivariate normals. Various other examples of this type can be worked out leading to the class of elliptical distributions as variance mixture normals, or beyond in the case of mean-variance mixture models. Chapter 3 in McNeil, Frey & Embrechts [2005] contains a detailed discussion of elliptical distributions; a nice summary with emphasis on applications to finance is Bingham & Kiesel [2002]. A useful set of results going back to the early development of QRM leads to the conclusion that within the class of *elliptical models*, standard questions asked concerning risk measurement and capital allocation are well understood and behave much as in the exact multivariate normal case. For a concrete statement of these results, see Embrechts, McNeil & Straumann [2002]. A meta-theorem, however, says that as soon as one deviates from this class of elliptical models, QRM becomes much more complicated. It also quickly becomes context- and application-dependent. For instance, in the elliptical world, VaR as a risk measure is subadditive meaning that the VaR of a sum of risks is bounded above by the sum of the individual VaRs. This property is often compared to the notion of diversification, and has a lot to do with some of the issues we discuss in our book. As an example we briefly touch upon the current important debate on the modelling of *operational risk* under the Advanced Measurement Approach (AMA) which is based on the Loss Distribution Approach (LDA); once more, for detailed references and further particulars on the background, we refer to McNeil, Frey & Embrechts [2005]. For our purposes it suffices to realize that, beyond the well-known risk categories for market and credit risk, under the new Basel Committee guidelines (so-called Basel II), banks also have to reserve (i.e. allocate regulatory risk capital) for operational risk. According to Basel II, *Operational Risk* is defined as the risk of loss resulting from inadequate or failed internal processes, people or systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk; for details on the regulatory framework, see www.bis.org/bcbs/. Under the LDA, banks are typically structured into eight business lines and seven risk categories based on the type of operational loss. An example is corporate finance (business line) and internal fraud (risk type). Depending on the approach followed, one has either a 7-, 8-, or 56-dimensional problem to model. Moreover, an operational risk capital charge is calculated on a yearly basis using VaR at the 99.9% level. Hence one has to model a 1 in 1000 year event. This by every account is *extreme*. The high dimensionality of 56, or for some banks even higher, is obvious. The subadditivity question stated above is highly relevant; indeed a bank can add up VaRs business line-wise, risk type-wise or across any relevant subdivision of the 8×7 loss matrix. A final crucial point concerns the reduction of these sums of VaRs taking "diversification effects" into account. This may (and typically does) result in a rather intricate analysis where concepts like risk measure coherence (see Artzner et al. [1999]), EVT and copulas (non-linear dependence) enter in a fundamental way. Does the multivariate extreme value theory as it is presented on the pages that follow yield solutions to the AMA-LDA discussion above? The reader will not find ready-made models for this discussion. However, the operational risk issue briefly outlined above makes it clear that higher dimensional models are called for, within which questions on extremal events are of paramount importance. We definitely provide a novel approach for handling such questions in the future. Admittedly, as the theory is written down so far, it still needs a considerable amount of work before concrete practical consequences emerge. This situation is of course familiar from many (if not all) methodological developments. Besides the references above, the reader who is in particular interested in the operational risk example, may consult Chavez-Demoulin, Embrechts & Nešlehová [2006] and Nešlehová, Embrechts & Chavez-Demoulin [2006]. From information on operational risk losses available so far, one faces models that are skew and (very) heavy-tailed. Indeed, it is the non-repetitive (low-frequency) but high-severity losses that are of main concern. This immediately rules out the class of elliptical distributions. Some of the models discussed in our book will come closer to relevant alternatives. We are not claiming that the theory presented will, in a not too distant future, come up with a useful 56-dimensional model for operational risk. What we are saying, however, is that the theory will yield a better understanding of quantitative questions asked concerning extremal events for high-dimensional loss portfolios.

Mathematicians are well advised to show humbleness when it comes to model formulation involving uncertainty, especially in the field of economics. In a speech entitled "Monetary Policy Under Uncertainty" delivered in August 2003 in Jackson Hole, Wyoming, Alan Greenspan started with the following important sentence: "Uncertainty is not just an important feature of the monetary policy landscape; it is the defining characteristic of that landscape." He then continued with some sentences which are occasionally referred to, for instance by John Mauldin, as *The Greenspan Uncertainty Principle*:

> "Despite the extensive efforts to capture and quantify these key macroeconomic relationships, our knowledge about many of the important linkages is far from complete and in all likelihood will always remain so. Every model, no matter how detailed and how well designed conceptually and empirically, is a vastly simplified representation of the world that we experience with all its intricacies on a day-to-day basis. Consequently, even with large advances in computational capabilities and greater comprehension of economic linkages, our knowledge base is barely able to keep pace with the ever-increasing complexity of our global economy."

And further,

> "Our problem is not the complexity of our models but the far greater complexity of a world economy whose underlying linkages appear to be in a continual state of flux... In summary then, monetary policy based on risk management appears to be the most useful regime by which to conduct policy. The increasingly intricate economic and financial linkages in our global economy, in my judgment, compel such a conclusion."

For many questions in practice, and in particular for questions related to the economy at large, there is no such thing as *the model*. Complementary to the quotes above, one can say that so often the road towards finding a model is far more important than the resulting model itself. We hope that the reader studying the theory presented in this book will enjoy the trip more than the goals reached so far. We have already discussed some of the places we will visit on the way.

One of the advantages of modern technology is the ease with which all sorts of information on a particular word or concept can be found. We could not resist googling "High Risk-Scenarios". Needless to say that we did *not* check all 11 300 000 entries which we obtained in 0.34 seconds. It is somewhat disturbing, or one should perhaps say sobering, that our book will add just one extra entry to the above list. The more correct search, keeping the three words linked as in the title of our book, yielded a massive reduction to an almost manageable 717. Besides the obvious connections with the economic and QRM literature, other fields entering included terrorism, complex real-time systems, environmental and meteorological disasters, biosecurity, medicine, public health, chemistry, ecology, fire and aviation, Petri nets or software development. Looking at some of these applications it becomes clear that there is no common understanding of the terminology. From a linguistic point of view, one could perhaps query the difference between "High-Risk Scenario" and "High Risk-Scenario". Rather than doing so, we have opted for the non-hyphenated version. In its full length "High Risk Scenarios and Extremes" presents a novel mathematical theory for the analysis of extremes in multi-dimensional space. Especially the econometric literature is full of attempts to describe such models. Relevant for our purposes are papers like Pesaran, Schuermann & Weiner [2004], Pesaran & Zaffaroni [2004], Dees et al. [2007], and in particular the synthesis paper Pesaran & Smith [2006] on the so-called global modelling approach. From a more mathematical finance point of view, Platen [2001] and Platen [2006], and Fergusson & Platen [2006] describe models to which we hope our theory will eventually be applicable. Further relevant publications in this context are Banner, Fernholz & Karatzas [2006] and Fernholz [2002].

In the preceding paragraphs, we explained some of our motivations behind the first part of the title: "High Risk Scenarios and Extremes". The next, more mathematical question is

"Why the second part of the title, A Geometric Approach?"

A full answer to this question will become clear as the reader progresses through the pages that follow. There are various approaches possible towards a multivariate theory of extremes, most of these being coordinatewise theories. This means that, starting from a univariate EVT, a multivariate version is developed which looks at coordinate maxima and their weak limit laws under appropriate scaling. Then the key question to address concerns the dependence between the components of the nondegenerate limit. In the pages that follow, we will explain that, from a mathematical point of view, a more geometrical, coordinate-free approach towards the stochastic modelling is not only mathematically attractive, but also very natural from an applied point of view. For this, first recall the portfolio link stated above. A portfolio is merely a linear combination of underlying risk factors $X_1, \ldots, X_d$ with weights $w_1, \ldots, w_d$. Here X_i stands for the future, one-period value of some underlying financial instrument. The hopefully rare event that the value of the portfolio

$$V(\boldsymbol{w}) = \sum_{i=1}^{d} w_i X_i$$

is low can be expressed as $\{V(\boldsymbol{w}) \le q\}$ where q is some value determined by risk management considerations. A value below q should only happen with a very small probability. Now, of course, the event $\left\{\sum_{i=1}^{d} w_i X_i \le q\right\}$ has an immediate geometric interpretation as the vector $(X_1, \ldots, X_d)$ hitting a halfspace determined by the portfolio weights $(w_1, \ldots, w_d)$ at the critical level q. Furthermore, depending on the type of position one holds, the signs of the individual w_i's will be different: in portfolio language, one moves from a long to a short position. Further, the world of financial derivatives allows for the construction of portfolios, the possible values of which lie in specific subspaces of $\mathbb{R}^d$. The first implication is that one would like to have a broad theory that yields the description of rare events over a wide range of portfolio positions. The geometry enters naturally through the description of this rare event set as a halfspace. A natural question then to ask is *what do we know about the stochastic behaviour of* $(X_1, \ldots, X_d)$ *given that such a rare event has occurred?* Thus a theory is needed which yields results on the conditional distribution of a random vector (the risk factors) given that a linear combination of these factors (a portfolio position or market index) surpasses a high (rare) value.

The interpretation of high or rare value depends on the kind of position taken, hence in the first instance, the theory should allow for resulting halfspaces to drift to infinity in a general (non-preferred) direction. This kind of isotropic limit nevertheless yields a rich theory covering many of the standard examples in finance and insurance.

At the same time, however, one also needs to consider theories for multivariate extremes where the rare event or high risk scenario corresponds to a "drifting off" to infinity in one specific direction. This of course is the case when one is interested in one particular portfolio with fixed weights over the holding period (investment horizon) of the portfolio. Another example concerns the operational risk problem discussed above. Here the one-year losses correspond to random variables $L_1, \ldots, L_d$ where, depending on the approach used, d can stand for eight business lines, seven loss types or fifty-six combinations of these. Under Basel II, banks have to come up with a risk measure for the total loss $L_1 + \cdots + L_d$ and hence a natural question to ask is the limiting behaviour of the conditional distribution of the vector $(L_1, \ldots, L_d)$ given that $L_1 + \cdots + L_d$ is large. This is an example where one is interested in the conditional behaviour of the risk factors $(L_1, \ldots, L_d)$ in the direction given by the vector $(1, \ldots, 1)$. The mathematics entering the theory of multivariate extremes in a particular direction in $\mathbb{R}^d$ is different from the "isotropic" theory mentioned above and translates into different invariance properties of classes of limit laws under appropriate transformations. Examples of research papers where the interplay between geometrical thinking and the discussion of multivariate rare events are to be found include for instance Hult & Lindskog [2002] and Lindskog [2004]. The latter PhD thesis also contains a nice summary of the various approaches available to the multivariate theory of regular variation and its applications to multivariate extreme value theory. Besides the various references on this topic presented later in the text, we also like to mention Fougères [2004] and Coles & Tawn [1991]. The necessary statistical theory is nicely summarized in Coles [2001].

Perhaps an extra remark on the use of geometric arguments, mainly linked to invariance properties and symmetry arguments is in order. It is no doubt that one of the great achievements of 19th century and early 20th century mathematics is the introduction of abstract tools which contribute in an essential way to the solution of applied problems. Key examples include the development of Galois Theory for the solution of polynomial equations or Lie groups for the study of differential equations. By now both theories have become fundamental for our understanding of natural phenomena like symmetry in crystals, structures of complex molecules or quantum behaviour in physics. For a very readable, non-technical account, see for instance, Ronan [2006]. We strongly believe that geometric concepts will have an important role to play in future applications to Quantitative Risk Management.

By now, we made it clearer why we have written this text; the motivation comes definitely from the corner of QRM in the realm of mainly banking and to some extent insurance. An alternative title could have been "Stress testing methodology for multivariate portfolios", though such a title would have needed a more concrete set of tools for immediate use in the hands of the (financial) portfolio manager. We are not yet at that level. On the other hand, the book presents a theory which can contribute to the discussion of stress testing methodology as requested, for instance,gp in statements of the type

> "Banks that use the internal models approach for meeting market risk capital requirements must have in place a rigorous and comprehensive stress testing program. Stress testing to identify events or influences that could greatly impact banks is a key component of a bank's assessment of its capital position."

taken from Basel Committee on Banking Supervision [2005]. Over the years, numerous applications of EVT methodology to this question of stress testing within QRM have been worked out. Several examples are presented in McNeil, Frey & Embrechts [2005] and the references therein; further references beyond these include Bensalah [2002], Kupiec [1998] and Longin [2000]. There is an enormous literature on this topic, and we very much hope that academics and practitioners contributing to and interested in this ever-growing field will value our contribution and indeed help in bringing the theory presented in this book to full fruition through real applications. One of the first tasks needed would be to come up with a set of QRM questions which can be cast in our geometric approach to high risk stress scenarios. Our experience so far has shown that such real applications can only be achieved through a close collaboration between academics and practitioners. The former have to be willing (and more importantly, capable) to reformulate new mathematical theory into a language which makes such a discussion possible. The latter have to be convinced that several of the current quantitative questions asked in QRM do require new methodological tools. In that spirit, the question

"For whom have we written this book?"

should in the first instance be answered by: "For researchers interested in understanding the mathematics of multivariate extremes." The ultimate answer should be "For researchers and practitioners in QRM who have a keen interest in understanding the extreme behaviour of multivariate stochastic systems under stress". A key example of such a system would be a financial market. At the same time, the theory presented here is not only coordinate-free, but also application-free. As a consequence, we expect that the book may appeal to a wider audience of “extreme value adepts”. One of the consequences of modern society with its increasing technological skills and information technology possibilities is that throughout all parts of science, large amounts of data are increasingly becoming available. This implies that also more information on rare events is being gathered. At the same time, society cares (or at least worries) about the potential impact of such events and the necessary steps to be taken in order to prevent the negative consequences. Also at this wider level, our book offers a contribution to the furthering of our understanding of the underlying methodological problems and issues.

It definitely was our initial intention to write a text where (new) theory and (existing) practice would go more hand in hand. A quick browse through the pages that follow clearly shows that theory has won and applications are yet to come. This of course is not new to scientific development and its percolation through the porous sponge of real applications. The more mathematically oriented reader will hopefully find the results interesting; it is also hoped that she will take up some of the scientific challenges and carry them to the next stage of solution. The more applied reader, we very much hope, will be able to sail the rough seas of mathematical results like a surfer who wants to stay near the crest of the wave and not be pulled down into the depths of the turbulent water below. That reader will ideally guide the former into areas relevant for real applications. We are looking forward to discuss with both.

Preview

The Preview presents a tour along the main results of these lecture notes. It introduces concepts and notation that will be used throughout the book. It should help the reader to follow the thrust of the ideas developed in the individual lectures, and to determine which lectures are of sufficient interest to merit a closer look.

A recipe

The question addressed in these lectures is simple: Given a multivariate sample cloud, what can one say about the underlying distribution in a region containing only one or two points of the sample?

Typically the region is a halfspace, and one is concerned about the eventuality of future data points lying far out in the region. We shall use the terminology of financial mathematics and speak of loss and risk. The data cloud could just as well contain data of insurance claims, or data from quality control, biomedical research, or meteorology. In all cases one is interested in the extremal behaviour at the edge of the sample cloud, and one may use the concepts of risk and loss. In a multivariate setting *risk* and loss may be formalized as functions which increase as one moves further out into the halfspace.

In first instance the answer to the question above is: "Nothing". There are too few points to perform a statistical analysis. However some reflection suggests that one could use the whole sample to fit a distribution, say a Gaussian density, and use the tails of this density to determine the conditional distribution on the given halfspace. In financial mathematics nowadays one is very much aware of the dangers of this approach. The Gaussian distribution gives a good fit for the daily log returns, but not in the tails. So the proper recipe should be: Fit a distribution to the data, and check that the tails fit too. If one can find a distribution, Gaussian say, or elliptic Student, that satisfies these criteria, then this solves the problem, and we are done. In that case there is no need to read further.

What happens if the data cloud looks as if it may derive from a normal distribution, but has heavy tails? There is a convex central black region surrounded by a halo of isolated points. The cloud does not exhibit any striking directional irregularities. Such data sets have been termed *bland* by John Tukey. Only statistical analysis is able to elicit information from bland clouds.

Rather than fitting a distribution to the whole cloud, we shall concentrate on the tails. We assume some regularity at infinity. In finite points regularity is expressed by the existence and continuity of a positive density at those points. Locally the distribution will then look like the *uniform distribution*; under proper scaling the sample cloud will converge vaguely to the standard Poisson point process on $\mathbb{R}^d$ as the number of points in the sample increases. We want to perform a similar analysis at infinity. Of course, in a multivariate setting there are many ways in which halfspaces may diverge. This problem is inherent to multivariate extremes. In order to obtain useful results, we have to introduce some regularity in the model setup.

Ansatz. Conditional distributions on halfspaces with relatively large overlap asymptotically have the same shape.

Let us make the content of the Ansatz more precise.

Definition. Two probability distributions (or random vectors Z and W) have the same *shape* or are of the same *type* if they are *non-degenerate*, and if there exists an affine transformation α such that Z is distributed like $\alpha(W)$. A random vector Z has a *degenerate* distribution if it lives on a hyperplane, equivalently, if there exists a linear functional $\xi \neq 0$ and a real constant c such that $\xi Z = c$ a.s.

For instance, all Gaussian densities have the same shape. Shape (or type) is a geometric concept. Given a non-degenerate Gaussian distribution on $\mathbb{R}^d$ one can find coordinates such that in these coordinates the distribution is standard Gaussian with density

$$e^{-(w_1^2+\cdots+w_d^2)/2}/(2\pi)^{d/2}, \quad w = (w_1, \ldots, w_d) \in \mathbb{R}^d.$$

A basic theorem in this setting is the Convergence of Types Theorem (CTT). It allows us to speak of a sequence of vectors as being asymptotically Gaussian. We write $Z_n \Rightarrow Z$ if the distribution functions (dfs) of Z_n converge weakly to the distribution function (df) of Z.

Theorem 1 (Convergence of Types). *If $Z_n \Rightarrow Z$ and $W_n \Rightarrow W$, where W_n and Z_n are of the same type for each n, then the limit vectors, if non-degenerate, are of the same type.*

Proof. See Fisz [1954] or Billingsley [1966]. □

At first sight the CTT may look rather innocuous. In many applied probability questions involving limit theorems it works like a magic hat from which new models may be pulled: In the univariate setting the Central Limit Problem for partial sums yields the stable distributions; the Extreme Value Problem for partial maxima yields the extreme value distributions. See Embrechts, Klüppelberg & Mikosch [1997], Chapters 2 and 3. In the multivariate setting, in Chapter II below, the CTT yields the well-known multivariate max-stable laws; in Chapter III the CTT yields a continuous one-parameter family of limit laws; and in Chapter IV the CTT yields two semi-parametric families of high risk limit laws, one for exceedances over (horizontal) linear thresholds, one for exceedances over elliptic thresholds.

Let us try to give the intuition behind the CTT in the case of a Gaussian limit.

A sequence of random vectors Z_n is asymptotically normal if there exist affine normalizations α_n such that

$$W_n := \alpha_n^{-1}(Z_n) \Rightarrow W,$$

where W is standard normal. The validity of the term asymptotic normality would seem to derive from geometric insight. In geometric terms one may try to associate with each Z_n an ellipsoid E_n that is transformed into the unit ball B by the normalization. These ellipsoids $E_n = \alpha_n(B)$ may be related to the expectation and covariance of Z_n (if these exist and converge), or to certain convex *level sets* of the density of Z_n (if the density exists and is *unimodal*). Perhaps the correct intuition is that large sample clouds from distributions that are asymptotically Gaussian are asymptotically elliptic, and that affine transformations that map the elliptic sample clouds into spherical sample clouds may be used to normalize the distributions. The normalizations are thus determined geometrically. The same geometric intuition forms the background to these lectures. Instead of convergence of the whole sample cloud we now assume convergence at the edge. Since we want to keep sight of individual sample points, we assume convergence to a point process.

Affine transformations are needed to pull back the distributions as the halfspaces diverge. Let us say a few words about the space $\mathcal{A} = \mathcal{A}(d)$ of affine transformations on $\mathbb{R}^d$. Recall that an *affine transformation* has the form

$$w \mapsto z = \alpha(w) = Aw + a, \tag{1}$$

where a is a vector in $\mathbb{R}^d$, and A an invertible matrix of size d. The inverse is

$$z \mapsto w = \alpha^{-1}(z) = A^{-1}(z - a).$$

The set $\mathcal{A}$ is a group since $\alpha^{-1} \in \mathcal{A}$, and the composition of two affine transformations is an affine transformation:

$$w \mapsto (\alpha\beta)(w) = \alpha(\beta(w)) = A(Bw + b) + a.$$

Convergence $\alpha_n \to \alpha$ means $a_n \to a$ and $A_n \to A$ componentwise, or, equivalently, $\alpha_n(w) \to \alpha(w)$ for all $w \in \mathbb{R}^d$, or for $w = 0, e_1, \dots, e_d$, where $e_1, \dots, e_d$ are linearly independent vectors. From linear algebra it is known that $A \mapsto A^{-1}$ is continuous on the group $\mathrm{GL}(d)$ of invertible matrices of size d. Hence $\alpha \mapsto \alpha^{-1}$ is continuous. Obviously $(\alpha, \beta) \mapsto \alpha\beta$ is continuous. So $\mathcal{A}$ is a topological group. It is even a Lie group. It is possible to see $\mathcal{A}(d)$ as a subgroup of $\mathrm{GL}(1 + d)$ by writing

$$\alpha(w) = Aw + a = z \iff \begin{pmatrix} 1 & 0 \\ a & A \end{pmatrix} \begin{pmatrix} 1 \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ a + Aw \end{pmatrix} = \begin{pmatrix} 1 \\ z \end{pmatrix}. \tag{2}$$

This representation makes it possible to apply standard results from linear algebra when working with affine transformations. We usually write Z for the random vector with components Z_i and

$$W = \alpha^{-1}(Z) = A^{-1}(Z - a) \tag{3}$$

for the normalized vector. Now assume $\alpha_n^{-1}(Z_n) \Rightarrow W$ where W is non-degenerate. The normalizations α_n are not unique. They may be replaced by normalizations β_n which are *asymptotic* to α_n:

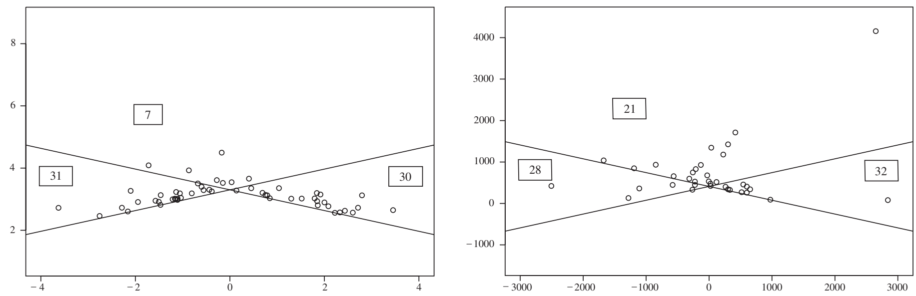
$$\beta_n \sim \alpha_n \iff \alpha_n^{-1}\beta_n \to \text{id}, \tag{4}$$

where id stands for the identity transformation. Asymptotic equality is an equivalence relation for sequences in $\mathcal{A}$.

Warning. If $\alpha_n \to \text{id}$ then $\alpha_n^{-1} \to \text{id}$. However $\alpha_n \sim \beta_n$ does not imply $\alpha_n^{-1} \sim \beta_n^{-1}$, not even in dimension $d = 1$. Here is a simple counterexample:

Example 2. Let X_n be uniformly distributed on the interval $(1, n+1)$. Properly normalized, the X_n converge in distribution to a rv U which is uniformly distributed on the interval $(0, 1)$. Indeed $(X_n - 1)/n \Rightarrow U$; but also $X_n/n \Rightarrow U$. Set $\alpha_n(u) = nu + 1$ and $\beta_n(u) = nu$. Then $\beta_n^{-1}\alpha_n(u) \to u$ but $\alpha_n\beta_n^{-1}(x) = x + 1$. So $\alpha_n \sim \beta_n$ does not imply $\alpha_n^{-1} \sim \beta_n^{-1}$. Indeed, asymptotic equality means that the normalized variables $\alpha_n^{-1}(X_n)$ and $\beta_n^{-1}(X_n)$ are close, not the approximations $\alpha_n(U)$ and $\beta_n(U)$. ◊

After this digression on shape, geometry and affine transformations, let us return now to the basic question of determining the distribution on a halfspace containing only a few (or no) points of the sample, and to our Ansatz that high risk scenarios



Standard normal   Spherical Cauchy

Exceedances over linear thresholds with varying direction, for 40 000 points. Boxed are the number of sample points in the halfplanes, and in the intersection. In the Cauchy sample many points lie outside the figure; the highest is (27 000, 125 000).

on halfspaces with relatively large overlap have distributions with approximately the same shape. Here a *high risk scenario* of a random vector Z for a given halfspace H is just the vector Z conditioned to lie in H. For halfspaces far out this corresponds to our interpretation of a rare or extreme event. The reader may wonder whether the Ansatz implies that all high risk scenarios asymptotically have the same shape. Note that the condition of a relatively large overlap is different for light tails and for heavy tails. For a Gaussian distribution the directions of two halfspaces far out have to be close to have a relatively large overlap; for a spherical *Cauchy distribution* there is considerable overlap even if the directions of the halfspaces are orthogonal. See the figure above.

In the univariate case the condition that high risk scenarios, properly normalized, converge leads to a one-parameter family of limit shapes, the *generalized Pareto distributions*. These *GPD*s may be standardized to form a continuous one-parameter family, indexed by $\tau \in \mathbb{R}$:

$$G_\tau(v) = 1 - (1 + \tau v)_+^{-1/\tau}, \quad v \geq 0. \tag{5}$$

By continuity G_0 is the standard exponential df. The associated univariate limit theory has been applied in many fields. It is our aim to develop a corresponding theory in the multivariate setting.

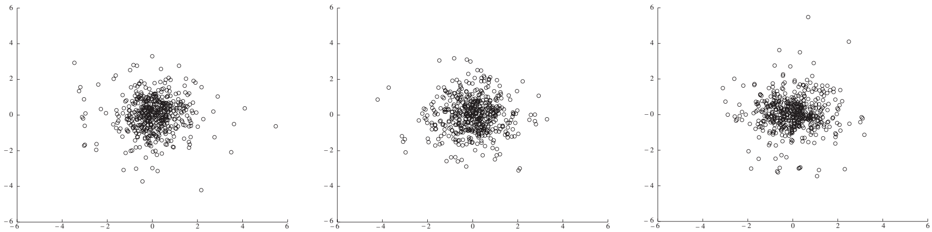
We shall denote the *high risk scenario* for Z associated with the halfspace H by Z^H. By definition, Z^H lives on H, and for any Borel set E

$$\mathbb{P}\{Z^H \in E\} = \mathbb{P}\{Z \in E \cap H\}/\mathbb{P}\{Z \in H\}.$$

Halfspaces are assumed to be closed, and $\mathbb{P}\{Z \in H\}$ is assumed positive.

One may impose the condition that the high risk scenarios Z^{H_n}, properly normalized, converge for any sequence of halfspaces H_n under the sole restriction that $\mathbb{P}\{Z \in H_n\}$ is positive and vanishes for $n \to \infty$. This assumption is quite strong. It presupposes a high degree of directional homogeneity in the halo of the sample cloud. In order to understand multivariate tail behaviour, a thorough analysis of the consequences of this strong assumption seems like a good starting point. This analysis is given in Chapter III, the heart of the book. As an illustration we exhibit below the three plane projections of a data cloud in $\mathbb{R}^3$, consisting of the log-returns of three stocks on the Dutch stock exchange AEX over the period from 2-2-04 until 31-12-05. The data were kindly made available by Newtrade Research.

One may also start with the weaker assumption that the high risk scenarios converge for halfspaces which diverge in a certain direction. This is done in Chapter IV for horizontal halfspaces. The Ansatz now holds only for horizontal halfspaces. Write $z = (x, y)$ where y is the vertical component of z and x the h-dimensional horizontal part, with $h = d - 1$. Similarly write $Z = (X, Y) \in \mathbb{R}^{h+1}$. High risk scenarios for *horizontal halfspaces* $H^y = \mathbb{R}^h \times [y, \infty)$ correspond to *exceedances* over *horizontal*



ING - Allianz   Allianz - ASML   ASML - ING

Bland sample clouds: bivariate marginals of daily log-returns ING - Allianz - ASML.

thresholds. Let $\alpha(y)$ be affine transformations mapping $H_+ = \{y \geq 0\}$, the *upper halfspace*, onto H^y. The vectors $W_y = \alpha(y)^{-1}(Z^{H^y})$ live on H_+. Now suppose that the $\alpha(y)$ yield a limit vector:

$$W_y := \alpha(y)^{-1}(Z^{H^y}) \Rightarrow W, \quad \mathbb{P}\{Y \geq y\} \to 0. \tag{6}$$

Assume the limit is non-degenerate. What can one say about its distribution? It is not difficult to see that $\alpha(y)$ maps horizontal halfspaces into horizontal halfspaces, and that this implies that the high risk scenarios $Y^{[y,\infty)}$ of the vertical coordinate, with the corresponding normalization, converge to the vertical coordinate V of the limit vector, $W = (U, V)$. By the univariate theory the vertical coordinate of the limit vector has a GPD, see (5).

One may prove more. Suppose (6) holds. Let $Z_1, Z_2, \ldots$ be independent observations from the distribution π of Z. Choose y_n so that $\mathbb{P}\{Y \geq y_n\} \sim 1/n$. Set $\alpha_n = \alpha(y_n)$ where $\alpha(y)^{-1}(Z^{H^y}) \Rightarrow W$ as above. Then the normalized sample clouds converge in distribution to a Poisson point process:

$$N_n := \{\alpha_n^{-1}(Z_1), \ldots, \alpha_n^{-1}(Z_n)\} \Rightarrow N_0. \tag{7}$$

The mean measures of the sample clouds N_n converge weakly to the *mean measure* ρ of the limiting Poisson point process N_0 on all horizontal halfspaces J on which ρ is finite:

$$\rho_n = n\alpha_n^{-1}(\pi) \to \rho \text{ weakly on } J, \quad \rho(J) < \infty. \tag{8}$$

The restriction of ρ to H_+ is a probability measure, the distribution of W.

The equivalence of the two limit relations (6) and (7) is the Extension Theorem in Section 14.6. It is a central result. The first limit relation is analytical. It raises questions such as:

1) What limit laws are possible?

2) For a given limit law, what conditions on the distribution of Z will yield convergence?

The second relation is more geometric. Here one may ask:

1) Does convergence in (8) also hold for halfspaces J which are close to horizontal?

2) Will the convex hull of the normalized sample cloud converge to the convex hull of N_0?

For the novice to the application of point process methodology to extreme value problems this all may seem to go a bit too fast. Modern *extreme value theory* with its applications to more involved problems in risk management, however, needs this level of abstraction. See McNeil, Frey & Embrechts [2005] for a good discussion of these issues. In the one-dimensional case one already needs such a theory for understanding the *Peaks Over Thresholds* method based on (5), or the limit behaviour of several order statistics, as will be seen in Section 6.4. So bear with us and try to follow the general scheme.

The vector W in (6) is the limit of high risk scenarios for horizontal halfspaces. It follows that the high risk scenarios W^{H^y} all have the same shape. This result follows from the trivial fact that a high risk scenario of a high risk scenario is again a high risk scenario, at least for horizontal halfspaces. In the univariate setting, the exponential distribution, the uniform distribution and the Pareto distributions all have the *tail property*: Any tail of the distribution is of the same type as the whole distribution. In fact this tail property characterizes the class of GPDs. It may also explain why these distributions play such an important role in applications in insurance and risk theory. In the multivariate setting the tail property is best formulated in terms of the infinite measure ρ in (8): There exists a *one-parameter group*[4] of affine transformations γ^t, $t \in \mathbb{R}$, such that

$$\gamma^t(\rho) = e^t \rho, \quad t \in \mathbb{R}. \tag{9}$$

These equations form the basis of the theory developed in the lectures below. The equations are simple. They succinctly express the stability inherent in the limit law in terms of symmetries of the associated infinite measure. The stability allows us to tackle our basic problem of describing the distribution tail on a halfspace that contains few sample points.

Definition. A measure on an open set in $\mathbb{R}^d$ is a *Radon measure* if it is finite for compact subsets. An *excess measure* is a Radon measure ρ on an open set in $\mathbb{R}^d$ that satisfies (9) and gives mass $\rho(J_0) = 1$ to some halfspace J_0.

Excess measures play a central role in this book. They are infinite, but have a simple probabilistic interpretation in terms of point processes. The significance of point processes for extreme value theory has been clear since the appearance of

[4] One-parameter groups of matrices should not frighten a reader who has had some experience with finite state Markov chains in continuous time or with linear differential equations in $\mathbb{R}^d$ of the form $\dot{x} = Ax$.

the book Resnick [1987]. In our more geometric theory the excess measure is the *mean measure* of the Poisson point process which describes the behaviour of the sample cloud, properly normalized, at its edge, as the number of data points tends to infinity. An example should make clear how excess measures may be used to tackle our problem of too few sample points.

Example 3. Suppose γ^t, $t \in \mathbb{R}$, is the group of vertical translations, $\gamma^t\colon (u,v) \mapsto (u, v+t)$ on $\mathbb{R}^{h+1}$. A measure ρ of the form $d\rho(u,v) = d\rho^*(u)e^{-v}dv$ will satisfy (9) for any probability measure ρ^* on $\mathbb{R}^h$. The halfspace $J_0 = H_+$ has mass one and ρ is an excess measure. Conversely one may show that any excess measure with the symmetries γ^t above has the form $d\rho(u,v) = d\rho^*(u)e^{-v}dv$ if one imposes the condition that $\rho(H_+) = 1$. ♢

The probability measure ρ^* in the example above is called the *spectral measure*. The product form of the excess measure in the example makes it possible to estimate the spectral measure even if the upper halfspace contains few points. One simply chooses a larger horizontal halfspace, containing more points. Something similar may be done for any excess measure for exceedances over horizontal thresholds. We shall not go into details here. Suffice it to say that such an excess measure is determined by its symmetry group γ^t, $t \in \mathbb{R}$, and a probability measure ρ^* on the horizontal coordinate plane $\mathbb{R}^h$. The spectral measure ρ^* may be interpreted as the conditional distribution of U given $V = 0$ for the limit vector $W = (U, V)$ on H_+ in (6). The exponential distribution on the vertical axis enters the picture via the Representation Theorem for the limit vector:

$$W = \gamma^T(U^*, 0), \tag{10}$$

where the vector U^* in $\mathbb{R}^h$ has distribution ρ^*, and T is standard exponential, independent of U^*. This decomposition of W reflects the symmetry of the excess measure expressed in (9). It enables one to build probability distributions on halfspaces H far out, and then to estimate probabilities $\mathbb{P}\{Z^H \in E\}$ for EßH, and expectations $\mathbb{E}\varphi(Z^H)$ for loss functions $\varphi\colon H \to [0,\infty)$. Here is the recipe. Assume $\alpha_H^{-1}(Z^H) \Rightarrow W$, where W lives on a halfspace J_0, and has a non-degenerate distribution that extends to an excess measure ρ, and where $H = \alpha_H(J_0)$ are halfspaces such that $\mathbb{P}\{Z \in H\} \to 0$.

Recipe. *Replace Z^H by $\alpha_H(W)$ and compute $\mathbb{P}\{\alpha_H(W) \in E\} = \rho(\alpha_H^{-1}(E))/\rho(J_0)$ and the integral $\mathbb{E}\varphi(\alpha_H(W)) = \int_{J_0} \varphi \circ \alpha_H \, d\rho/\rho(J_0)$ in terms of the excess measure. Given the symmetry group and the normalization α_H one only needs to know the spectral measure ρ^* to compute these quantities. The spectral measure may be estimated from data points lower down in the sample cloud.* ♢

The spectral measure is dispensable. It is the symmetries in (9) that do the job.

These allow us to replace a halfspace containing few observations by a halfspace containing many observations, and on which the distribution has the same shape[5].

Given the recipe, it is clear what one should do to develop the underlying theory: determine the *one-parameter* groups γ^t in $\mathcal{A}(d)$, and for each determine the excess measures (if any) and their domains of attraction. This is done in Section 18.8 for $d = 2$. For linear transformations the program has been executed by Meerschaert and Scheffler in their book Meerschaert & Scheffler [2001] on limit laws for sums of independent random vectors. Let us give a summary of the theory in MS.

We may restrict attention to one-parameter matrix groups by (2). Such one-parameter groups are simple to handle. The group γ^t, $t \in \mathbb{R}$, is determined by its generator C. One may write $\gamma^t = e^{tC}$, where the right hand side is defined by its power series. There is a one-to-one correspondence between matrices C of size d and one-parameter groups of linear transformations on $\mathbb{R}^d$. If one chooses coordinates such that C has *Jordan form*, one may write down the matrices for γ^t by hand for any dimension d. See Section 18.12 for details. Now we have to choose ρ. Let ρ be a Radon measure on an open set Oß$\mathbb{R}^d$ that satisfies $\gamma^t(\rho) = e^t\rho$, $t \in \mathbb{R}$. For an excess measure there still has to be a halfspace J_0 of mass one.

Example 4. *Lebesgue measure* on $\mathbb{R}^d$ satisfies (9) with $\gamma^t = \mathrm{diag}(a_1^t, \ldots, a_d^t)$ for any diagonal matrix with $a_i > 0$, and $a_1 \ldots a_d = 1/e$. However there are no halfspaces of measure one. There are, if one restricts the measure to an orthant, or a *paraboloid.* ◊

If γ^t for $t > 0$ maps the horizontal halfspace J_0 onto a proper subset of itself, then any probability measure ρ^* on $\mathbb{R}^h$ may act as spectral measure for a measure ρ that satisfies (9) and gives mass one to J_0. Similarly, if the γ^t for $t > 0$ are linear expansions, and the image of the open unit ball $B = \{\|w\| < 1\}$ contains the closed unit ball, a probability measure ρ^* on the sphere $\partial B = \{\|w\| = 1\}$ will generate a measure ρ on $\mathbb{R}^d \setminus \{0\}$ that satisfies (9), and gives mass one to the complement of the ball B. In the second case there are many halfspaces of finite mass: $\rho(J)$ is finite for any halfspace J that does not contain the origin. Constructing excess measures is not difficult!

Given the excess measure ρ, the halfspace J_0 of mass one, and the one-parameter group γ^t in (9), we still have to determine the domain of attraction. Recall that Z lies in the *domain of attraction* of W if (6) holds. We write $Z \in \mathcal{D}^h(W)$, and call $\mathcal{D}^h(W)$ or $\mathcal{D}^h(\rho)$ the *domain* of W or ρ *for exceedances over horizontal thresholds*. Let $Z \in \mathcal{D}^h(\rho)$ have distribution π. The basic limit relation (8) for horizontal

[5] Coles and Tawn, in their response to the discussion of their paper Coles & Tawn [1994] write: "Anderson points out that our point process model is simply a mechanism for relating the probabilistic structure within the range of the observed data to regions of greater extremity. This, of course, is true, and is a principle which, in one guise or another, forms the foundation of all extreme value theory."

halfspaces J may be reformulated as

$$e^t \beta(t)^{-1}(\pi) \to \rho \quad \text{weakly on } J,\ \rho(J) < \infty. \tag{11}$$

The $\beta(t)$ belong to the group $\mathcal{A}^h$ of affine transformations mapping horizontal halfspaces into horizontal halfspaces. One may choose $\beta\colon [0,\infty) \to \mathcal{A}^h$ to be continuous, and to *vary like* γ^t:

$$\beta(t_n)^{-1}\beta(t_n + s_n) \to \gamma^s, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{12}$$

In a slightly different terminology this states that β varies regularly with index C, where C is the generator of the symmetry group γ^t. Section 18.1 contains a brief introduction to multivariate regular variation. Regular variation of linear transformations is treated in more detail in Chapter 4 of Meerschaert & Scheffler [2001]. The central result, the Meerschaert Spectral Decomposition Theorem, states that one may restrict attention to one-parameter groups γ^t for which all complex eigenvalues of γ have the same absolute value. See Section 18.4 for details.

Does one really need the theory of multivariate regular variation to handle high risk scenarios?

There are good reasons for using *regular variation* to study multivariate extremes. We list four:

1) The theory is basic. Nothing essential is lost if one assumes $t_n = n$ and $s_n = 1$ in (12). In the final resort, regular variation is about sequences of the form:

$$\beta(n) = \beta(0)\gamma_1 \dots \gamma_n, \quad \gamma_n \to \gamma. \tag{13}$$

One gets back the original curve β, or a curve asymptotic to the original curve, by interpolation. Details are given in Section 18.2.

2) The theory contains a number of deep results that clarify important issues in applications. We give two examples of questions that may be resolved by the Meerschaert *Spectral Decomposition Theorem*, the fundamental result in the multivariate theory of regular variation.

i) Suppose γ^t, $t \in \mathbb{R}$, is a group of *linear* transformations. Is it possible to choose the origin in z-space and normalizations $\bar{\beta}(t) \sim \beta(t)$, mapping w into z, that are linear in the new coordinates?

ii) Suppose the symmetries γ^t, $t \in \mathbb{R}$, in appropriate coordinates in w-space, are diagonal. Is it possible to choose coordinates in z-space and normalizations $\bar{\beta}(t) \in \mathrm{GL}$, asymptotic to $\beta(t)$ for $t \to \infty$, that are diagonal in the new coordinates?

The answer to i) is "Yes" if γ is an expansion, or a contraction ; see Lemmas 15.15 and 16.13 below. This result explains why univariate extreme value theory is so much simpler for heavy tails than for distributions in the domain of the Gumbel law. Univariate linear normalizations are non-zero scalars! The answer to ii) is "Yes" if the diagonal entries in γ are distinct.

3) Regular variation enables us to construct simple continuous densities in the domain of attraction of excess measures with continuous densities. By the transformation theorem the density g of the excess measure in (9) satisfies

$$g(\gamma^t(w)) = g(w)/q^t, \quad q = e|\det A|, \quad \gamma^t(w) = b(t) + A^t w. \tag{14}$$

For $\gamma^t \in \mathcal{A}^h$ one also has the decomposition

$$g(u, v) = g_v(u)\tilde{g}(v), \tag{15}$$

where $\tilde{g}$ is the density of a univariate GPD, see (5), and the conditional densities g_v all have the shape of the density g^* of the spectral measure ρ^*.

Example 5. The *Gauss-exponential* density $e^{-u^T u/2}e^{-v}/(2\pi)^{h/2}$ determines an excess measure ρ on $\mathbb{R}^{h+1}$. Vertical translations $\gamma^t : (u, v) \mapsto (u, v+t)$ are symmetries. The spectral density is standard Gaussian. For any curve $\beta : [0, \infty) \to \mathcal{A}^h$ that varies like γ^t there exists a vector $Z = (X, Y)$ with distribution π, and continuous density

$$f(x, y) = f_y(x)\tilde{f}(y) \tag{16}$$

such that $e^t \beta(t)^{-1}(\pi) \to \rho$ weakly on all horizontal halfspaces. The density f satisfies

$$\frac{f(\beta(t_n)(w_n))}{f(\beta(t_n)(0))} \to e^{-u^T u/2}e^{-v}, \quad t_n \to \infty, \; w_n \to (u, v) \in \mathbb{R}^{h+1}. \tag{17}$$

The density $\tilde{f}$ of Y satisfies the von Mises condition for the univariate Gumbel domain, see Section 6.6; the conditional densities f_y of X given $Y = y$ in (16) are Gaussian. ◊

A continuous density f as above will be called a *typical density* for g^* and β.

4) Multivariate regular variation has a strong geometric component. This is particularly clear if the excess measure is symmetric in a geometric sense. Let $\rho = \rho_\tau$ be a *Euclidean Pareto* measure on $\mathbb{R}^d \setminus \{0\}$. These measures are spherically symmetric with densities $c/\|w\|^{d+\lambda}$, where $\lambda = 1/\tau > 0$ describes the decay rate of the tails, and $c = c_\tau > 0$ may be chosen so that $\rho(J_0) = 1$ for the halfspace $J_0 = \mathbb{R}^h \times [1, \infty)$.

Definition. Bounded convex sets F_n and E_n of positive volume are *asymptotic* if

$$F_n \sim E_n \iff |F_n \cap E_n| \sim |F_n \cup E_n|. \tag{18}$$

Here $|A|$ denotes the volume of the set A.

Exercise 6. The reader is invited to investigate what a sequence of centered ellipses E_n in the plane looks like if $E_{n+1} \sim E_n$, and if the area $|E_n|$ is constant. ◊

Example 7. Start with a sequence of open centered ellipsoids $E_0, E_1, \dots$ such that $E_{n+1} \sim 2E_n$. Also assume E_n contains the closure of E_{n-1} for $n \geq 1$. For $c > 1$ one may use interpolation to construct a *unimodal* function f with elliptic *level sets* such that $\{f > 1/c^n\} = E_n$ for $n \geq 1$, and such that $\{f = 1\}$ is the closure of E_0. For $c > 2^d$ the function f is integrable, say $\int f d\lambda = C < \infty$. Suppose the ellipsoids $E(t) = \{f > 1/c^t\}$ *vary regularly*:

$$E(t_n + s_n) \sim 2^s E(t_n), \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{19}$$

The probability distribution π with density f/C lies in the domain of the *Euclidean Pareto* excess measure ρ_τ, where $c = 2^{d+1/\tau}$. One may choose linear transformations $\beta(t)$, depending continuously on $t \geq 0$, such that $\beta(t)(B) = E(t)$, and such that β varies like $\gamma^t : w \mapsto 2^t w$. Then

$$c^{-t}\beta(t)(\pi) \to \rho_\tau \text{ weakly on } \mathbb{R}^d \setminus \varepsilon B, \quad t \to \infty,\ \varepsilon > 0.$$

Details are given in Section 16. ◇

The reader will notice that densities occupy a central position in our discussions. In the multivariate situation densities are simple to handle. Densities are geometric: sample clouds tend to evoke densities rather than distribution functions. If the underlying distribution has a singular part, this will be reflected in irregularities in the sample cloud. Such irregularities, if they persist towards the boundary, call for a different statistical analysis. In the theory of coordinatewise maxima dfs play an all-important role. Densities have been considered too, see de Haan & Omey [1984] or de Haan & Resnick [1987], but on the whole they have been treated as stepchildren. In our more geometric approach densities are a basic ingredient for understanding asymptotic behaviour. From our point of view the general element in the domain of attraction of an excess measure with a continuous density is a perturbation of a probability distribution with a typical density. From a naive point of view we just zoom in on the part of the sample cloud where the vertical coordinate is maximal, adapting our focus as the number of sample points increases. Under this changing focus the density with which we drape the sample cloud should converge to the density of the limiting excess measure.

Proper normalization is essential for handling asymptotic behaviour and limit laws in probability theory. The geometric approach allows us to ignore numerical details, and concentrate on the main issues. Let us recapitulate: In order to estimate the distribution on a halfspace containing few sample points one needs some form of stability. The stability is formulated in our *Ansatz*: High risk scenarios far out in a given direction have the same shape. If one assumes a limit law, then there is an excess measure. The symmetries of the excess measure make it possible to estimate the distribution on halfspaces far out by our recipe above. The symmetries also impose conditions on the normalizations. These conditions have a simple formulation in terms of regular variation. One may choose the normalizing curve β in (11) to vary like γ^t. Roughly speaking, the group of symmetries γ^t of the excess measure enforces regular variation on the normalizations.

The four arguments above should convince the reader that regular variation is not only a powerful, but also a natural tool for investigating the asymptotic behaviour of distributions in the domain of attraction of excess measures.

In these notes we take an informal approach to regular variation, dictated by its applications to extremes. Attention is focussed on three situations:

1) for coordinatewise extremes the symmetries γ^t and the normalizations α_n are coordinatewise affine transformations (CATs);

2) for exceedances over *horizontal thresholds* the symmetries γ^t and normalizations α_n belong to the group $\mathcal{A}^h$: they map horizontal halfspaces into horizontal halfspaces;

3) for exceedances over *elliptic thresholds* the symmetries γ^t, $t > 0$, are linear expansions, and so are the renormalizations $\alpha_n^{-1}\alpha_{n+1}$.

The theory of coordinatewise extremes is well known, and there exist many good expositions. Our treatment in Chapter II is limited to essentials. Exceedances are treated in Chapter IV. Exceedances over horizontal thresholds describe high risk scenarios associated with a given direction; exceedances over elliptic thresholds may be handled by linear expansions. The theory developed in MS is particularly well suited to exceedances over elliptic thresholds. Arguments for using elliptic thresholds for heavy tailed distributions are given in Section 16.1. The basic limit relation (8) now reads

$$n\alpha_n^{-1}(\pi) \to \rho \ \text{ weakly on } \varepsilon B^c, \quad \varepsilon > 0, \tag{20}$$

where B is the open unit ball, and α_n are linear expansions. If (20) holds we say that π lies in the *domain* of ρ *for exceedances over elliptic thresholds*, and write $\pi \in \mathcal{D}^\infty(\rho)$. Example 7 is exemplary. It treats an excess measure on $\mathbb{R}^d \setminus \{0\}$ with a spectral measure ρ^* which is uniformly distributed over the unit sphere, and a symmetry group of scalar expansions

$$\gamma^t(w) = e^{\tau t} w, \quad t \in \mathbb{R}. \tag{21}$$

If we allow ρ^* to have any distribution on the unit sphere ∂B, but assume the normalizations $\beta(t)$ to be scalar, then the ellipsoids $\beta(t)(B)$ are balls. The limit relation for the high risk scenarios simplifies:

$$Z^r/r \Rightarrow W, \quad r \to \infty, \tag{22}$$

where Z^r is the vector Z conditioned to lie outside the open ball rB. In this situation it is natural to use polar coordinates and write $Z = R\zeta$ with $R = \|Z\|$. The distribution of $(\zeta, R/r)$, conditional on $R \geq r$, converges to a product measure $d\rho^* \times dG$ on

$\partial B \times [1, \infty)$, where ρ^* is the *spectral measure*, and G a Pareto distribution on $[1, \infty)$ with density $\lambda / r^{\lambda+1}$, $\lambda = 1/\tau$. The spectral measure gives an idea of the directions in which the data extremes cluster; the parameter τ in (21) describes the decay rate of the tails.

Here we have another example of the close relation between symmetry and independence! In this model it is again obvious how to estimate the distribution of the high risk scenarios Z^r for values of r so large that only one or two sample points fall in the complement of the ball rB.

Asymptotic *independence* is *not* the subject of these lectures. Our theory is based on concepts like scale invariance, self-similarity and symmetry. It is geometric and local. Independence is a global analytic assumption. It allows one to draw far-reaching conclusions about extremes, but the techniques are different from those developed here.

So far we have assumed convergence of a one-parameter family of high risk scenarios indexed by horizontal halfspaces H^y, $y \uparrow y_\infty$, or by an increasing family of ellipsoids $E_t = \alpha_t(B)$. These situations yield a limit measure with a one-parameter family of symmetries, the excess measure described in (9). Let us now return to high risk scenarios Z^H where the halfspaces H are allowed to diverge in any direction. For simplicity assume Z has a density. Assume convergence of the normalized high risk scenarios to a non-degenerate vector W on H_+: For each halfspace H of positive mass there exists an affine transformation α_H mapping H_+ onto H such that

$$\alpha_H^{-1}(Z^H) \Rightarrow W, \quad \mathbb{P}\{Z \in H\} \to 0. \tag{23}$$

The limit describes the tail asymptotics in every direction. In Section 13 we shall exhibit a continuous one-dimensional family of excess measures ρ_τ, $\tau \geq -2/h$, $h = d - 1$, corresponding to the multivariate GPDs. The densities, standardized to have a simple form, and without norming constant, are

$$\begin{aligned}
g_0(u, v) &= e^{-(v + u^T u/2)}, && w = (u, v) \in \mathbb{R}^{h+1}, \\
& && \tau = 0,\ J_0 = \{v \geq 0\}; \\
g_\tau(w) &= 1/\|w\|^{d+\lambda}, && w \neq 0, \\
& && \tau = 1/\lambda > 0,\ J_0 = \{v \geq 1\}; \\
g_\tau(u, v) &= (-v - u^T u/2)_+^{\lambda-1}, && v < -u^T u/2, \\
& && \tau = -1/(h/2 + \lambda) < 0,\ J_0 = \{v \geq -1\}.
\end{aligned}$$

The reader may recognize the Gauss-exponential and the Euclidean Pareto excess measure from the examples above. In all cases the vertical coordinate of the high risk limit distribution determined by the restriction of ρ_τ to J_0 has a univariate GPD, with Pareto parameter τ, see (5). For $\tau < 0, \lambda = 1$, the excess measure ρ_τ is Lebesgue measure on the paraboloid $\{v < -u^T u/2\}$. For $\tau = -2/h$ the excess measure is singular. The symmetry of the excess measures ρ_τ, $\tau \geq -2/h$, is impressive. Instead of the one-parameter group γ^t, $t \in \mathbb{R}$, in (9) there now is a symmetry group of dimension $2, 4, 7, \ldots$ for $d = 2, 3, 4, \ldots$. Many halfspaces have finite mass. The probability distributions $d\rho^J = 1_J d\rho/\rho(J)$ associated with such halfspaces all have the same shape. The measure ρ has the *tail property* to an excessive degree. The domains of attraction, $\mathcal{D}(\tau)$, of the measures ρ_τ are investigated in Chapter III.

Before giving a detailed description of the contents of the various chapters we still want to consider two topics: the relation to the multivariate theory of coordinatewise maxima, and the range where the theory will apply.

How do coordinatewise maxima fit in?

The subject of this book may be described as *geometric extreme value theory* since we are looking at the behaviour of the extreme points of sample clouds as the number of data points increases without bound. We are concerned with the convex hull, but also with the points of the cloud below the surface. Since we are zooming in at the scale of individual sample points, the limit, if we assume convergence, has to be a Poisson point process whose mean measure is finite and positive on some halfspace. Such limits were first considered by Eddy [1980].

The geometric approach and the analytic, coordinatewise approach are complementary. The geometric theory is interested in linear combinations of coordinates, the analytic theory in maxima of coordinates. There is a difference in interpretation. In the geometric theory for exceedances over horizontal or elliptic thresholds there is one variate (the vertical, or the radial) that measures risk, and an h-dimensional ancillary vector; in the analytic theory all d coordinates play an equal role. The geometric approach looks at exceedances, the analytic approach at maxima. In the univariate situation the two theories are equivalent. In higher dimensions the relation between extremes and exceedances is most clearly seen in the behaviour of sample clouds, and the limiting Poisson point process. In the geometric approach the mean measure of the limit process is called an excess measure, in the analytic approach, it is called an exponent measure; but actually these two terms denote the same[6] object.

One might say that the theory of coordinatewise maxima is concerned with high risk scenarios on sets that are not halfspaces, but complements of shifted negative orthants. Instead of divergent sequences of halfspaces H_n one considers sets of the form

$$[-\infty, \infty)^d \setminus [-\infty, a_n), \quad a_n \in \mathbb{R}^d, \tag{24}$$

where the points a_n increase towards the upper endpoint of the df. Since the complement of a shifted negative orthant contains many halfspaces, convergence of the coordinatewise maxima implies convergence of high risk scenarios on many half-

[6] Exponent measures may give mass to hyperplanes at $-\infty$; excess measures live on open subsets of $\mathbb{R}^d$. The differences will be discussed more fully at various points in these lectures.

spaces. Below we formulate a result that expresses these ideas. One has to distinguish between heavy and light tails. The upper tail of a df F on $\mathbb{R}$ is *light* if

$$t^m(1 - F(t)) \to 0, \quad t \to \infty,\ m = 1, 2, \ldots. \tag{25}$$

It is *heavy* if there exists an integer $m \ge 1$ such that $t^m(1 - F(t)) \to \infty$ for $t \to \infty$.

If all components have heavy upper tails the relation between coordinate maxima and exceedances is simple. One assumes that Z has non-negative components. The exponent measure lives on $[0, \infty)^d \setminus \{0\}$. It is an excess measure, whose symmetries γ^t are linear diagonal expansions for $t > 0$. The max-stable limit $G = \lim F^n \circ \alpha_n$ has the form $G = e^{-R}$, where R is the distribution function of the excess measure. So the df G determines the mean measure, and hence the distribution, of the Poisson point process describing the asymptotic behaviour of the sample clouds. The normalizations α_n are diagonal matrices. Weak convergence $F^n \circ \alpha_n \to G$ implies weak convergence $n\alpha_n^{-1}(dF) \to dR$ on the complement of any ε-ball εB centered in the origin.

Vectors whose components have light upper tails have exponent measures that may charge planes and lines at $-\infty$. The normalizations are *CAT*s, *coordinate affine transformations*, $z_i = a_i w_i + b_i$, $i = 1, \ldots, d$. Let us show how coordinatewise extremes for light tails fit in.

Proposition 8. *Let Z have df F with marginals F_i having light upper tails. Suppose Z lies in the domain of attraction $\mathcal{D}^{\vee}(\rho)$ for coordinatewise maxima: There exist CATs α_n such that*

$$\begin{aligned} &F^n(\alpha_n(w)) \to G(w) = e^{-R(w)} \text{ weakly,} \\ &R(w) = \rho([-\infty, \infty)^d \setminus [-\infty, w)), \quad w \in \mathbb{R}^d. \end{aligned} \tag{26}$$

Choose $q \in \mathbb{R}^d$ such that $H_i(q_i) < 1$ for $i = 1, \ldots, d$, and set $a_n = \alpha_n(q)$ and $Q = [-\infty, \infty)^d \setminus [-\infty, q)$. Then $\alpha_n^{-1}(Z^{(-\infty, a_n)^c}) \Rightarrow W$, where W has distribution $1_Q d\rho/\rho(Q)$. Let $J = \{\xi \ge c\}$ß$\mathbb{R}^d$ with $\xi \in (0, \infty)^d$. If $\mathbb{P}\{W \in \mathbb{R}^d\}$ is positive, then

$$\alpha_n^{-1}(Z^{H_n}) \Rightarrow W^J, \quad H_n = \alpha_n(J), \tag{27}$$

where W^J has distribution $1_J d\rho/\rho(J)$.

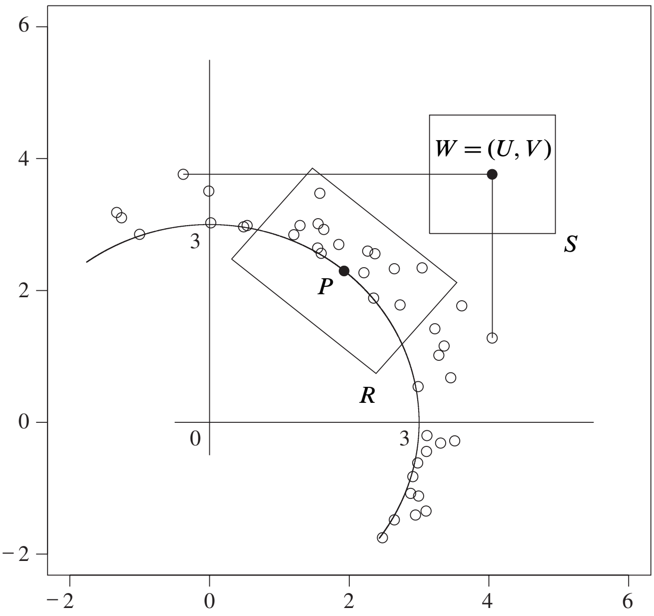
Proof. Relation (26) is standard; see Theorem 7.3. In the limit relation (27) the crucial point is that the condition $\mathbb{P}\{W \in \mathbb{R}^d\} > 0$ ensures that $\rho(J)$ is positive. This implies

$$\mathbb{P}\{Z \in H_n\}/\mathbb{P}\{Z \notin (-\infty, a_n)\} \to \rho(J)/\rho(Q).$$

For JßQ the result follows by a simple conditioning argument. The general case follows by the symmetry of ρ. □

What happens if $\mathbb{P}\{W \in \mathbb{R}^d\} = 0$ in the proposition above?

The figure below suggests that a more geometric approach which zooms in on a boundary point of the sample rather than on the max-*vertex* may be useful in certain situations.



The rectangle R around the point P contains more information about the edge of this 10 000 point sample from the normal distribution than the square S around the max-vertex $W = (U, V)$.

The standard normal distribution on the plane lies in the domain $\mathcal{D}^{\vee}(\rho)$ of the max-stable df $\exp -(e^{-u} + e^{-v})$ (independent *Gumbel* marginals). In order to describe the coordinatewise maximum, the sample cloud is enclosed in a coordinate rectangle. The coordinatewise maximum is the upper right hand corner of the rectangle, the max-vertex. Now the scaling is crucial. For a heavy tailed distribution, a *spherical Student* distribution for instance, the scaling preserves the origin, which remains in the picture. For the light tailed *Gaussian distributions*, however, the normalization zooms in on a small (empty) square around the max-vertex. It fails even to see the shape of the sample cloud. As a result all bivariate Gaussian densities with standard normal marginals yield the same bivariate extreme asymptotics.

Let us now say a few words on the applicability of the theory presented in this book.

Our approach to risk is that of an observer, rather than a risk manager. Given a multivariate data set describing the past behaviour, and a loss function, our aim is to describe the tail behaviour of the distribution underlying the data set. Such a description enables one to construct large synthetic samples, and to study the behaviour of the associated random losses. This procedure is known as *stress testing* and *scenario analysis*. We are not concerned with the problem of changing the parameters of the underlying distribution, redirecting the dynamics which produced the data set, or altering the loss function by a suitable form of risk transfer. These issues are treated in McNeil, Frey & Embrechts [2005] for financial risk; and for risk in the realm of reliability engineering in Bedford & Cooke [2001].

Einstein showed that the erratic movement of pollen grains suspended in a drop of water, as observed by Brown at the beginning of the 19th century, could be described by smooth probability distributions exhibiting a large degree of symmetry. Complex dynamical systems may give rise to symmetric probability distributions. Symmetries in a data set may reflect regularities inherent in the dynamical system which produces the data. If so, the symmetries are likely to persist. The validity of our model depends on this persistence of the symmetry. For Brownian motion as a model for the movement of pollen grains, Einstein [1906] imposed a bound of 10^{-7} seconds for applicability. So too, in financial or meteorological or biological applications the symmetry will break down at a certain level[7].

To fix ideas let us posit an ultimate probability p_0 in the range 10^{-99} to 10^{-20}. Halfspaces with probabilities below this value have no reality for risk management. Replacing the conditional distribution on such a halfspace by any other distribution does not influence the policy of the risk manager. This means that the endless variety of the ever slower pirouettes performed by the sequence of ellipses E_n in Exercise 6 above, forms part of the mathematical theory, but has no bearing on risk management, since the probabilities $\mathbb{P}\{Z \in E_n^c\}$ lie below the threshold value p_0 after a few hundred terms. By the same argument the existence of moments falls outside the range of realistic risk theory. (Convergence of integrals is determined by the behaviour of the distribution on invisible halfspaces.) This collateral result is not as disturbing as it may seem on first sight. For heavy tails the value of the exponent where the moment first fails to exist, is a convenient measure of risk, but for a realist the difference between a Gaussian distribution and a *Cauchy distribution* is established by the behaviour of samples of size a hundred. She is not interested in the tail behaviour at risk levels 10^{-99}. The assumption of an ultimate probability p_0 also has advantages. It provides

[7]A finite universe does not preclude a model with infinite upper tail for spatial variables. String theory tells us that four-dimensional space-time may break down at magnitudes of 10^{-30} meters to reveal six hidden dimensions curled up into compact sets. This does not mean that models with continuous densities for spatial or temporal quantities are invalid.

us with the liberty to choose the behaviour of the distribution on invisible halfspaces to suit our fancy. We find it convenient to assume convergence of *high risk scenarios*.

Having said this much on applicability, we may now proceed with our proper task, the mathematical investigation of the consequences of the assumption that high risk scenarios converge.

Contents

The book consists of twenty lectures, grouped into five chapters. There is a basics chapter on point processes, a final chapter listing open problems, and in between there are three chapters covering three different topics: coordinatewise extremes, multivariate GPDs, and exceedances over thresholds (horizontal and elliptic).

Chapter II treats the basic univariate extreme value theory, and provides an overview of the theory of coordinatewise maxima. Our focus is on exponent measures rather than max-stable dfs. The Chapters III and IV form the body of the book. They present two different views on *high risk scenarios*. In Chapter III the high risk scenarios Z^H converge to a common limit law, in whatever direction the halfspaces H diverge. This restricts the class of limit laws. We present a one-parameter family of limit laws, the multivariate GPDs. It is not known whether other limit laws exist. Chapter III is a relatively self-contained account of what is known about the domains of attraction of the multivariate GPDs. The elegant theory of multivariate GPDs, and their domains of attraction, should be useful in situations where the sample cloud is bland or where the dimension is high, and where one is interested in the overall extremal behaviour rather than the asymptotic behaviour in a particular direction. Such an approach may be of interest to the supervisor or regulator; it allows a diversified view of the extremal behaviour of widely varying positions in the underlying market. The theory presented in Chapter IV is different. In this chapter we look at exceedances. For linear thresholds this means that we look at halfspaces moving off in a given direction. Such a model is of interest to the trader or risk manager taking directional positions in the underlying market. For simplicity we assume the thresholds horizontal. For heavy tails, elliptic thresholds are more natural since there is no difference between the local and the global theory. This is explained in Section 16.1. Heavy tailed vectors are normalized by linear contractions. The theory of exceedances presented in Chapter IV has the same structure as the theory of coordinatewise maxima. The limit laws are known. The excess measure, like the exponent measure, satisfies a one-parameter group of symmetries. The normalizations are more complex than the CATs used for coordinatewise extremes, and call for a geometric approach. The asymptotics may be handled by regular variation. A complete characterization of the domains of attraction is available for a number of limit laws. It is presented in Sections 15.2 and 16.7.

These notes offer probability theory rather than statistics. If one accepts the idea that excess measures may occur as the mean measure of a limiting Poisson point process describing the asymptotic behaviour of sample clouds at their edge, then good estimates of the excess measure and the normalizations allow one to simulate large samples that may then be used in risk analysis. The task of the probabilist is to analyse the model. What do excess measures look like? For any given excess measure, what does the domain of attraction look like? What normalizations are allowed? What moments will converge? What does the convex hull of the sample cloud look like? Does it converge? These are some of the questions that will be addressed in the present text.

Chapter I treats point processes. The first four sections are standard theory: An intuitive introduction; the Poisson point process as a limit of superpositions of sparse point processes; the distribution of point processes; and their convergence. In Section 5 extremes enter the scene. We consider the n-point sample cloud N_n from the probability distribution $\pi_n = \alpha_n^{-1}(\pi)$ on $\mathbb{R}^d$, and assume vague convergence $N_n \Rightarrow N_0$ to a Poisson point process N_0. In applications the mean measure ρ of the limit process is an infinite Radon measure on an open set O, for instance $\mathbb{R}^d$ or $\mathbb{R}^d \setminus \{0\}$. We are interested in convergence of the convex hulls of the sample clouds. That means that we have to determine the halfspaces HßO on which ρ is finite, and on which the mean measures $n\pi_n$ converge weakly to ρ. The class of such halfspaces determines two open cones in the dual space, the *intrusion cone* Δ and the *convergence cone* Γ. Modulo some minor regularity conditions convex hulls converge if Δ is non-empty, and $\Gamma = \Delta$. We also discuss loss functions, and approximate their integrals by sample sums.

Chapter II treats the theory of maxima. It consists of two sections. Section 6 treats the univariate situation. The domains of attraction for exceedances and maxima coincide. The domain of attraction $\mathcal{D}^+(0)$ of the exponential law is described in terms of densities which satisfy a von Mises condition. The section also contains an elementary proof of Bloom's basic theorem on self-neglecting functions. The second part, Section 7, assumes some acquaintance with the theory of coordinatewise extremes. We concentrate on the domain of max-stable distributions with standard exponential marginals on $(-\infty, 0)$. This allows us to treat exponent measures that charge coordinate planes in $-\infty$. The sample copula yields a simple tool for investigating the dependency structure. Non-linear normalization of the coordinates provides a direct link to copula theory.

Chapter III starts with an extensive introductory section treating applications, examples, and the general asymptotics of high risk scenarios. For coordinatewise maxima, powers of distribution functions play an important role; in the theory of high risk scenarios one encounters powers of densities. Unimodal densities (with convex level sets) seem to reflect quite well the shape of the sample clouds to which the theory applies. Pointwise convergence of densities often is a first step towards the derivation of limit theorems for distributions. We establish simple asymptotic expressions for excess probabilities, $\mathbb{P}\{Z \in H\}$, in terms of densities.

There is a continuous one-parameter family of multivariate GPDs, indexed by a shape parameter τ. As in the univariate case the family falls apart in three power families, corresponding to the sign of the parameter τ. Excess measures of the heavy tailed distributions, corresponding to $\tau > 0$, have a spherically symmetric density. Tails of distributions in $\mathcal{D}(\tau)$ for $\tau > 0$ may be approximated by tails of elliptic Student distributions. Distributions in $\mathcal{D}(\tau)$ for $\tau < 0$ have bounded support; the convex hull of the support is egg-shaped. The latter distributions receive only cursory treatment; their role in risk theory is limited.

Special attention is given to $\mathcal{D}(0)$, the domain of the Gauss-exponential law. As in the univariate setting, the parameter value $\tau = 0$ is the most interesting mathematically. Section 9 introduces the class $\mathcal{RE}$ of *rotund-exponential* densities. These have the form

$$f(z) \propto e^{-\psi \circ n_D(z)}$$

where the function $e^{-\psi}$ satisfies the von Mises condition for the univariate domain of attraction $\mathcal{D}^+(0)$. The function n_D is the gauge function of a *rotund* set D. Such a set is egg-shaped: convex, open, and bounded, it contains the origin, and the boundary is C^2 with continuously varying positive definite curvature. One may think of the gauge function as a norm, generated by the set D, when $-D = D$. The rotund-exponential densities extend the class of spherical *Weibull densities* $ce^{-\|z\|^r}$, $r > 0$. They allow us to treat sample clouds whose central part is egg-shaped rather than elliptic. Their simple structure should make them tractable for statistical analysis. The normalizations α_H may be written down explicitly in terms of ψ and D. In Section 9 we prove pointwise convergence as H diverges:

$$f(\alpha_H(w))/f(\alpha_H(0)) \to e^{-(v+u^Tu/2)}, \quad w = (u, v) \in \mathbb{R}^{h+1}; \tag{28}$$

in Section 10 we prove $\mathbf{L}^1$-convergence of these quotients for unimodal densities. Section 11 introduces flat functions. Flat functions play the same role in the multivariate theory as slowly varying functions do in the univariate theory. Finally it will be shown that the normalizations induce a Riemannian metric on the convex open set $O = \{f > 0\}$. Conversely, the metric determines the normalizations, and hence the global structure of distributions in the domain $\mathcal{D}(0)$.

We mention two results from Section 13 that should give an impression of the scope and of the limitations of the theory of multivariate GPDs.

• Let A be a linear map from $\mathbb{R}^d$ onto $\mathbb{R}^m$. If the vector $Z \in \mathbb{R}^d$ lies in $\mathcal{D}(\tau)$, then so does $A(Z)$.

• A vector $Z \in \mathcal{D}(\tau)$ with independent components is Gaussian (and $\tau = 0$).

Chapter IV treats exceedances over horizontal and elliptic thresholds. The first two sections treat horizontal thresholds. The first is theoretical. We prove the Extension Theorem: If the high risk scenarios Z^H for horizontal halfspaces H, properly normalized, converge in distribution to a non-degenerate limit vector W, then there is an excess measure ρ, and the normalized sample clouds converge in distribution to a Poisson point process N_0 with mean measure ρ weakly on all horizontal halfspaces on which ρ is finite. This is the step from (6) to (11) and (12). Next we determine the limit laws and excess measures for exceedances over horizontal thresholds. Up to a non-essential multiplicative constant, the excess measure ρ is determined by its symmetry group γ^t, $t \in \mathbb{R}$, and a probability measure ρ_0^* on $\mathbb{R}^h$, the *spectral probability measure*, which is the conditional distribution of U given $V = 0$, where $W = (U, V)$ is the high risk limit vector on H_+. Section 14.9 describes the situation in $\mathbb{R}^3$.

A question that is important for applications is: To what extent may one relax the condition that the halfspaces be horizontal? The excess measure is finite for a horizontal halfspace J_0 by definition. Is it also finite for non-horizontal halfspaces close to J_0? Does weak convergence $n\alpha_n^{-1}(\pi) \to \rho$ in (8) hold on such halfspaces? A related question is whether the convex hull of the normalized sample cloud converges to the convex hull of the limiting Poisson point process. The book gives partial answers to these questions.

A considerable part of Chapter IV is taken up by the analysis of specific examples, and a discussion of the relation to the limit theory for coordinatewise extremes. Section 15 investigates the excess measures and domains of attraction for three simple symmetry groups γ^t: vertical translations, scalar contractions, and scalar expansions. For vertical translations a complete description of the domains of attraction is given.

The next two sections of Chapter IV treat heavy tailed distributions normalized by linear contractions. Here we shall work with elliptic thresholds. Section 16 presents the basic theory. The introduction to this section gives more information. The main result is a complete characterization of the domain of attraction $\mathcal{D}^\infty(\rho)$ for excess measures with a continuous positive density. Section 17 contains examples, and a more detailed analysis of the domain of attraction in the case of scalar symmetries. We also give a careful analysis of the relation between limit laws for exceedances over elliptic thresholds and multivariate regular variation. For very heavy tails sample sums are determined asymptotically by the extremes, and domains of attraction for *operator stable* distributions and for excess measures coincide: $\mathcal{D}^{OS}(\rho) = \mathcal{D}^\infty(\rho)$. In this situation excess measures may be interpreted as Lévy measures for multivariate stable processes without a Gaussian component. The theory for exceedances over elliptic thresholds coincides with the limit theory for sums of independent vectors developed in MS. The theory for exceedances may thus provide a simple introduction to the limit theory for operator stable distributions.

The theoretical results on regular variation in groups and regularly varying multivariate probability distributions developed by Meerschaert and Scheffler in their monograph MS are essential for a deeper understanding of the domain $\mathcal{D}^\infty(\rho)$. The Spectral Decomposition Theorem (*SDT*) clears up the mysterious disparity between the domains of attraction of excess measures with scalar symmetries and those with diagonal non-scalar symmetries. Section 18 contains a brief introduction to the theory of multivariate regular variation, and to the SDT. The second half of this section treats the general theory of excess measures on $\mathbb{R}^d$. Finally Section 18.14 presents an example that shows that the three approaches to the asymptotics of multivariate sample extremes developed in these notes – coordinatewise maxima, exceedances over linear thresholds, and exceedances over elliptic thresholds – may yield conflicting results.

Chapter V lists some fifty open problems. Together with the hundred examples scattered throughout the text these serve to enliven the presentation, and to mark the boundary of our present knowledge. The second part of the chapter describes some of the difficulties that a statistician may encounter if she decides to apply the theory to concrete data sets.

We have provided this lengthy introduction because the book does not have a clear linear structure. It is a collection of essays. Moreover, there is a certain ambiguity in the subject matter. Basically the book is about high risk scenarios. Chapter III may be read from this point of view without bothering about point processes. The reader will then observe that each of the limit laws extends naturally to an infinite measure, and he will observe that this excess measure has an extraordinary degree of symmetry. The excess measure is infinite, but has a simple probabilistic interpretation: The normalizations that are used to obtain a non-degenerate limit law for the *high risk scenarios* may be applied to the sample clouds to yield a limiting Poisson point process. The excess measure is the mean measure of this Poisson point process. This convergence of sample clouds is the second point of view. From this point of view it is natural to start with an overview of point processes, Sections 1–5. The point process approach unifies the univariate theory of extremes and exceedances in Section 6. Section 7 treats the limit theory for coordinatewise maxima under linear and non-linear coordinatewise normalizations from the same point of view. A natural counterpart to the limit theory for high risk scenarios developed in Chapter III is the limit theory for exceedances over thresholds developed in Chapter IV. The two sections on horizontal thresholds are only loosely connected, as are the next two sections on exceedances over elliptic thresholds. In fact it might be more instructive to start with one of the examples in Section 15 or in Section 17 in order to gain an impression of this part of the theory, rather than working through the technicalities leading up to the Extension Theorem in Section 14. Similarly the open problems in Chapter V should give a good impression of the scope of geometric extreme value theory, as treated in this monograph. Sections 5 and 18 have a special standing. They contain background material. Section 5 looks into the question: How does one describe convergence of sample clouds to a limiting Poisson point process in terms of halfspaces? Section 18 treats multivariate regular variation and the general theory of excess measures and their symmetries. It contains subsections on the Meerschaert Spectral Decomposition Theorem, on Lie groups, and on the Jordan form of a matrix.

The book treats only a part of extreme value theory. For extremes of stationary processes, of Gaussian fields, or of time series, the reader may consult Berman [1992], Davis & Resnick [1986], Dieker [2006], Finkenstädt & Rootzén [2004] or Leadbetter, Lindgren & Rootzén [1983], and the references cited therein. For extremes in Markov sequences see Perfekt [1997]; for exceedances see Smith, Tawn & Coles [1997]. Extremes in function spaces and for stochastic processes are treated in Giné, Hahn & Vatan [1990], de Haan & Lin [2003] and Hult & Lindskog [2005]. Limit behaviour of convex hulls has been investigated in Eddy & Gale [1981], Groeneboom [1988], Brozius & de Haan [1987], Baryshnikov [2000], Bräker, Hsing & Bingham [1998], and Finch & Hueter [2004]. Statistics for coordinatewise extremes are treated in de Haan & Ferreira [2006]; for heavy tails see Resnick [2006].

Interest in exceedances over linear thresholds is not new. We mention early papers by de Haan [1985], de Haan & de Ronde [1998], and Coles & Tawn [1994]. The last two contain nice applications to meteorological data. Exceedances over linear thresholds seem to fit snugly within the framework of coordinatewise extremes, as is shown by Proposition 8. It is only by taking a geometric point of view that one becomes aware of the limitations imposed by the coordinatewise approach, due to the restriction to CATs in the normalization. The strong emphasis on coordinates in multivariate extreme value theory so far may also explain why the relevance of the theory of multivariate regular variation developed in MS has not been realized before.

Notation

Halfspaces are closed, and denoted by $H = \{\zeta \geq c\}$ or J. Horizontal halfspaces have the form $\{y \geq c\}$, or $\{\eta \geq y\}$, where η is the vertical coordinate. We often use the decomposition $z = (x, y) \in \mathbb{R}^{h+1}$ into a *vertical component* $y \in \mathbb{R}$ and a horizontal part $x \in \mathbb{R}^h$. So $h = d - 1$ is the dimension of the horizontal coordinate plane $\{y = 0\}$.

The set of affine transformations $\alpha\colon w \mapsto Aw + b$ on $\mathbb{R}^d$ is denoted by $\mathcal{A} = \mathcal{A}(d)$. If the linear part is a diagonal matrix with positive entries we call α a *CAT* (coordinatewise affine transformation). CATs are simple to handle, and they are the transformations used in coordinatewise extreme value theory. The CATs form a closed subgroup of $\mathcal{A}$. So do the translations $w \mapsto w + b$ and the set $\mathcal{A}^h$ of affine transformations that map horizontal halfspaces into horizontal halfspaces. For the closed subgroups of linear transformations, and the compact subgroups of orthogonal and special orthogonal transformations (with determinant one) we use the standard notation GL(d), O(d), and SO(d). We write GL, O and SO if the dimension is not specified.

In general π denotes the distribution of a vector $Z = (X, Y)$ in $\mathbb{R}^{h+1}$. We assume that Z lies in the domain of attraction of a limit vector $W = (U, V)$. This means that

$W_n = \alpha_n^{-1}(Z^{H_n}) \Rightarrow W$ for certain sequences of halfspaces H_n where Z^H denotes the vector Z conditioned to lie in the halfspace H. We regard α_n as transformations from (u, v)-space to (x, y)-space, and hence use the inverse α_n^{-1} to normalize. This corresponds to the usual practice in the univariate case where one subtracts a location parameter and divides by a scale parameter.

The table below lists the domains of attraction introduced in the text:

$\mathcal{D}^+(\tau),\ \tau \in \mathbb{R}$	(6)	domain of the univariate GPD G_τ
$\mathcal{D}(\tau),\ \tau \geq -1/2h$	(8)	. . . of the multivariate GPD π_τ for high risk scenarios
$\mathcal{D}^\vee(\rho)\quad \mathcal{D}^\wedge(W)$	(7)	. . . for coordinatewise maxima and minima, normed by CATs
$\mathcal{D}^\uparrow(\rho)\quad \mathcal{D}^\uparrow(C)$	(6,7)	. . . . . . normed by monotone transformations
$\mathcal{D}^h(\rho)$	(14)	. . . for exceedances over *horizontal thresholds*
$\mathcal{D}^\infty(\rho)$	(16)	. . . for exceedances over *elliptic thresholds*
$\mathcal{D}^{OS}(\rho)$	(17)	domain of operator stable vectors with Lévy measure ρ

Domains of attraction (in brackets the section in which they are introduced).

The argument of $\mathcal{D}$ in the table above is the *Pareto parameter* τ, or the excess measure, exponent measure, or Lévy measure ρ, or the limit vector W, or the max-stable copula C. One could add a number of extra parameters, the dimension d, the generator of the symmetry group, restrictions on the normalizations in the form of a subgroup (CATs, linear maps, diagonal maps, scalar maps, translations, etc.); one could specify that convex hulls converge, that densities converge, or that densities be unimodal. Since the theory is still in flux we restrict the notation to essentials.

We mention three possible sources of confusion:

1) The high risk limit vector W lives on a halfspace J_0. In the limit relation $\alpha_H^{-1}(Z^H) \Rightarrow W$ it is assumed that $\mathbb{P}\{Z \in H\}$ is positive – in order to have well-defined high risk scenarios –, that $\mathbb{P}\{Z \in H\} \to 0$ – in order to have an interesting limit relation –, and that $\alpha_H(J_0) = H$ – in order to ensure that the normalized high risk scenarios $W_H = \alpha_H^{-1}(Z^H)$ live on J_0. Often $J_0 = H_+$, the upper halfspace, but it may sometimes be convenient to choose some other halfspace, for instance $J_0 = \{v \geq j_0\}$ with $j_0 = 1$ or $j_0 = -1$, or to leave the precise form of J_0 unspecified. This confusion already exists in the univariate case where Pareto distributions may be standardized to live on $[0, \infty)$ or on $[1, \infty)$.

2) The *spectral measure* is a finite measure, which together with the one-parameter group of symmetries γ^t, $t \in \mathbb{R}$, determines the excess measure. One may take it to be a probability measure by dividing ρ by a harmless positive constant. The spectral measure lives on $\mathbb{R}^h$ or on the unit sphere. It has the advantage over the excess measure that it is arbitrary. The *spectral measure* bears no relation with the *spectral decomposition*. The latter concerns the symmetries and the normalizations.

3) Exceedances over *elliptic thresholds* is an alternative to exceedances over linear thresholds which is particularly well suited to distributions with heavy tails. Actually, as explained in Section 16.1, we shall hardly consider *high risk scenarios* of the form Z^{E^c}. The limit relation $\alpha_E^{-1}(Z^{E^c}) \Rightarrow W$ only occurs in Section 17.7. The really interesting relation is

$$e^t\beta(t)^{-1}(\pi) \to \rho \text{ weakly on } \varepsilon B^c, \quad t \to \infty,\ \varepsilon > 0.$$

In the terminology of MS this just says that the probability measure π *varies regularly* with exponent C, where the one-parameter linear expansion group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, satisfies $\gamma^t(\rho) = e^t\rho$.

A function $f \geq 0$ on $\mathbb{R}^d$ is called *unimodal* if the level sets $\{f > c\}$ are convex for $c > 0$.

Limits are often one-sided. If y_∞ is the upper endpoint of a distribution on $\mathbb{R}$ then $y \to y_\infty$ always means convergence from below. In limits for sequences indexed by n we implicitly assume $n \to \infty$.

B is the open centered Euclidean unit ball, E_n are open ellipsoids.

N_n and N are point processes, usually on $\mathbb{R}^d$, or on an open set Oß$\mathbb{R}^d$.

$d = h+1$ is the dimension of the vectors $Z = (X, Y)$ and $W = (U, V)$ in $\mathbb{R}^{h+1}$.

τ is the Pareto (shape) parameter, ρ the excess measure, π a probability distribution. The vertical component of the measure ρ on $\mathbb{R}^{h+1}$ is $\tilde{\rho}$, and $\tilde{\alpha}$ for $\alpha \in \mathcal{A}^h$ is the univariate affine transformation of the vertical coordinate induced by α.

The abbreviations rv, df and iid are standard in probability theory and statistics.

The relation $\stackrel{d}{=}$ denotes equality in distribution; $\Rightarrow$ denotes convergence in distribution, $p := \mathbb{P}\{X_n \geq c\}$ defines p as the probability of a certain event. We use the notation $a_n \ll b_n$ or $a_n = o(b_n)$ to signify that $a_n/b_n \to 0$; $a_n \sim b_n$ means that $a_n/b_n \to 1$ and $a_n \asymp b_n$ means that the quotients a_n/b_n and b_n/a_n are bounded eventually. We use $\text{int}(E)$ and $\text{cl}(E)$ to denote the interior and the closure of a set E, and $c(E)$ to denote the *convex hull*.

The basic terminology and notation have been introduced in the Preview. Special notation could not be avoided completely. One may always consult the Index at the end of the book. In the text itself the index entry is in bold face or emphasized by printing it in italics to contrast with the surrounding text. In the case of multiple entries, the bold face page number in the Index will guide the reader to the formal definition. In the Bibliography the numbers in square brackets refer to the pages on which the item is cited.

The table of contents is detailed. Starred sections may be skipped. They are technical or treat subjects which are not used in the remainder of the text. There are a number of sections treating specialized subjects: Sections 2.6 and 17.6 treat Lévy processes and convergence to operator stable processes: Section 6.7 discusses self-neglecting functions; Section 5.3 treats halfspaces and convex sets; Section 18.1 gives a brief introduction to multivariate regular variation; Sections 16.9 and 18.4 discuss the Spectral Decomposition Theorem; Section 18.8 describes the excess measures on the plane; Section 18.9 treats orbits of one-parameter groups of affine transformations on $\mathbb{R}^d$; Section 18.13 treats Lie groups, and Section 18.12 treats the Jordan form, and the spectral decomposition of one-parameter groups.

EKM and MS denote two books which will be cited frequently. Embrechts, Klüppelberg & Mikosch [1997] contains the fundamental material on which this monograph is built, and is an excellent guide to applications in finance and risk theory. Meerschaert & Scheffler [2001] contains an in-depth exposition of the analytic theory of multivariate regular variation for linear transformations, functions, and measures.

I Point Processes

In the geometric theory of multivariate extremes, sample clouds and limiting Poisson point processes play a prominent role. This chapter provides a basic introduction to the theory of point processes. We start with a constructive description of a point process. Then we turn to Poisson point processes as limits for superpositions of a large number of independent sparse point processes. Essential concepts such as the distribution of a point process, and convergence in distribution in the vague and weak topology are treated in Sections 3 and 4. In the last section the theory is applied to a sequence of sample clouds in $\mathbb{R}^d$ converging to a Poisson point process. We discuss convergence of convex hulls, and of stochastic integrals of loss functions.

This set-up allows us to focus on the more analytic question of weak convergence of high risk scenarios, suitably normalized, and vague convergence of the associated Radon measures, in the remaining chapters. Interpretations in terms of point processes, convex hulls, stochastic loss integrals, and of the error terms associated with such integrals, are essential when handling real-life data. Such interpretations will be pointed out repeatedly in various settings throughout the book. Their validity follows from the theory developed in this chapter.

Starred sections may be skipped. They are not essential for the theory developed in Chapters II–IV.

1 An intuitive approach

This section introduces a number of important concepts in a rather intuitive fashion. Precise definitions will follow later. Impatient readers may follow the bold face page entries in the index. These lead to the formal definition.

1.1 A brief shower

Example 1.1. Suppose there has been a very brief rain shower. You are on an open space covered by tiles of one square meter, unit squares. There are only a few drops on each tile, say an average of $c = 5$ drops per tile. For any given tile the number of rain drops on the tile is random. Can one say anything about the distribution of this random integer, K? Can one describe the configuration of these K drops? Let C be the circle of radius $1/2$ inscribed in the tile. What is the probability that there are no drops inside the circle, and exactly one drop in each of the four regions outside the circle?

We divide the tile into $n = 4^m$ congruent squares. Write

$$K = K_1 + \cdots + K_n$$

where K_i denotes the number of drops in the ith subsquare. It may happen that one of these subsquares contains two drops or more, but the probability of this event decreases to zero as m grows, so we shall neglect it and assume K_i is zero or one. The expected number of drops in each of these n subsquares is the same, and hence equals c/n. Let p_n denote the probability of a drop in a subsquare. Under the simplifying assumption of at most one drop per subsquare we find

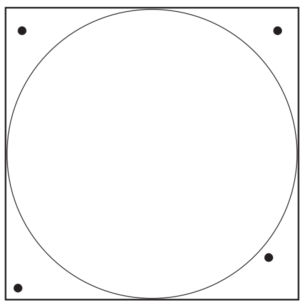
$$p_n = c/n.$$

The number of drops in any subsquare is not influenced by the drops which fall outside this subsquare. We therefore assume the 0-1 variables K_i to be independent. Their sum K then has a binomial distribution

$$\begin{aligned}\mathbb{P}\{K = k\} &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n \ldots (n + 1 - k)}{k!} \frac{c^k}{n^k} (1 - p_n)^{-k} (1 - c/n)^n \to \frac{c^k}{k!} e^{-c}.\end{aligned}$$

For $n \to \infty$ the limit is a Poisson distribution with expectation c. This answers our first question. The number of rain drops on the tile is Poisson with expectation $c = 5$.

Given that there falls precisely one rain drop on the tile, by assumption it is uniformly distributed over the tile. If there are precisely k drops, then each of these is uniformly distributed over the tile, at least if we order them sequentially in time. Again it is reasonable to assume independence. So conditionally on $K = k$ the rain drops on the tile form a sample cloud of k points from the *uniform distribution*.



Now consider the probability of no drops inside the circle C and one drop in each of the four regions outside the circle. In total there are four drops which have to be divided over four regions of equal probability $p := (1 - \pi/4)/4$. There are 4! ways to arrange this. The probability that four independent uniformly distributed points on the unit square each lie in one of the corner regions thus is $4!p^4$. Since $\mathbb{P}\{K = 4\} = e^{-c}c^4/4!$, the desired probability is

$$e^{-c}(pc)^4 \approx 0.000035, \quad c = 5.$$

Alternatively one may argue that under the assumption of non-interference the number of drops in the five regions are independent Poisson distributed variables with means pc (four times) and $c - 4pc$ (once), yielding the probability $(pce^{-pc})^4 e^{-(c-4pc)}$. ◊

What we want to do is to construct a mathematical model for the shower, or rather the trace of the shower on a tile. Let us first try to decide what further examples our theory should be able to handle.

1.2 Sample cloud mixtures. Suppose we do not look at the trace on one tile, but at the trace of a brief shower of intensity c on the first quadrant, which we assume to be divided into tiles. By assumption there is no interference between drops, so the number of drops on disjoint tiles are independent, as are their positions. The point process of rain drops on the first quadrant could be constructed from a sequence of Poisson random integers with expectation c, and a sequence of uniform random variables, all independent. Similarly one could construct a point process with intensity c on the whole plane.

Instead of points on a square one may consider points in a cube – think of raisins in an Easter bread – and use such cubes to construct a point process of intensity c in space. One might also consider point processes with a variable intensity – think of stars in space, or diamonds in the earth. Indeed let μ be any Radon measure on the plane. By definition a *Radon measure* is finite on compact sets. Dissect the plane into a sequence of tiles T_i, and let $d\mu_i = 1_{T_i} d\mu$ be the restriction of μ to the ith tile. Let K_i be a Poisson random integer with expectation $c_i = \mu_i(\mathbb{R}^2)$, and let $\pi_i = \mu_i / c_i$ denote the associated probability measure on the tile. We need only consider tiles of positive mass. Construct a point process with mean measure μ_i on the ith tile by first letting K_i choose an integer k, and then distributing k independent random points from the distribution π_i over the tile. Do this independently for each tile. We thus obtain a point process with state space $\mathbb{R}^2$ and mean measure μ.

The state space of a point process may be $\mathbb{R}^d$, an open subset of $\mathbb{R}^d$, or a d-dimensional manifold, with a Radon measure μ on the Borel sigma-field. Henceforth we shall assume that the state space is a separable metric space $\mathcal{X}$, and μ a σ-finite measure on the Borel sigma-field $\mathcal{B}$ of $\mathcal{X}$.

(One may throw out the topology altogether and consider σ-finite measures on a measurable space, say $(\Xi, \mathcal{E})$. If the underlying measure is infinite, and without atoms, and if the σ-field is generated by a countable collection of sets and separates points, then $(\Xi, \mathcal{E}, \mu)$ is isomorphic (as measure space) to the space $([0,\infty), \mathcal{B}, \lambda)$ where λ is Lebesgue measure on the Borel field $\mathcal{B}$ on $[0,\infty)$. One may embed Ξ in the halfline $[0,\infty)$ to produce an isometry between the Hilbert spaces $\mathbf{L}^2(\Xi, \mathcal{E}, \mu)$ and $\mathbf{L}^2([0,\infty), \mathcal{B}, \lambda)$; see Parthasarathy [1967]. So the most general example of a Poisson point process ends up as the standard Poisson process on the halfline $[0,\infty)$.)

Point processes on $[0,\infty)$ occur in many stochastic models. Think of renewal processes, survival processes, queueing theory, insurance claims, credit losses, or operational risk losses. Such point processes have the property that the points may be ordered in a canonical way. This allows one to introduce concepts such as interarrival times, compensators and martingales, which do not generalize easily to higher dimensions. We do not exclude such point processes from our theory, but typically our point processes live on open subsets of $\mathbb{R}^d$, with $d = 2$ or larger. On the other hand it should be noted that there is no strict division between point processes on the line and on $\mathbb{R}^d$ since one may add marks to the points on the line. In the opening example of a brief shower on a tile, one may take a dynamic approach and add a time coordinate, creating a three-dimensional state space. One could then describe this point process with points (X_k, Y_k, T_k), $1 \le k \le K$, as a marked point process on a time interval $[0, t_0)$, the duration of the shower, where T_k is the time when the drop hits the tile, and the mark for each time point T_k is the vector (X_k, Y_k) denoting the location of the corresponding drop on the tile. For the corresponding shower on the first quadrant such a description is not available.

The point process describing a shower with finite mean measure $\mu \neq 0$ is a mixture of sample clouds from the probability distribution $\pi = \mu/\|\mu\|$. An n-point *sample cloud* from a distribution π on $\mathcal{X}$ is a sample of n independent observations $X_1, \dots, X_n$ from this distribution. Such a sample cloud is a point process (with a fixed number of points). A mixture of sample clouds from a given distribution π is a *sample cloud mixture*. It has the form $X_1, \dots, X_K$ where $X_1, X_2, \dots$ are independent observations from the distribution π and $K \ge 0$ is a random integer, independent from the sequence (X_n). The mixing distribution $p_k = \mathbb{P}\{K = k\}$ for the shower above is Poisson, with mean $\|\mu\|$. As a special case of a sample cloud mixture we mention the *zero-one point process* determined by a finite measure μ of total mass $p = \|\mu\| \le 1$. With probability $1 - p$ this point process contains no points. If p is positive then the probability distribution $\pi = \mu/p$ is well defined and determines the location of the unique point of the 0-1 point process conditional on there being a point. Such 0-1 point processes crop up naturally if one has a random vector and restricts attention to a subset of the underlying vector space.

1.3 Random sets and random measures. One may regard a *point process* in $\mathbb{R}^d$ as a *random set*, locally finite and closed. In this set-up special care has to be taken of multiple points. A point process is *simple* if there are almost surely no multiple points. This will be the case for sample cloud mixtures if the underlying probability distribution π is *diffuse* (i.e. has no *atoms*). Stochastic geometry treats random closed sets in locally compact spaces. The convex hull of a sample cloud in $\mathbb{R}^d$ is a closed random set. Random subsets of the lattice $\mathbb{Z}^2$ are used in Ising models, infinite particle systems, percolation theory and image analysis. The theory of point processes may be regarded as a subtheory of stochastic geometry. It models the distribution of diamonds (rather than oil fields), the distribution of stars (rather than interstellar gas clouds), and the distribution of red blood cells (rather than liver tissue).

One may also regard a *point process* in $\mathbb{R}^d$ as a *random* (*counting*) *measure*, finite on compact sets. Sample clouds are essentially the same as empirical processes. Random (probability) measures are well known from the theory of conditional probabilities. Point processes are simpler to handle than random measures, but basically one needs the same set-up. One has to impose the condition that for bounded Borel sets in the state space $\mathcal{X}$ the number of points in this Borel set is a random variable. One may then define the *stochastic integral* with respect to a point process N for non-negative Borel functions f on the state space $\mathcal{X}$ as

$$\int f dN = \sum_{x \in N} f(x) = \sum f(X_k), \quad N(B) = \int 1_B dN,$$

where $X_1, X_2, \ldots$ is an enumeration of the points of N, using repetition to take multiplicity into account. The second sum is obviously measurable. It is less obvious that there exists an enumeration of the points of N (as a sequence of random elements in $\mathcal{X}$).

We shall adopt the second point of view in these notes, and treat point processes as random integer-valued measures.

1.4 The mean measure. The mean measure of a point process plays the same role as the expectation of a random variable. It always exists, but may be infinite. The reader should check that it indeed is a measure, non-negative and sigma-additive.

Definition. Let N be a point process on the separable metric space $\mathcal{X}$. Then

$$\nu : B \mapsto \mathbb{E}N(B), \quad B ß \mathcal{X} \text{ a Borel set},$$

is the *mean measure* of N.

Proposition 1.2. *Let N be a point process on the separable metric space $\mathcal{X}$ with mean measure ν, and let $f : \mathcal{X} \to [0, \infty]$ be measurable. Then*

$$\mathbb{E} \int f dN = \int f d\nu.$$

Proof. This is obvious if $f = c 1_E$ for $c \in (0, \infty)$. Now write $f = \sum c_n 1_{E_n}$ with E_n measurable and $c_n \in (0, \infty)$. This decomposition is always possible. The monotone convergence theorem gives the desired result. □

Example 1.3. The mean measure may be infinite for many sets. Let π be the standard normal distribution on $\mathbb{R}^2$ and let N be the corresponding sample cloud mixture, with the mixing sequence $p_k = 1/(k+1)(k+2) = \mathbb{P}\{N(\mathbb{R}^2) = k\}$, $k = 0, 1, \ldots$. Then N is finite almost surely, but νE is infinite or zero for each Borel set E in the plane. ♢

Exercise 1.4. If the mean measure ν is infinite, but N is finite on bounded sets, one may still construct a finite measure ν^* with the same null sets as ν, since ν is a sum of finite measures. (Partition $\mathcal{X}$ into bounded Borel sets, then condition on the number of points in the set.) ♢

Exercise 1.5. Suppose N is a point process on the separable metric space $\mathcal{X}$. Let N^B denote the restriction of N to the Borel set B:

$$N^B(E) = N(B \cap E), \quad E \in \mathcal{B}.$$

Show that N^B is a sample cloud mixture if N is. ◊

1.5* Enumerating the points. Let $X_1, X_2, \dots$ be random points in the metric space $\mathcal{X}$. If $d(x, X_n) \to \infty$ holds a.s. for some $x \in \mathcal{X}$, then this holds for every x, and the sequence (X_n) is the enumeration of a point process N on $\mathcal{X}$ which is finite on bounded Borel sets. Conversely any such point process N on $\mathcal{X}$ may be enumerated (in many ways). We describe one such enumeration. The reader is invited to construct his own.

Definition. A *partition* $\mathcal{C}$ of a separable metric space $\mathcal{X}$ is a countable (finite or infinite) collection of disjoint non-empty Borel sets $C_1, C_2, \dots$, the *atoms*, which cover $\mathcal{X}$. A sequence of partitions $\mathcal{C}_n$ is *increasing* if each atom of $\mathcal{C}_n$ is contained in an atom of $\mathcal{C}_{n-1}$, her mother. The σ-algebra generated by $\mathcal{C}_n$ then contains the σ-algebra generated by $\mathcal{C}_{n-1}$. The increasing sequence $(\mathcal{C}_n)$ *separates points* if for each pair of points $x \neq y$ in $\mathcal{X}$ there exists an index n such that x and y lie in different atoms of $\mathcal{C}_n$.

A standard example is the sequence of cubic partitions of $\mathbb{R}^d$. Here $\mathcal{C}_n$ consists of the cubes

$$K = [(k_1 - 1)/2^n, k_1/2^n) \times \cdots \times [(k_d - 1)/2^n, k_d/2^n), \quad k_i \in \mathbb{Z},\ i = 1, \dots, d. \tag{1.1}$$

There are many others. A separable metric space has a countable base, $U_1, U_2, \dots$, of bounded open sets. Let $\mathcal{C}_n$ be the partition generated by the sets $U_1, \dots, U_n$. This sequence of partitions is increasing, and has the *small set property*:

(SS) for each $z \in \mathcal{X}$ and each $\delta > 0$ there exist $n \geq 1$ and $j \geq 1$ such that

$$z \in C_{nj} \text{ß} B^{\delta}(z) := \{x \in \mathcal{X} \mid d(x, z) < \delta\}. \tag{1.2}$$

The small set condition (SS) ensures that $\bigcup \mathcal{C}_n$ generates the Borel σ-field on $\mathcal{X}$, since any open set is the union of the atoms C_{nj} which it contains.

An increasing sequence of partitions $\mathcal{C}_n$ which separates points of $\mathcal{X}$ yields an *enumeration of the points* of a point process N on $\mathcal{X}$: With each $x \in \mathcal{X}$ associate a unique positive integer sequence $j_1, j_2, \dots$ such that $x \in C_{nj_n}$ for $n = 1, 2, \dots$. The address (j_n) allows us to order the points of N lexicographically.

1.6 Definitions. It should now be clear what a point process is. A random measure N on a separable metric space $\mathcal{X}$ with values in the set $\{0, 1, \ldots, \infty\}$. We impose the condition that small sets almost surely contain only finitely many points of N. Here small may mean that the closure is compact. It may also mean that the set is bounded. Or that the mean measure of the set is finite. These concepts do not coincide. At this moment we shall assume that the state space is a separable metric space, without being too specific about small sets. This allows us to include point processes on separable infinite-dimensional Banach spaces, and on arbitrary Borel sets in $\mathbb{R}^d$. It allows us to construct Poisson point processes whose mean measure is any given σ-finite measure on $\mathcal{X}$. It is only when we consider convergence of point processes in Section 4, and when we apply the theory to point processes on open subsets of $\mathbb{R}^d$, that we restrict attention to point processes on locally compact spaces, whose realizations are Radon measures, with values in $\{0, 1, \ldots, \infty\}$.

Definition. A *point process* N on a separable metric space $\mathcal{X}$, the *state space*, is a random measure $\omega \mapsto N(\omega)$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The following conditions hold:

1) $N(\omega)$ is a measure on the Borel σ-field $\mathcal{B}$ of $\mathcal{X}$ for each $\omega \in \Omega$;

2) $N(\omega)(E) \in \{0, 1, \ldots, \infty\}$ for all Borel sets Eß$\mathcal{X}$ and all $\omega \in \Omega$;

3) $\{N(E) = k\}$ is an event for each Borel set Eß$\mathcal{X}$ and for each integer $k = 0, 1, \ldots$;

4) the space $\mathcal{X}$ is covered by a sequence of bounded Borel sets on which N is finite a.s.

The choice of the sequence of bounded Borel sets depends on the situation. For Poisson point processes it is natural to choose sets of finite mean measure; for point processes on separable metric spaces one may choose open balls; if the space is locally compact, one may choose compact sets.

Basic references are Daley & Vere-Jones [2003] for point processes on separable metric spaces, finite on bounded sets; Kallenberg [2002], Chapter 14 and Appendix, for point processes on locally compact separable metric spaces, finite on compact sets; and Kingman [1993] who concentrates on Poisson point processes, and imposes minimal conditions. Kallenberg occasionally treats simple point processes as discrete closed sets. For statistics for point processes, see Reiss [1993], Karr [1991], or Jacobsen [2006].

Exercise 1.6. Prove that $\{N(\mathcal{X}) = \infty\}$ is an event. $\Diamond$

2 Poisson point processes

Poisson point processes will be used to describe the extremal behaviour of sample clouds. We begin with a constructive description in terms of sample cloud mixtures.

2.1 Poisson mixtures of sample clouds. Let N be a Poisson mixture of sample clouds on $\mathcal{X}$ with finite mean measure $\mu \neq 0$. So the mixing sequence is $e^{-c}c^n/n!$ with $c = \mu(\mathcal{X})$. The underlying distribution for the sample clouds is $\pi = \mu/c$. Let $E_0, \dots, E_m$ be a partition of $\mathcal{X}$ (with E_i measurable) and set

$$K_i = N(E_i), \quad K = N(\mathcal{X}) = K_0 + \cdots + K_m, \quad c_i = \mu E_i = cp_i, \quad p_i = \pi E_i.$$

Let $k_0, \dots, k_m$ be non-negative integers with sum $k \geq 0$. Then conditional on $N(\mathcal{X}) = k$ the k points of N form a sample cloud from the distribution π (by definition), and hence the number of points in the $m + 1$ subsets E_i of the partition have a multinomial distribution:

$$\mathbb{P}(K_0 = k_0, \dots, K_m = k_m \mid N(\mathcal{X}) = k) = p_0^{k_0} \dots p_m^{k_m} \cdot k!/(k_0! \dots k_m!).$$

Since $\mathbb{P}\{K = k\} = e^{-c}c^k/k!$ we find

$$\begin{aligned}\mathbb{P}\{K_0 = k_0, \dots, K_m = k_m\} &= \mathbb{P}(K_0 = k_0, \dots, K_m = k_m \mid K = k)\mathbb{P}\{K = k\} \\ &= (k!/k_0! \dots k_m!) p_0^{k_0} \dots p_m^{k_m} \cdot e^{-c}c^k/k! \\ &= \prod_{j=0}^{m} e^{-c_j} c_j^{k_j}/k_j!\,.\end{aligned}$$

This proves

Proposition 2.1. *Let N be a Poisson mixture of sample clouds with finite mean measure μ, and let $E_1, \dots, E_m$ be disjoint Borel sets. Then the variables*

$$K_j := N(E_j), \quad j = 1, \dots, m$$

are independent, and Poisson distributed with expectation $c_j = \mu E_j$.

Does the converse hold? Suppose N is a point process on $\mathcal{X}$ with finite mean measure μ, and

$$\begin{aligned}&\mathbb{P}\{N(E_1) = k_1, \dots, N(E_m) = k_m\} \\ &\qquad = \prod_{j=1}^{m} e^{-\mu E_j} (\mu E_j)^{k_j}/k_j!, \quad k_i \geq 0,\ i = 1, \dots, m,\end{aligned}$$

for all finite sequences of disjoint Borel sets $E_1, \dots, E_m$ in $\mathcal{X}$. Does it follow that N is a Poisson sample cloud mixture?

The real question here is: When do two point processes have the same distribution? In the next lecture, Section 3, we shall introduce the space of counting measures to define the distribution of a point process. Here we take a more intuitive approach.

2.2 The distribution of a point process. Raindrops on a tile are not ordered (unless we see them fall). We have to distinguish between a finite point process and a finite sequence of points. Configurations of rain drops on the tile are described by constructing finite partitions of the tile, and counting how many drops each of the atoms of the partition contains. On a tile it is not hard to devise an enumeration of the drops – introduce coordinates, and use a lexicographical ordering, like reading a page of a book. However such an ordering is artificial, and destroys the geometry of the configuration. When working with point processes, one has to get used to counting points in sets.

Suppose we have two finite point processes M and N on the separable metric space $\mathcal{X}$. For simplicity assume both have exactly k points. Suppose for any finite collection of disjoint Borel sets $E_1, \dots, E_m$ in $\mathcal{X}$ and non-negative integers $k_1, \dots, k_m$

$$\mathbb{P}\{N(E_1) = k_1, \dots, N(E_m) = k_m\} = \mathbb{P}\{M(E_1) = k_1, \dots, M(E_m) = k_m\}. \tag{2.1}$$

Does it follow that M and N have the same distribution?

Let $X = (X_1, \dots, X_k)$ be an enumeration of N, and $Y = (Y_1, \dots, Y_k)$ an enumeration of M. If $M = N$ the finite sequences X and Y need not have the same distribution. (Think of a finite sample in $\mathbb{R}$, and the corresponding order statistics.) Introduce a random permutation S, uniformly distributed over the group $\mathcal{S}_k$ of all $k!$ permutations of $\{1, \dots, k\}$, and independent of X and of Y, and define $X_i^* = X_{S(i)}$ for $i = 1, \dots, k$, and Y^* similarly. The starred k-tuples are *exchangeable*. If they have the same distribution then (2.1) holds. We shall show the converse: (2.1) implies that X^* and Y^* have the same distribution.

Set

$$E_0 = \mathcal{X} \setminus (E_1 \cup \dots \cup E_m), \quad k_0 = k - (k_1 + \dots + k_m).$$

Consider k-tuples $(F_1, \dots, F_k)$ of sets F_i which contain k_j copies of E_j for $j = 0, \dots, m$. There are $n = k!/(k_0! \dots k_m!)$ distinct sequences $(F_1, \dots, F_k)$ of this kind. By exchangeability the products $F_1 \times \dots \times F_k$ all have the same probability with respect to X^* and the union U of these n disjoint product sets has the same probability for X and for X^* since the union is invariant under permutations of the coordinates:

$$\mathbb{P}\{X \in U\} = \mathbb{P}\{X^* \in U\} = \mathbb{P}\{N(E_1) = k_1, \dots, N(E_m) = k_m\}.$$

Now observe that there is a one-one correspondence between the distribution of a k-point point process N on $\mathcal{X}$ and the distribution of an exchangeable k-tuple X^* since one may write

$$\mathbb{P}\{X_1^* \in F_1, \dots, X_k^* \in F_k\} = \frac{k_0! \dots k_m!}{k!} \mathbb{P}\{N(E_1) = k_1, \dots, N(E_m) = k_m\}. \tag{2.2}$$

The left side determines the distribution of X^*. (Indeed the distribution of a k-tuple X is determined by the probabilities $\mathbb{P}\{X_1 \in B_1, \dots, X_k \in B_k\}$, where $B_1, \dots, B_k$ are Borel sets in $\mathcal{X}$. If $E_0, \dots, E_m$ is a partition of $\mathcal{X}$ and each Borel set B_i is a union of atoms E_i, then the event $\{X_1 \in B_1, \dots, X_k \in B_k\}$ is a disjoint union of events $\{X_1 \in F_1, \dots, X_k \in F_k\}$, with $F_i \in \{E_0, \dots, E_m\}$.)

Conclusion. There is a one-one correspondence between the distribution of k-point point processes N on $\mathcal{X}$, and the distribution of exchangeable k-tuples $X^* \in \mathcal{X}^k$. The correspondence is given by (2.2).

2.3 Definition of the Poisson point process. We now have an alternative, analytic description of Poisson point processes.

Definition. Let μ be a σ-finite measure. A point process N with *mean measure* μ is called a *Poisson point process* if $N(E)$ is a Poisson random variable with expectation μE for every measurable set E of finite mass $\mu(E)$, and if the random variables

$$N(E_1), \dots, N(E_m)$$

are independent whenever the sets $E_1, \dots, E_m$ have finite mass and are disjoint.

Proposition 2.2. *The sum of independent Poisson point processes is a Poisson point process.*

Proof. Let N have mean measure ν and let M have mean measure μ. By independence $(N+M)(E)$ is a Poisson random integer with mean measure $(\nu+\mu)(E)$, and independence of $N(E_1), \dots, M(E_m)$ implies independence of $(N+M)(E_1), \dots, (N+M)(E_m)$. □

Proposition 2.3. *The restriction of a Poisson point process to a measurable set is a Poisson point process. Restrictions to disjoint sets are independent.*

Proof. The first statement follows immediately from the definition. For the second statement partition each of the sets, and observe that the numbers of points in the atoms are independent variables. □

Proposition 2.4. *If $\mathcal{X}$ is covered by an increasing sequence of Borel sets B_n, and the point process N on $\mathcal{X}$ has the property that each restriction $dN^{B_n} = 1_{B_n} dN$ is a finite Poisson point process, then N is a Poisson point process.*

Proof. First observe that N^{B_n} has finite measure μ_n which lives on B_n, and that μ_{n+1} extends μ_n. This defines a σ-finite measure μ on $\mathcal{X}$. For any Borel set E of finite mass the random variable $N(E)$ is Poisson with expectation $\mu(E)$ as limit of the increasing Poisson variables $N(E \cap B_n)$ with expectation $\mu(E \cap B_n)$. Independence of $N(E_i)$ for disjoint Borel sets $E_1, \dots, E_k$ of finite mass follows from the independence of $N_{1n}, \dots, N_{kn}$ for fixed n, with $N_{in} = N(E_i \cap B_n)$, and the convergence $\mathbb{P}\{N_{in} = m\} \to c_i^m e^{-c_i}/m!$ with $c_i = \mu(E_i)$ for $n \to \infty$. $\square$

Exercise 2.5. Check that the point process of rain drops on the plane with Radon measure μ defined in terms of a sequence of tiles as described in Section 1 is indeed the Poisson point process with mean measure μ. $\Diamond$

2.4 Variance and covariance. Independence is the most prominent characteristic of Poisson point processes. In whatever way we cut up the underlying space, the restrictions are independent. This *independence* allows us to establish a simple formula for the variance of stochastic integrals with respect to Poisson point processes.

Let N be a Poisson point process on the separable metric space $\mathcal{X}$ with σ-finite mean measure μ. If f is a step function, $f = c_1 1_{B_1} + \cdots + c_m 1_{B_m}$, with B_i disjoint Borel sets of finite measure, then $\int f dN$ is a linear combination of the independent Poisson variables $N(B_i)$, with variance $\mu(B_i)$, and hence one may write

$$\operatorname{var}\left(\int f dN\right) = c_1^2 \mu(B_1) + \cdots + c_m^2 \mu(B_m) = \int f^2 d\mu.$$

If $0 \le f_n \uparrow f$ then $Z_n := \int f_n dN \uparrow Z := \int f dN$, and both $\mathbb{E}(Z_n) \to \mathbb{E}(Z)$ and $\mathbb{E}(Z_n^2) \to \mathbb{E}(Z^2)$ by monotone convergence. So $\operatorname{var}(Z_n) \to \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2$ if $\int f d\mu$ is finite. We find

Proposition 2.6. *Let N be a Poisson point process with mean measure μ, and f a Borel function. If $\int |f| d\mu$ is finite then*

$$\mathbb{E} \int f dN = \int f d\mu, \quad \operatorname{var} \int f dN = \int f^2 d\mu \le \infty.$$

Proof. For real valued f write $f = f 1_E + f 1_{E^c}$ with $E = \{f > 0\}$, and use independence of $\int f 1_E dN$ and $\int f 1_{E^c} dN$. $\square$

The polarization formula, $2ab = (a+b)^2 - a^2 - b^2$, yields the *covariance*

$$\operatorname{cov}\left(\int f dN, \int g dN\right) = \int fg d\mu, \quad f, g \in (\mathbf{L}^1 \cap \mathbf{L}^2)(\mu). \tag{2.3}$$

Exercise 2.7. Let N_n be the n-point sample cloud from the distribution π on $\mathcal{X}$. Then

$$\mathbb{E}\int f dN_n = n\int f d\pi, \quad \operatorname{var}\Big(\int f dN_n\Big) = n\Big(\int f^2 d\pi - \Big(\int f d\pi\Big)^2\Big). \tag{2.4}$$

Write $\int f dN = f(X_1) + \cdots + f(X_n)$, and use independence! ◊

If $\mu(\mathcal{X})$ is infinite it is possible that $\int f^2 d\mu$ is finite while $\int f d\mu = \infty$. The variance of $\int f dN$ exists, but the expectation is infinite! This situation may be handled by introducing the *compensated* Poisson point process $N - \mu$. Partition $\mathcal{X}$ into Borel sets B_n, of finite measure, and set

$$Z_n = \int f 1_{B_n} dN - \int f 1_{B_n} d\mu = \int f 1_{B_n} d(N - \mu).$$

The random variables Z_n are independent and centered, and $\sum \mathbb{E}(Z_n^2) = \int f^2 d\mu$ is finite. Hence the martingale $X_n = Z_1 + \cdots + Z_n$ converges in $\mathbf{L}^2$ and almost surely. We write the limit X as

$$X = \int f d(N - \mu), \quad \mathbb{E}(X) = 0, \quad \mathbb{E}X^2 = \int f^2 d\mu.$$

These integrals play a role in the theory of Lévy processes.

2.5* The bivariate mean measure. Let us say a few words about the second order theory in general. For a point process N on $\mathcal{X}$ with points $X_1, X_2, \ldots$ define the bivariate integral for non-negative Borel functions h on $\mathcal{X}^2$ by

$$\int h dN^{(2)} = \sum_{i \neq j} h(X_i, X_j).$$

This determines a measure $\mu^{(2)}$ on $\mathcal{X} \times \mathcal{X}$, the *bivariate mean measure*, by

$$\int h d\mu^{(2)} = \mathbb{E}\int h dN^{(2)} = \sum_{i \neq j} \mathbb{E}h(X_i, X_j). \tag{2.5}$$

For non-negative real Borel functions f and g on $\mathcal{X}$ set $(f \otimes g)(x, y) = f(x)g(y)$. Then

$$\int f dN \int g dN = \int f \otimes g dN^{(2)} + \int fg dN.$$

Assume the mean measure μ is σ-finite on $\mathcal{X}$, and f and g are integrable. Taking expectation we find

$$\operatorname{cov}\Big(\int f dN, \int g dN\Big) + \int f d\mu \int g d\mu = \int f \otimes g d\mu^{(2)} + \int fg d\mu.$$

Now restrict to indicator functions of disjoint Borel sets F and G of finite mean measure. The formula for the covariance simplifies

$$\mathrm{cov}(N(F), N(G)) = \mu^{(2)}(F \times G) - \mu(F)\mu(G),$$

$$\mathrm{var}(N(F)) = \mu^{(2)}(F \times F) + \mu(F)(1 - \mu(F)).$$

Bivariate mean measures may be defined locally by restricting the point process to small sets.

Let N be a point process on the open set Oß$\mathbb{R}^d$ with finitely many points, $X_1, \ldots, X_K$. Assume that $(X_1, \ldots, X_K)$ conditional on $K = k$ has density f_k on O^k. Let f_{kij} denote the bivariate density of (X_i, X_j) on O^2 conditional on $K = k$. The bivariate *Janossi density* is defined by

$$j_2(x, y) = \sum_k p_k \sum_{1 \le i \ne j \le k} f_{kij}(x, y), \quad p_k = \mathbb{P}\{K = k\}. \tag{2.6}$$

Proposition 2.8. *With the notation and assumptions above,*

$$\mathbb{P}\{N(dx) = 1, N(dy) = 1\} = j_2(x, y)dxdy = \mu^{(2)}(dxdy)$$

for $x \ne y$, $x, y \in O$ß$\mathbb{R}^d$.

Proof. We may assume $K = k > 1$. Let f be the density of $X = (X_1, \ldots, X_k)$ on O^k, and f_{ij} the density of (X_i, X_j). Set

$$V = v(X), \quad v(x) = \min_{i \ne j} \|x_j - x_i\|, \quad x \in O^k.$$

Then v is positive a.e. and hence $f_n = f 1_{\{v > 1/n\}} \uparrow f$ a.e. This also holds for the bivariate marginals f_{nij}, and their sum $f_n^{(2)}$ over $1 \le i \ne j \le k$. Let F and G be Borel sets in O of diameter less than $\delta = 1/2n$, and at distance more than δ from each other. Then

$$\{X_i \in F, X_j \in G, V > 2\delta\}\text{ß}\{X_i \in F, X_j \in G; X_k \in (F \cup G)^c, k \ne i, j\}$$
$$\text{ß}\{X_i \in F, X_j \in G\}.$$

On adding the probabilities for $1 \le i \ne j \le k$ we find that $\mathbb{P}\{N(F) = 1, N(G) = 1\}$ is enclosed between the integral of $f_n^{(2)}$ and $f^{(2)} = j_2$ over the set $F \times G$ in O^2. This yields the first equality. The second follows from (2.5) and the definition of j_2. □

The Janossi densities j_m for $m > 2$ may be defined similarly. See Daley & Vere-Jones [2003] for details.

2.6 Lévy processes. Given a Poisson point process on $\mathcal{X}$ with mean measure ρ one can add a time coordinate, and consider the Poisson point process N on $\mathcal{X} \times [0, \infty)$ with mean measure $d\rho(x)dt$ and points (X_k, T_k). If $\mathcal{X}$ is a subset of $\mathbb{R}^d$, the spatial components X_k of the points in N may be added. Sums for disjoint time slices are independent by the basic property of Poisson point processes. So if the sums

$$S(t) = \sum \{X_k \mid T_k \leq t\}, \quad t \geq 0,$$

converge, the process $S\colon [0, \infty) \to \mathbb{R}^d$ has independent increments, and the increments are stationary since the distribution of N on the time slice $(0, s]$ and $(t, t + s]$ is the same as far as the space coordinate is concerned. By summing (or integrating) a Poisson point process we obtain a Lévy process. In the theory of Lévy processes it is often easier to use the "derivative" of the process, the underlying Poisson point process, then the process itself. A Lévy process may also have a Brownian component and a drift, but if the jump part generated by the Poisson point process is absent one will usually call the process a Brownian motion. Since Poisson point processes are simple, so are Lévy processes, unless one asks hard questions, or has to deal with the Brownian part.

Definition. A random process $X\colon [0, \infty) \to \mathbb{R}^d$ is a *Lévy process* if

1) it has independent increments;
2) the increments are stationary;
3) $X(0) = 0$ almost surely;
4) the sample functions are right-continuous with left-hand limits.

Theorem 2.9. *Let ρ be a Radon measure on $O = \mathbb{R}^d \setminus \{0\}$ such that*

$$\int x^T x \wedge 1 d\rho(x) < \infty.$$

Let $\chi\colon \mathbb{R}^d \to [0, 1]$ be a Borel function which is one on a neighbourhood of the origin and vanishes outside a bounded set. Let N be the Poisson point process on $O \times [0, \infty)$ with mean measure $d\rho(x)dt$. There exists a Lévy process $X\colon [0, \infty) \to \mathbb{R}^d$ such that for $t > 0$

$$X(t) = \int_{O \times [0,t]} x(dN - \chi(x)d\rho(x))dt \quad a.s.$$

If we replace χ by a different function $\bar{\chi}$ with the same properties, then for all ω outside a null set

$$\bar{X}(t, \omega) = X(t, \omega) + bt, \quad t \geq 0, \ b = \int \bar{\chi} - \chi d\rho.$$

Proof. Choose $r > 0$ so that χ and $\bar{\chi}$ are one on rB, where B is the open unit ball. We split the integral for X into two parts:

$$X_1(t) = \int_{rB^c \times [0,t]} x dN - tq, \quad q = \int_{rB^c} x\chi(x) d\rho(x),$$
$$X_0(t) = \int_{(rB\setminus\{0\})\times[0,t]} x d\tilde{N}, \quad d\tilde{N} = dN - d\rho ds.$$

The process X_1 is the sum of a linear drift $-tq$ and a step process

$$Z_1 1_{[S_1,\infty)} + Z_2 1_{[S_2,\infty)} + \cdots,$$

where (Z_n, S_n) are the points of N outside rB ordered so that $S_1 < S_2 < \cdots$. The S_n form a Poisson point process on $[0,\infty)$ with intensity $\rho(rB^c)$, and hence $S_n \to \infty$ a.s. The process X_0 is an $\mathbf{L}^2$-martingale, and $\mathbb{E}\|X_0(t)\|^2 = t \int_{rB\setminus\{0\}} x^T x d\rho(x) < \infty$. It is well-known that there is a version with sample functions which are right-continuous with left limits. See for instance Kallenberg [2002], Theorem 6.27. Now observe that both X_0 and X_1 satisfy the four conditions of a Lévy process. So does their sum since they are independent. Since $\overline{X}_0 = X_0$ we have $\overline{X}(t) - \bar{q}t = X(t) - qt$. □

Without proof we mention

Theorem 2.10. *A Lévy process is the sum of a process X as above, a linear drift, and a centered Brownian motion independent of X.*

Proof. See Kallenberg [2002], Theorem 13.4. □

It should be emphasized that the drift term depends on the compensator function χ. It is customary to take $\chi = 1_B$, but sometimes it is convenient to have a continuous compensator. If $\int \|x\| \wedge 1 d\rho(x)$ is finite, no compensator is needed. The Poisson point process may go over into white noise, the "derivative" of Brownian motion. If for instance ρ_n is concentrated on $B \setminus \{0\}$, $\rho_n \to 0$ vaguely on $B \setminus \{0\}$ and $\int \xi_i \xi_j d\rho_n \to c_{ij}$ for $1 \le i, j \le d$, then the associated Lévy processes converge to a Brownian motion W with *covariance* $\mathbb{E} W_i(t) W_j(t) = c_{ij} t$.

Exercise 2.11. Let $d\rho(x) = 1_{(0,\infty)}(x) dx/x$ on $\mathbb{R} \setminus \{0\}$. The Lévy process is an asymmetric *stable process* of index one. The sample functions have infinitely many jumps on any interval (s,t) with $0 \le s < t$. The jumps $S(t) - S(t-0)$ are positive and the sum of the jumps over (s,t) is infinite almost surely. On the other hand, right-continuous functions with left-hand limits are bounded on any interval $[0,t]$. Describe the sample functions. ◊

Exercise 2.12. If the Lévy measure ρ is infinite, as in the exercise above, the sequence of time points $T_1, T_2, \ldots$ is dense on the time axis $[0,\infty)$. Prove that there are no multiple points: $T_i \ne T_j$ holds almost surely for $i \ne j$. Show that any Borel set Eß$[0,\infty)$ a.s. contains infinitely many points, or none. ◊

2.7 Superpositions of zero-one point processes. Let N be the sum (or superposition) of 0-1 point processes N_n with mean measures μ_n. Then N has mean measure $\mu = \sum \mu_n$. If the measures μ_n are small and the processes N_n independent, then N is close to a Poisson point process. The proof is based on an elementary inequality.

Lemma 2.13. *Let N be a* 0-1 *point process with mean measure μ. If the underlying probability space is sufficiently rich, there exists a Poisson point process M with mean measure μ, such that*

$$\mathbb{P}\{M \neq N\} \leq p^2, \quad p = \|\mu\|.$$

Proof. We may assume that p is positive. Let $X_1, X_2, \dots$ be an iid sequence from the distribution $\pi = \mu/p$. Then $N = \sum_{j=1}^{N_0} \delta_{X_j}$ where δ_X denotes the point mass in X, and N_0 is a 0-1 variable, independent of (X_n) with $\mathbb{P}\{N_0 = 1\} = p$. Let M_0 be a Poisson variable independent of (X_n) with expectation p and set $M = \sum_{j=1}^{M_0} \delta_{X_j}$. Then $\mathbb{P}\{M = N\} = \mathbb{P}\{M_0 = N_0\}$. Since

$$\mathbb{P}\{M_0 = 0\} = e^{-p} > 1 - p = \mathbb{P}\{N_0 = 0\},$$

we may choose M_0 such that $\{M_0 = 0\} \supset \{N_0 = 0\}$. Then $\mathbb{P}\{M_0 \leq 1\} = (1 + p)e^{-p}$ gives

$$\mathbb{P}\{M_0 \neq N_0\} = e^{-p} - (1 - p) + (1 - (1 + p)e^{-p}) = p(1 - e^{-p}) < p^2.$$

We need some freedom to construct the Poisson variable. On a non-trivial two-point probability space one can define a non-zero 0-1 point process, but not a non-zero Poisson point process, on $\mathcal{X} = \{x_0\}$. □

Proposition 2.14. *Let N_n be independent* 0-1 *point processes with mean measures μ_n on $\mathcal{X}$. Assume $\mu = \sum \mu_n$ is σ-finite. If the underlying probability space is sufficiently rich, there exists a Poisson point process M with mean measure μ such that*

$$\mathbb{P}\{M \neq N_1 + N_2 + \cdots\} \leq \sum \|\mu_i\|^2. \tag{2.7}$$

Proof. Construct independent Poisson point processes $M_1, M_2, \dots$ with mean measures $\mu_1, \mu_2, \dots$ so that $\mathbb{P}\{M_n \neq N_n\} \leq \|\mu_n\|^2$, and let M be the sum of the point processes M_n. □

Here is an application. Large sample clouds locally look like Poisson point processes, since the binomial distribution for small values of p looks like a Poisson distribution. Let N be the n-point sample cloud from the distribution π on $\mathbb{R}^d$. Let O be open with $n\pi(O) = c > 0$. If the underlying probability space is rich enough, Proposition 2.14 yields a Poisson point process M on O with mean measure $d\mu = n1_O d\pi$ such that

$$\mathbb{P}\{M \neq N\} \leq \frac{c}{n}.$$

In fact one can do better, using Barbour's Poisson approximation to the binomial distribution: For independent events A_n there is a Poisson variable M such that

$$\mathbb{P}\{N \neq M\} \leq \sum p_n^2(1 - e^{-\lambda})/\lambda, \quad \mathbb{E}N = \mathbb{E}M,$$
$$N = \sum 1_{A_n}, \quad p_n = \mathbb{P}(A_n), \quad \lambda = \sum p_n.$$

See Barbour, Holst & Janson [1992]. For Poisson point process approximations to dependent sums of Bernoulli variables see Barbour & Chryssaphinou [2001]; for bounds on the Hellinger distance see Falk, Hüsler & Reiss [2004].

Exercise 2.15. N is a 0-1 point process on a sufficiently rich probability space. There is a Poisson point process $M \geq N$ such that $\mathbb{P}\{M \neq N\} \leq p^2$ for $p = \mathbb{P}\{\|N\| = 1\} \leq 1/2$. ◊

The role of the Gaussian distribution for sums of random variables is taken over by the Poisson point process for sums of independent point processes. The next two sections are needed to understand the theorem below, a kind of central limit theorem for point processes.

Theorem 2.16 (Grigelionis, Superposition of Point Processes). *For $n = 1, 2, \dots$ let $N_{n1}, N_{n2}, \dots$ be independent point processes on the separable metric space $\mathcal{X}$. Let μ be a measure on $\mathcal{X}$ which is finite on bounded Borel sets. Suppose for each bounded Borel set E with $\mu(\partial E) = 0$ we have*

$$\begin{aligned} \mathbb{P}\{N_{ni}(E) > 0\} &\to 0, & n + i &\to \infty, \\ \textstyle\sum_i \mathbb{P}\{N_{ni}(E) > 0\} &\to \mu(E), & n &\to \infty, \\ \textstyle\sum_i \mathbb{P}\{N_{ni}(E) > 1\} &\to 0, & n &\to \infty. \end{aligned}$$

There exists a Poisson point process M with mean measure μ such that $M_n := \sum_i N_{ni} \Rightarrow M$, in the sense that

$$\int \varphi dN_n \Rightarrow \int \varphi dM$$

for all bounded μ-a.e. continuous Borel functions $\varphi \colon \mathcal{X} \to \mathbb{R}$ for which $\{\varphi \neq 0\}$ is bounded.

Proof. To follow. See the end of Section 4. □

Here we only want to observe that the boundary of a Borel set E may contain E (if E has no interior points). However, for open or closed sets, the condition that $\mu(\partial E)$ equals 0 is the rule rather than the exception. Recall that the complement of a countable set is dense in $\mathbb{R}$.

Lemma 2.17. *Let μ be a σ-finite measure on the separable metric space $\mathcal{X}$, and $f: \mathcal{X} \to \mathbb{R}$ continuous. Define the open and closed sets*

$$O_t = \{f < t\}, \quad F_t = \{f \leq t\}, \quad t \in \mathbb{R}.$$

The boundaries ∂O_t, $t \in \mathbb{R}$, of the open sets are disjoint, as are the boundaries of the closed sets. There is a countable set Aß$\mathbb{R}$ such that

$$\mu(\partial O_t) = \mu(\partial F_t) = 0, \quad t \in \mathbb{R} \setminus A.$$

Proof. Both ∂O_t and ∂F_t are contained in the closed set $\{f = t\}$, and these sets are obviously disjoint. Write μ as the sum of a sequence of finite measures μ_n and define A as the union of the finite sets $A_{nm} = \{t \in \mathbb{R} \mid \mu_n\{f = t\} > 1/m\}$. □

2.8 Mappings. Poisson point processes are preserved under measurable mappings, at least if the image of the mean measure is σ-finite again.

Theorem 2.18 (Mapping Theorem). *Let $\mathcal{X}$ and $\mathcal{Y}$ be separable metric spaces, and $y: \mathcal{X} \to \mathcal{Y}$ a measurable map. Let N be a Poisson point process on $\mathcal{X}$ with points $X_1, X_2, \ldots$, and σ-finite mean measure ν. Suppose the image $\mu = y(\nu)$ on $\mathcal{Y}$ is also σ-finite. Then $M = y(N)$, the image process, with points $Y_i = y(X_i)$, is a Poisson point process with mean measure $\mu = y(\nu)$.*

Proof. We first check that $M(B)$ is Poisson with expectation $\mu(B)$ for Borel sets B in $\mathcal{Y}$ of finite mass. Let $E = y^{-1}(B)$. Then $\mu(B) = \nu(E)$ and $M(B) = N(E)$. Hence $\nu(E)$ is finite and $M(B) = N(E)$ is Poisson with expectation $\mu(B)$. Next observe that independence of the variables $M(B_i) = N(E_i)$ for $B_1, \ldots, B_m$ disjoint Borel sets in $\mathcal{Y}$ holds since the sets $E_i = y^{-1}(B_i)$ are disjoint and N is Poisson. □

Example 2.19. The map $y(t) = (e^{2\pi i t}, e^{it})$ maps the halfline $[0, \infty)$ onto a dense subset of the torus, $S \times S$, where S is the unit circle in the complex plane. For any non-empty open set U in the torus, the inverse image is open. It contains infinitely many intervals of length greater than ε for some $\varepsilon > 0$. It follows that the image of the standard Poisson point process N on $[0, \infty)$, with points $T_1 < T_2 < \cdots$, is dense in the torus. By the Mapping Theorem $y(N)$ is a Poisson point process on the torus. ◊

2.9* Inverse maps. Poisson point processes exhibit an unlimited amount of independence. If one partitions $\mathcal{X}$ into sets C_n the restrictions $dN^{C_n} = 1_{C_n} dN$ are independent. Conversely suppose one has an increasing sequence of partitions $\mathcal{C}_n = C_{n1}, C_{n2}, \ldots$, and the point process N on $\mathcal{X}$ has the property that for each partition $\mathcal{C}_n$ the restrictions $N^{C_{n1}}, N^{C_{n2}}, \ldots$ are independent. Does it follow that N is a Poisson point process?

Example 2.20. Let $N = 2N_0$ where $N_0 \neq 0$ is a Poisson point process. Each point is doubled! Restrictions to disjoint sets are independent, but N is not Poisson. $\diamondsuit$

We now formulate a partial converse to the mapping theorem.

Theorem 2.21. *Let $\mathcal{X}$ and $\mathcal{Y}$ be separable metric spaces, and let $y : \mathcal{X} \to \mathcal{Y}$ be Borel measurable. Let N be a point process on $\mathcal{X}$ whose image $M = y(N)$ is a Poisson point process on $\mathcal{Y}$, with mean measure which is diffuse and σ-finite. For each finite collection of disjoint bounded Borel sets $B_1, \dots, B_m$ in $\mathcal{Y}$ let the restriction point processes $N^{E_1}, \dots, N^{E_m}$ with $E_i = y^{-1}(B_i)$ be independent. Then N is a Poisson point process.*

Proof. The point process N is a superposition of independent restrictions. One may use the Superposition Theorem to prove that N is a Poisson point process. We shall derive it from a more technical result, Proposition 2.25 below, for which we give an elementary proof. $\square$

Example 2.22. The result above is remarkable. Start with a point process N on $\mathbb{R}^d$ with points $Z_1, Z_2, \dots$. Suppose the vertical coordinates $Y_n = y(Z_n)$ form a simple Poisson point process on the vertical axis. If restrictions of N to disjoint horizontal slices are independent then the restrictions to disjoint Borel sets in $\mathbb{R}^d$ are independent (since N is Poisson by the theorem above).

In the case of the rain shower the assumption of independence for disjoint time intervals forces the shower to be a Poisson point process. Independence for disjoint horizontal slices implies independence for disjoint vertical cylinders! This also holds for a rain shower on the whole plane. Each horizontal slice then contains infinitely many points. The only condition is that no two drops hit the earth simultaneously, and at no moment is there a positive probability of a drop hitting the earth. $\diamondsuit$

The results below are a validation of the heuristic arguments used for the rain shower in Section 1.1. Recall that we used partitions of the unit square, and argued that one might assume that eventually all subsquares would contain at most one point. We also needed the probability of a rain drop in a subsquare to vanish. For the uniform rain shower on a tile we could write $p_n = c/n$, but we also considered point processes with varying intensity. Partitions are defined in Section 1.5.

Lemma 2.23. *Let M, with points $Y_1, \dots, Y_K$, be a finite simple point process on $\mathcal{Y}$ without fixed points:*

$$\mathbb{P}\{Y_i = Y_j\} = 0, \quad i \neq j; \quad \mathbb{P}\{Y_k = y\} = 0, \quad k \geq 1, \ y \in \mathcal{Y}. \tag{2.8}$$

Let $\mathcal{C}_n = (C_{nj})$ be increasing partitions, and suppose $\bigcup \mathcal{C}_n$ separates points. Set $K_{nj} = M(C_{nj})$ and $K_n = \max_j K_{nj}$. Then

$$\mathbb{P}\{K_n > 1\} \to 0, \quad n \to \infty,$$
$$\mathbb{P}\{K_{nj} > 0\} \to 0, \quad n + j \to \infty.$$

Proof. Let $M(\omega)$ have points $y_1, \dots, y_k$. There exists an index $J(\omega)$ such that these k points lie in distinct sets in the partition $\mathcal{C}_n$ for $n \geq J(\omega)$. We take $J(\omega)$ minimal. Then $\{K_n > 1\} = \{J > n\}$. Since J is finite the first limit relation holds.

Let $\varepsilon > 0$. We have to show that $\mathbb{P}\{K_{nj} > 0\} > 2\varepsilon$ occurs only finitely often. Choose k so large that $\mathbb{P}\{M(\mathcal{Y}) > k\} < \varepsilon$, and let $M^* = M$ on $\{M(\mathcal{Y}) \leq k\}$ and $M^* = 0$ elsewhere. So M^* has at most k points, say $Y_1, \dots, Y_K$ with $K \leq k$. Set $K^*_{nj} = M^*(C_{nj})$. It suffices to show that $\mathbb{P}\{K^*_{nj} > 0\} > \varepsilon$ holds only finitely often, or even that

$$\mathbb{P}\{Y_i \in C_{nj}\} > \varepsilon/k \tag{2.9}$$

holds finitely often for $i = 1, \dots, K$. We shall do this for Y_1.

For given n there are at most k/ε sets C_{nj} for which the inequality (2.9) holds with $i = 1$. Also note if $\mathbb{P}\{Y_1 \in C\} > \varepsilon/k$, then this inequality also holds for the mother set $C' \supset C$. Suppose $p_{nj} = \mathbb{P}\{Y_1 \in C_{nj}\} > \varepsilon/k$ is infinite for infinitely many C_{nj}. Then there is such a set $C_1 = C_{1j_1}$ with infinite offspring. One of the daughters $C_2 = C_{2j_2}$ also has infinite offspring. Proceeding thus we find a sequence of sets $C_n = C_{nj_n}$ with $p_{nj_n} > \varepsilon/k$ so that C_n is the mother of C_{n+1}. The events $E_n = \{Y_1 \in C_n\}$ decrease to E, and $\mathbb{P}(E) \geq \varepsilon/k$. Let $\omega_0 \in E$, and $y_0 = Y_1(\omega_0)$. Then $Y_1(\omega) = y_0$ for all $\omega \in E$, since $\bigcup \mathcal{C}_n$ separates points. Hence $\mathbb{P}\{Y_1 = y_0\} \geq \mathbb{P}(E) \geq \varepsilon/k$. This contradicts (2.8). □

We will need a simple inequality for sums of independent 0-1 variables.

Lemma 2.24. *Let $L_1, L_2, \dots$ be independent zero-one variables with sum $S \leq \infty$. Suppose $C \geq 4$. Then $\mathbb{P}\{S > C\} < 1/2$ implies $\mathbb{E}S < 2C$.*

Proof. Assume $\mathbb{E}S = 2C$, decreasing L_n if need be. The Bienaymé–Chebyshev inequality gives

$$\mathbb{P}\{S \leq C\} \leq \mathrm{var}(S)/C^2 \leq 2/C \leq 1/2,$$

since $\mathrm{var}(L_i) \leq \mathbb{E}L_i$ implies $\mathrm{var}(S) \leq \mathbb{E}S$. □

Proposition 2.25. *Let $\mathcal{X}$ and $\mathcal{Y}$ be separable metric spaces and $y\colon \mathcal{X} \to \mathcal{Y}$ a measurable map. Let N be a point process on $\mathcal{X}$ with points $X_1, X_2, \dots,$ which is finite on B_n, for an increasing sequence of bounded Borel sets B_n which cover $\mathcal{X}$. Suppose the points $Y_k = y(X_k)$ satisfy (2.8). Let $\mathcal{C}_n = (C_{nj}, j \geq 0)$ be increasing partitions of $\mathcal{Y}$ such that $\bigcup \mathcal{C}_n$ separates points. If for fixed n the restrictions*

$$N^{D_{nj}}, \quad j \geq 0; \quad D_{nj} = y^{-1}(C_{nj})$$

are independent, then N is a Poisson point process.

Proof. We may assume N is finite by restricting to B_n. Then use Proposition 2.4. Set

$$K_{nj} = M(C_{nj}) = N(D_{nj}), \quad K_n = \max_j K_{nj}.$$

By Lemma 2.23

$$q_n = \mathbb{P}\{K_n > 1\} \to 0, \quad n \to \infty,$$
$$p_{nj} = \mathbb{P}\{K_{nj} > 0\} \to 0, \quad n + j \to \infty.$$

Define 0-1 point processes: $N^*_{nj} = N^{D_{nj}}$ for $K_{nj} = 1$ and $N^*_{nj} = 0$ else. For each index n Lemma 2.13 allows us to choose independent Poisson point processes N^0_{nj} so that $\mathbb{P}\{N^0_{nj} \neq N^*_{nj}\} < p^2_{nj}$. Set

$$N^*_n = \sum_j N^*_{nj}, \quad N^0_n = \sum_j N^0_{nj}.$$

There is a constant $C \geq 4$ such that $\mathbb{P}\{N(\mathcal{X}) > C\} < 1/2$. Lemma 2.24 applied to $L_j = K_{nj} \wedge 1$ with sum $S \leq N(\mathcal{X})$ gives $\sum_j p_{nj} = \mathbb{E}S < 2C$. So

$$\sum_j p^2_{nj} \leq 2Cp_n, \quad p_n = \max_j p_{nj} \to 0,$$

and $\mathbb{P}\{N^0_n \neq N\} \leq 2Cp_n + q_n \to 0$. We may choose a subsequence $k_n \to \infty$ so that

$$\mathbb{P}\{N_n \neq N\} < 1/2^n, \quad n = 1, 2, \dots, \quad N_n = N^0_{k_n}.$$

Then $N_n \to N$ almost surely. The N_n are Poisson point processes on $\mathcal{X}$. Hence $N_n(E)$ is Poisson for any Borel set E in $\mathcal{X}$, and $N_n(E_1), \dots, N_n(E_m)$ are independent for disjoint Borel sets $E_1, \dots, E_m$. This then also holds for N. Hence N is a Poisson point process on $\mathcal{X}$. □

The next result formalizes the construction of Section 1.1 for simple point processes without fixed points. We do not need the rather artificial assumption of an intensity.

Proposition 2.26. *Let N be a point process on $\mathcal{Y}$ with points $Y_1, Y_2, \dots$. Let $\mathcal{C}_n = (C_{n1}, C_{n2}, \dots)$, $n \geq 0$, be an increasing sequence of partitions on $\mathcal{Y}$ which separate points. If (2.8) holds, if $N(C_{0j})$ is a.s. finite for all j, and if the j random variables $N(C_{n1}), \dots, N(C_{nj})$ are independent for each n and j, then N is a Poisson point process.*

Proof. We may assume N finite by restricting to an atom $A \in \bigcup \mathcal{C}_n$. With C_{nk} associate three random integers, $N_{nk} = N(C_{nk})$, $L_{nk} = 1_{\{N_{nk}=1\}}$, and M_{nk}, a Poisson rv with expectation $p_{nk} = \mathbb{P}\{N_{nk} = 1\}$, such that $\mathbb{P}\{M_{nk} \neq L_{nk}\} \leq p^2_{nk}$, see Lemma 2.13. For fixed n one may choose the Poisson variables M_{nk} independent, since the N_{nk} are independent by assumption, and hence the L_{nk} too. Since $S_n = \sum_k L_{nk} \leq N(\mathcal{X})$ and $\mathbb{P}\{N(\mathcal{X}) > C\} < 1/2$ for some $C \geq 4$, Lemma 2.24 gives $\mathbb{E}S_n < 2C$. Moreover $p_{nk} \to 0$ for $n + k \to \infty$ by Lemma 2.23. Hence $\sum_k p_{nk}^2 \to 0$, and $N(\mathcal{X}) = N(A)$ is Poisson. This holds for each atom A. Together with the assumed independence this makes N into a Poisson point process; see relation 4) in Theorem 3.2. □

2.10* Marked point processes. Given a point process N on the separable metric space $\mathcal{X}$ with points $X_1, X_2, \dots$, one may add a random mark U_n to each point X_n. In first instance we assume that (U_n) is an iid sequence of uniformly distributed random variables in $\mathrm{U} = [0, 1]$, independent of $\mathcal{F}$, where $\mathcal{F}$ is a σ-field on the underlying probability space on which the point process N and the points X_n are measurable. One obtains a new point process M with points (X_n, U_n). This construction does not depend on the enumeration. If X'_n also is an enumeration of the points, say $X'_n = X_{I_n}$, then the corresponding marks $U'_n = U_{I_n}$ again form an iid sequence of uniform-(0, 1) distributed random variables, at least if the renumeration is $\mathcal{F}$-measurable. Indeed, conditional on $I_1 = i_1, \dots, I_n = i_n$ the random variables $U'_1, \dots, U'_n$ are distributed like $U_{i_1}, \dots, U_{i_n}$. If $K = N(\mathcal{X})$ is finite with positive probability, condition on $K = k$ and restrict to $n \leq k$.

Example 2.27. Consider a point process on $[0, \infty)^2$ concentrated around the diagonal, where the points first are enumerated according to their vertical coordinate, and later according to their horizontal coordinate. The renumeration sequence (I_n) is measurable on the σ-field generated by N. We assume some regularity conditions here: the projections on the two axes should yield simple point processes with finitely many points on bounded intervals. ◊

Let M be a point process on $\mathcal{Y}$ with points $Y_1, Y_2, \dots$. For each of these points toss a coin to decide whether the point is retained or deleted. The tosses are independent. The resulting point process M^0 is called the *thinned point process*. One may think of a forest in which certain trees are felled. In general the probability for retaining the tree may depend on the site. So we have a function $p \colon \mathcal{Y} \to [0, 1]$ which determines the probability of heads for the coin which is tossed to decide whether a tree at site y is to remain. Such a thinned point process M^0 is simple to construct. First construct the marked point process N on $\mathcal{X} = \mathcal{Y} \times [0, 1]$ with points (Y_k, U_k). Restrict to the set below the graph of p, and project onto $\mathcal{Y}$.

Instead of marks in $[0, 1]$ one may have marks in a separable metric space $\mathcal{M}$. Let $g \colon \mathcal{Y} \times [0, 1] \to \mathcal{M}$ be measurable. The map $f : (y, u) \mapsto (y, g(y, u))$ maps the point process N on $\mathcal{Y} \times [0, 1]$ above, into a point process on $\mathcal{Y} \times \mathcal{M}$, with points $(Y_n, g(Y_n, U_n))$. If N on $\mathcal{Y} \times [0, 1]$ is Poisson with mean measure ν, the new point process is Poisson with mean measure $f(\nu)$. In the thinning above $\mathcal{M} = \{0, 1\}$; a tree is felled if its mark is 0, and $g(y, u) = 1$ for $u < p(y)$.

Proposition 2.28. *Let M be a Poisson point process on $\mathcal{Y}$ with mean measure μ and points $Y_1, Y_2, \dots$. Let $\mathcal{F}$ be a sigma-field on the underlying probability space* *on which the point process M and the variables $Y_1, Y_2, \dots$ are measurable. Let $U_1, U_2, \dots$ be independent uniformly distributed random variables. Assume the sequence (U_n) is independent of $\mathcal{F}$. The point process N on $\mathcal{Y} \times [0,1]$ with points $(Y_1, U_1), (Y_2, U_2), \dots$ is a Poisson point process with mean measure $d\mu(y)du$.*

Proof. This may seem a good opportunity to apply Proposition 2.25. However it is simpler to observe that in the case of a finite mean measure μ one may construct a Poisson point process N^* on $\mathcal{Y} \times [0,1]$ with mean measure $\mu \times \lambda$ as a Poisson mixture of sample clouds from the probability distribution $\pi \times \lambda$ where $\pi = \mu/\mu(\mathcal{Y})$. This is clearly a marked point process. Since the distribution does not depend on the enumeration, the processes N and N^* have the same distribution. For infinite σ-finite mean measures, use Proposition 2.4. □

Marked point processes may be used to compare Poisson point processes on $\mathcal{Y}$ whose mean measures μ_i are close. Let $d\mu_i(dy) = p_i(y)d\mu(y)$ for $i = 1, 2$, where $d\mu = d\mu_1 \vee d\mu_2$ (and hence $p_1 \vee p_2 = 1$). Define a Poisson point process N on $\mathcal{Y} \times [0,1]$ with mean measure $d\mu(y)du$. Let M_i be the projection on $\mathcal{Y}$ of the restriction of N to the area below the graph of p_i. Then

$$\mathbb{P}\{M_1 \neq M_2\} \leq \int |p_2 - p_1| d\mu = \|\mu_2 - \mu_1\|_1. \tag{2.10}$$

3 The distribution

3.1 Introduction. To specify the distribution of a random vector $X \in \mathbb{R}^d$ it does not suffice to give the distributions of the d components. One has to introduce a probability measure π on the Borel sigma-field of $\mathbb{R}^d$. Often the distribution may be described in terms of a density function. One may always use the d-dimensional distribution function F to describe the distribution of X, but in the multivariate setting the df is not very informative about the shape of sample clouds.

There are alternative ways to specify the distribution. One may give the characteristic function $\varphi(\xi) = \mathbb{E}e^{i\xi X}$, defined for $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$. Equivalently one may specify the probability $\pi(H)$ for all closed halfspaces $H = \{\xi \geq c\}$. For vectors X with thin tails, the moment generating function (mgf) $M(\xi) = \mathbb{E}e^{\xi X}$ may exist on a neighbourhood of the origin. The mgf is an analytic function. Its restriction to any non-empty open set in $\mathbb{R}^d$ determines the distribution. If the vector has non-negative components, the *Laplace transform* exists:

$$L(\xi) = \mathbb{E}e^{-\xi X}, \quad \xi \geq 0;$$

and if the components are non-negative integers, then one may describe the distribution in terms of the probability generating function. These transforms are related.

That the probability distribution is uniquely determined by each of these transforms, is not trivial.

3.2* The Laplace transform. For a point process N on a separable metric space $\mathcal{X}$ the *Laplace transform* is defined by

$$\mathcal{L}(f) = \mathbb{E}(e^{-\int f dN}), \quad f \geq 0,$$

for non-negative Borel functions f on $\mathcal{X}$. The integral $\int f dN$ may be infinite for certain values of ω. We set $e^{-\infty} = 0$. The larger the integral $\int f dN$, the smaller the Laplace transform $\mathcal{L}(f)$.

Let us compute the Laplace transform of the Poisson point process N with mean measure μ. Start with a step function,

$$f = c_1 1_{B_1} + \cdots + c_m 1_{B_m}, \quad B_i \in \mathcal{B} \text{ disjoint}, \ \mu(B_i) < \infty, \ c_i > 0.$$

Then $\int f dN = c_1 N(B_1) + \cdots + c_m N(B_m)$. This stochastic integral is a sum of m independent terms $X_i = c_i N(B_i)$, where $N(B_i)$ has a Poisson distribution with expectation $b_i = \mu B_i$. So let us first compute the Laplace transform of $\int g dN$ with $g = c 1_E$, and $\mu(E) = b < \infty$:

$$\begin{aligned} \mathcal{L}(g) = \mathbb{E}(e^{-\int g dN}) = \mathbb{E}e^{-cN(E)} &= \sum_n e^{-cn} p_n = e^{-b} \sum_n (be^{-c})^n / n! \\ &= e^{-(1-e^{-c})\mu(E)} = e^{-\int 1 - e^{-g} d\mu}, \quad p_n = e^{-b} b^n / n!. \end{aligned}$$

By independence $\mathcal{L}(f)$ is the product of $\mathcal{L}(f_i)$, with $f_i = c_i 1_{B_i}$. The terms in the exponent add, and

$$(1 - e^{-c_1})\mu B_1 + \cdots + (1 - e^{-c_m})\mu B_m = \int (1 - e^{-f}) d\mu$$

gives

$$\mathcal{L}(f) = e^{-\int (1-e^{-f}) d\mu}. \tag{3.1}$$

Theorem 3.1. *The Laplace transform of the Poisson point process with σ-finite mean measure μ is given by* (3.1) *for Borel functions $f : \mathcal{X} \to [0, \infty)$.*

Proof. Any Borel function $f \geq 0$ is pointwise limit of an increasing sequence of integrable step functions $f_n \geq 0$. The monotone convergence theorem applied ω-wise gives

$$\int f_n dN \to \int f dN,$$

and Lebesgue's theorem with majorant 1 gives

$$\mathcal{L}(f_n) = \mathbb{E}e^{-\int f_n dN} \to \mathbb{E}e^{-\int f dN} = \mathcal{L}(f).$$

On the right $\int (1 - e^{-f_n}) d\mu \to \int (1 - e^{-f}) d\mu$ by monotone convergence. □

3.3 The distribution. We shall now give a number of equivalent characterizations of the distribution of a point process N on a separable metric space $\mathcal{X}$. Since we want to develop the theory for separable metric spaces here, we assume in this subsection:

$$N(E) \text{ is finite for bounded Borel sets } E. \tag{3.2}$$

Let $\mathcal{C}_n = \{C_{n0}, C_{n1}, \dots\}$ be increasing partitions on $\mathcal{X}$ for which the small set condition (SS) holds, see (1.2). Let (U_n) be bounded open sets which form a basis for the open sets of $\mathcal{X}$, for instance the set of all open balls $B^r(a)$, with center a in some countable dense set Aß$\mathcal{X}$, and radius $r \in \{1, 1/2, 1/3, \dots\}$. Each open set Oß$\mathcal{X}$ is a union of such sets U_n. Let N and M be finite point processes on the separable metric space $\mathcal{X}$, and let $B_1, \dots, B_m$ be Borel sets, and f a Borel function on $\mathcal{X}$. Consider the equalities in distribution

$$(N(B_1), \dots, N(B_m)) \stackrel{\mathrm{d}}{=} (M(B_1), \dots, M(B_m)) \tag{3.3}$$

$$\int f dN \stackrel{\mathrm{d}}{=} \int f dM. \tag{3.4}$$

Theorem 3.2. *Let N and M be point processes on the separable metric space $\mathcal{X}$ which satisfy* (3.2). *With the notation above the following are equivalent:*

1) *The first relation holds for bounded Borel sets.*

2) *The first relation holds for disjoint bounded closed sets.*

3) *The first relation holds for disjoint sets which are finite unions of sets U_n in a countable base of bounded open sets $U_1, U_2, \dots$.*

4) *The first relation holds for bounded atoms $B_i \in \mathcal{C}_m$, $m \geq 1$.*

5) *The second relation holds for bounded Borel functions which vanish outside a bounded set.*

6) *The second relation holds for non-negative Borel step functions with bounded steps.*

7) *The second relation holds for non-negative uniformly continuous functions with bounded support.*

8) *The second relation holds for non-negative step functions $f = c_1 1_{C_1} + \dots + c_m 1_{C_m}$ with C_i bounded atoms in $\mathcal{C}_n$ for $n \geq 1$.*

Proof. First assume N and M are finite point processes, and $\mathcal{X}$ is bounded. Note that the first four criteria refer to finite-dimensional distributions of the process $N(B)$, $B \in \mathcal{B}$, the second four to stochastic integrals. There is a simple relation: The multivariate distribution of $(N(B_1), \dots, N(B_m))$ is determined by the univariate distributions of the random variables $Y = c_1 N(B_1) + \dots + c_m N(B_m)$, where we may restrict the c_i to be non-negative. (Indeed it suffices to know the distribution of one such linear combination Y provided the coefficients c_i are rationally independent.) In particular, if (3.3) holds for all non-negative step functions f for a certain class of steps, and the integrals are finite, it holds for all real valued step functions on this class of steps. Hence

$$1) \iff 6) \quad \text{and} \quad 4) \iff 7).$$

The first and fifth conditions are strong, the remaining ones are weak:

$$1) \Rightarrow 2), 3), 4), \quad \text{and} \quad 5) \Rightarrow 6), 7), 8).$$

Some implications are easy:

6) $\Rightarrow$ 5). First note that 1) is equivalent to 1′) where in 1′) we assume the first relation to hold for disjoint bounded Borel sets B_i. Moreover 6) is equivalent to 6′), where in 6′) we assume (3.3) to hold for all real-valued step functions on bounded Borel sets. Now observe that any Borel function is pointwise limit of a sequence of step functions on disjoint bounded Borel sets.

3) $\Rightarrow$ 2) If $F_1, \dots, F_m$ are disjoint closed sets one may find decreasing sequences of open sets $O_{ni} \downarrow F_i$ with $O_{11}, \dots, O_{1m}$ disjoint. Hence $N(O_{ni}) \downarrow N(F_i)$ for $n \to \infty$, $i = 1, \dots, m$, and 3) $\Rightarrow$ 2).

2) $\Rightarrow$ 1) If relation 1) holds for $m = 1$, then already M and N have the same mean measure, and determine the same null sets. With N one may associate a finite measure ν^* which has the same null sets as the mean measure, and similarly for M, see Exercise 1.4. For any Borel set E there exists an increasing sequence of bounded closed sets F_nßE such that $(\nu^* + \mu^*)(E \setminus F_n) \to 0$. Then $N(F_n) \uparrow N(E)$ almost surely, and similarly $M(F_n) \uparrow M(E)$.

7) $\Rightarrow$ 3) If the B_i are disjoint bounded open sets and $c_i > 0$ there is a sequence of uniformly continuous functions $f_n \uparrow f = c_1 1_{B_1} + \cdots + c_m 1_{B_m}$.

So 8), 3), 2), 1), 5) and 6) are equivalent, as are 4) and 7). It remains to observe that step-functions on bounded atoms in $\mathcal{C}$ are dense in $\mathbf{L}^1(\mathcal{X}, \mathcal{B}, \rho)$ where $\rho = \mu^* + \nu^*$, see above.

What if $\mathcal{X}$ is not bounded, and N and M not finite? Cover $\mathcal{X}$ by an increasing sequence of open balls. The eight conditions are formulated so that they may be checked for the restrictions of the point processes to a sufficiently large ball. On such a ball the point processes are finite. □

Definition. Two point processes N and M on the separable metric space $\mathcal{X}$ which satisfy (3.2) have the same *distribution*, and we write $N \stackrel{d}{=} M$, if

$$(N(B_1), \dots, N(B_m)) \stackrel{d}{=} (M(B_1), \dots, M(B_m))$$

holds for any finite sequence of disjoint bounded Borel sets $B_1, \dots, B_m$.

By the theorem above we have at least seven other equivalent definitions for equality in distribution of two point processes.

Exercise 3.3. Let N be a point process on the separable metric space $\mathcal{X}$ with mean measure μ which is finite on bounded sets. Suppose $\mathcal{L}(f) = e^{-\int 1-e^{-f} d\mu}$ for non-negative bounded continuous functions f on $\mathcal{X}$ for which $\{f > 0\}$ is bounded. Show that N is a Poisson point process. $\diamondsuit$

Let $\mathcal{N} = \mathcal{N}(\mathcal{X})$ be the set of all counting measures ξ on $\mathcal{X}$ which are finite for open balls. On the set $\mathcal{N}$, define the sigma-field $\mathcal{E}$ generated by the sets

$$E^k := \{\xi \in \mathcal{N} \mid \xi(E) = k\}, \quad k \geq 0, \ E \in \mathcal{B}_b,$$

where $\mathcal{B}_b$ denotes the class of bounded Borel sets in $\mathcal{X}$. If two probability measures agree on the class of finite intersections of such sets E^k in $\mathcal{N}$, then they agree on the sigma-field $\mathcal{E}$. A point process N on $\mathcal{X}$ (with underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$) is a measurable map from $(\Omega, \mathcal{A})$ to $(\mathcal{N}, \mathcal{E})$. Recall that we assume point processes to be finite on bounded sets. The *distribution* of the point process N is the image (under N) of the probability measure $\mathbb{P}$. Just as the distribution of a vector X is the image $X(\mathbb{P})$.

Exercise 3.4. Let N be a point process on $\mathcal{X}$ and E a Borel set. Prove that the restriction N^E is a point process on $\mathcal{X}$. $\diamondsuit$

Exercise 3.5. Suppose $\mathcal{X}$ is the union of an increasing sequence of Borel sets B_n. Let N and M be point processes on $\mathcal{X}$. If the restrictions N^{B_n} and M^{B_n} have the same distribution for each n, then N and M have the same distribution. $\diamondsuit$

In Section 2 we defined the distribution of a finite point process in terms of exchangeable k-tuples by conditioning on the event $\{N(\mathcal{X}) = k\}$. This is a useful characterization for constructing point processes. The analytic characterizations above are effective for checking that two point processes have the same distribution.

3.4* The distribution of simple point processes. In this subsection point processes are assumed finite on bounded sets. A point process N on a finite space $\mathcal{X} = \{1, \dots, d\}$ is just a random vector X with non-negative integer components. Its distribution is given by $\mathbb{P}\{N\{1\} = k_1, \dots, N\{d\} = k_d\}$. For simple point processes less information suffices.

Proposition 3.6. *Let N be a simple point process on the separable metric space $\mathcal{X}$. Let μ be a diffuse measure on $\mathcal{X}$ which is finite on open balls. Suppose*

$$P\{N(E) = 0\} = e^{-\mu(E)}, \quad E \in \mathcal{B} \text{ bounded}.$$

Then N is a Poisson point process with mean measure μ.

Recall that a point process with points $X_1, X_2, \ldots$ is *simple* if $\mathbb{P}\{X_i = X_j\} = 0$ for $i \neq j$. We shall show that for simple point processes the probabilities

$$p_E = P\{N(E) = 0\}, \quad E \in \mathcal{B} \text{ bounded}, \tag{3.5}$$

determine the distribution of N.

Lemma 3.7. *Let $B_n \uparrow B$. Then $p_{B_n} \downarrow p_B$.*

Proof. $\{N(B_n) = 0\} \downarrow \{N(B) = 0\}$. □

First consider a point process N on a countable set. We do not assume N is simple.

For any finite set FßE, write p_E^F for the probability that each $x \in F$ has positive mass, and the remaining points of E have no mass. If $F = \{x\}$ we write p_E^x, and if F is empty then $p_E^F =: p_E$. Observe that

$$p_E = p_{E\cup\{x\}}^x + p_{E\cup\{x\}}, \quad x \notin E,$$

since the event of no points in E is divided into two disjoint events according to whether x has positive mass or no mass. Similarly, for FßE and $x \notin E$,

$$p_E^F = p_{E\cup\{x\}}^{F\cup\{x\}} + p_{E\cup\{x\}}^F.$$

So if we know p_E^F for all n-point sets F and all $E \supset F$, we also know p_E^F for all $n+1$-point sets F and $E \supset F$.

Let $\mathcal{R}$ be the ring of sets generated by a countable partition of $\mathcal{X}$, and $\mathcal{S}$ the sigma-ring. The sets C of the partition will be called atoms. As above for F in $\mathcal{R}$ and Fß$E \in \mathcal{S}$ define p_E^F as the probability that $E \setminus F$ contains no points and each of the atoms C making up F contains at least one point. If we know p_E for all $E \in \mathcal{R}$, then we know p_E for all $E \in \mathcal{S}$, and by the argument above we know p_E^F for all $E \in \mathcal{S}$ and all $F \in \mathcal{R}$, FßE. We define p_E^n as the probability that E contains precisely n atoms of positive mass. This is the sum of the probabilities p_E^F with FßE containing precisely n atoms.

Now suppose we have an increasing sequence of partitions $\mathcal{C}_n = (C_{nj})$ which separates points. Let $p_E^m(n)$ be the probabilities associated with the nth partition. Let $E \in \mathcal{S}_{n_0}$. Then E is a countable union of atoms in $\mathcal{C}_{n_0}$ and hence $E \in \mathcal{S}_n$ for all $n \geq n_0$. For $n \geq n_0$ let $A_E^m(n)$ be the event that E contains m atoms $C \in \mathcal{C}_n$ of positive mass. Then $A_E^m(n) \to \{N(E) = m\}$ for $n \to \infty$ since $(\mathcal{C}_n)$ separates points. Hence

$$p_E^m(n) \to p_E^m = \mathbb{P}\{N(E) = m\}.$$

Therefore one may derive the probabilities $\mathbb{P}\{N(E) = m\}$ from the probabilities $\mathbb{P}\{N(E) = 0\}$.

Similarly one may determine the probability $\mathbb{P}\{N(E_1) = m_1, \dots, N(E_k) = m_k\}$ from the probabilities of the events

$$\{N(E_1) = 0, \dots, N(E_k) = 0\} = \{N(E) = 0\}, \quad E = E_1 \cup \dots \cup E_k$$

for given $k \geq 2$ for all collections of disjoint $E_1, \dots, E_k$ in $\bigcup \mathcal{C}_n$. (For this one has to check that $p^{F_1,\dots,F_k \cup \{x\}}_{E_1,\dots,E_k}$ may be computed as above for disjoint subsets $E_1, \dots, E_k$ of a countable set and for points $x \notin E = E_1 \cup \dots \cup E_k$.) We have shown:

Theorem 3.8. *Suppose N and M are simple point processes on $\mathcal{X}$ which satisfy* (3.2). *Let $\mathcal{C}_n$ be an increasing sequence of partitions which separates points. Let $\mathcal{R}$ be the ring generated by $\bigcup \mathcal{C}_n$. If*

$$P\{N(R) = 0\} = P\{M(R) = 0\}, \quad R \in \mathcal{R},$$

then N and M have the same distribution. □

4 Convergence

4.1 Introduction. Recall the usual criteria for *convergence in distribution* for random vectors in $\mathbb{R}^d$, which is written as $X_n \Rightarrow X_0$. In abstract arguments one uses the definition

$$\mathbb{E}\varphi(X_n) \to \mathbb{E}\varphi(X_0)$$

for all bounded continuous functions φ on $\mathbb{R}^d$ (or all continuous functions φ with compact support, or all complex functions $\varphi_\xi : x \mapsto e^{i\xi x}$, or all indicator functions $\varphi = 1_Q$ where $Q = (-\infty, c]$ is the intersection of closed lower halfspaces $\{x_i \leq c_i\}$, $i = 1, \dots, d$, whose boundary planes $\{x_i = c_i\}$ are not charged by the limit distribution). Often it is convenient to use coordinates: $X_n \Rightarrow X_0$ if and only if all linear combinations of the coordinates converge:

$$c_1 X_n^{(1)} + \dots + c_d X_n^{(d)} \Rightarrow c_1 X_0^{(1)} + \dots + c_d X_0^{(d)}.$$

A family of random vectors $X(t)$ (or their probability distributions) on $\mathbb{R}^d$ is *tight* if for any $\varepsilon > 0$ there exists a cube $K ß \mathbb{R}^d$ such that $P\{X(t) \in K\} > 1 - \varepsilon$ for all t. For this it suffices that the d components $X^{(i)}(t)$, $t \in T$, are tight. In a tight family $X(t), t \in T$, each sequence (t_n) contains a subsequence (t_{k_n}) such that $X(t_{k_n})$ converges in distribution.

In this section similar concepts will be developed for point processes.

4.2 The state space. Henceforth the state space of our point process, a separable metric space $\mathcal{X}$, is assumed to be locally compact, and the sample functions are assumed to be integer-valued Radon measures. The theory will be applied to point processes on $\mathbb{R}^d$, or on open subsets of $\mathbb{R}^d$. Occasionally on open subsets of $[0, \infty)^d$.

Weak convergence of probability measures is a matter of topology rather than metric. If the separable metric space $\mathcal{X}$ is a subset of a compact metric space $\mathcal{Y}$ (which is automatically separable), then it is simple to determine whether a sequence of probability measures π_n on $\mathcal{X}$ converges. Check whether the sequence of reals $\int \varphi d\pi_n$ converges in $\mathbb{R}$ for each continuous function φ on $\mathcal{Y}$. If so, there is a probability measure μ on $\mathcal{Y}$ such that $\pi_n \to \mu$ weakly on $\mathcal{Y}$. Now check whether μ lives on $\mathcal{X}$. If so, then $\pi_n \to \mu$ weakly on $\mathcal{X}$; if not, the sequence π_n does not converge on $\mathcal{X}$. For compact metric spaces $\mathcal{Y}$ the theory of weak convergence is particularly elegant since the space $\mathcal{P}(\mathcal{Y})$ of probability measures on $\mathcal{Y}$ is itself a compact metrizable space.

Similarly there is an elegant theory for vague convergence of Radon measures on locally compact separable metric spaces $\mathcal{Z}$. Here too, it is the topology rather than the metric which is important. So one should rather speak of *lcscH spaces*: spaces which are locally compact, second countable Hausdorff. These are topological spaces with a countable base of open sets U_n, such that each pair of distinct points $x, y \in \mathcal{Z}$ has disjoint neighbourhoods U_i and U_j. The sets U_n may be chosen to have compact closures. These conditions make it possible to introduce a metric on the space. One may choose the metric d so that the compact sets are precisely the bounded closed sets. In this metric, for any $a \in \mathcal{Z}$, the open balls $B^n(a) = \{x \in \mathcal{Z} \mid d(x, a) < n\}$ cover $\mathcal{Z}$, and are relatively compact. Radon measures on $\mathcal{Z}$ are precisely the measures which are finite on these balls.

Actually it is quite simple to construct a metric on a locally compact separable metric space such that bounded closed sets are compact. All one needs is a continuous positive function χ on the space, with the property that the sets $\{\chi \leq c\}$ are compact for $c > 0$. On an open proper subset O of $\mathbb{R}^d$, for each point $x \in O$ there exists a maximal $r = r(x) > 0$ such that $B^r(x)$, the open ball of radius r around x, still lies in O. The function r measures the distance to the boundary. It is easy to see that this function is positive on O, and continuous. The function $\chi(x) = \|x\| + 1/r(x) > 0$ has the desired property: it is continuous, and the sets $\{\chi \leq c\}$ are compact. Now define $d(x, y) = \|x-y\| + |\chi(x)-\chi(y)|$. This is a metric on O: it is the natural metric on the graph of χ. It agrees with the topology: $x_n \to x_0$ implies $d(x_n, x_0) \to 0$, and conversely. Any closed ball $\{x \mid d(x, a) \leq n\}$ is compact, since it is contained in the set $\{\chi(x) \leq \chi(a) + n\}$.

In most of our applications $\mathcal{Z} = O$ is an open set in $\mathbb{R}^d$. So $\mathcal{Z}$ is a locally compact separable metric space in the Euclidean norm. Unfortunately in the Euclidean norm bounded subsets of O need not be relatively compact, unless $\mathcal{Z} = \mathbb{R}^d$. Typically the points of the realization $N(\omega)$ of a point process N on O will cluster at the boundary.

If O is the open unit ball B, then a point process N with the property that the closure of $N(\omega)$ contains the unit sphere ∂B a.s. is not exceptional!

For point processes on open subsets of $\mathbb{R}^d$, the Euclidean metric is confusing! For a compact space it makes no difference what metric one uses; for a locally compact space is does. One may replace the Euclidean metric by a metric d in which bounded closed sets are compact, as we saw above, but it is simpler to forget about the metric altogether and use only topological concepts like convergent, closed, open, compact and continuous.

It is well-known that any separable metric space $\mathcal{X}$ may be embedded in a compact metric space $\mathcal{Y}$. One may choose $\mathcal{Y}$ to be the Hilbert cube $[0,1]^\infty$, with the embedding $x \mapsto (y_1, y_2, \dots)$ where $y_n = d(x, a_n) \wedge 1$ for a dense sequence a_n in $\mathcal{X}$. Add the coordinate $y_0 = d(x, a_0)$ where a_0 is any point in $\mathcal{X}$ to embed $\mathcal{X}$ in the lcscH space $\mathcal{Z} = [0, \infty) \times \mathcal{Y}$. A set Eß$\mathcal{X}$ is bounded in $\mathcal{X}$ if and only if it is relatively compact in $\mathcal{Z}$. If $(\mathcal{X}, d)$ is complete, the image in $\mathcal{Z}$ is the intersection of a decreasing sequence of open subsets of $\mathcal{Z}$. This follows from a characterization of Polish spaces.

Theorem 4.1. *For subsets E of a compact metric space $\mathcal{Y}$ the following are equivalent:*

1) *E is Polish;*

2) *there exists a metric d on E such that (E, d) is complete;*

3) *E is the intersection of a decreasing sequence of open sets in $\mathcal{Y}$.*

Proof. See Kuratowski [1948] I, p. 337. □

In an abstract sense the theory for point processes on separable metric spaces which are finite on bounded sets is the same as the theory for point processes on lcscH spaces which are finite on compact sets. As far as convergence is concerned there is a crucial difference. On an lcscH space it suffices to check that $\int \varphi dN_n$ converges in distribution for each continuous function φ with compact support. This ensures that there is a limit process. For separable metric spaces one has to prove existence separately, for instance by showing that the distributions of the point processes are tight.

Example 4.2. The space $\mathcal{X} = \mathbb{R} \times (0, \infty)$ with the Euclidean metric is a subset of the lcscH space $\mathcal{Z} = \mathbb{R} \times [0, \infty)$. Bounded sets in $\mathcal{X}$ are relatively compact in $\mathcal{Z}$. Let N_1 be the Poisson point process on $\mathcal{X}$ with mean measure $e^{-y}dxdy$ and points (X_k, Y_k). The point process N_n with points $(X_k, Y_k/n)$ is a Poisson point process with mean measure $ne^{-ny}dxdy$ on $\mathcal{X}$. On $\mathcal{Z}$ there is an a.s. limit process N, with points $(X_k, 0)$. This is the standard Poisson point process on the horizontal axis. On $\mathcal{X}$ the sequence N_n does not converge a.s. Does it converge in distribution? Is there a point process M on $\mathcal{X}$ such that $\int \varphi dN_n \Rightarrow \int \varphi dM$ for uniformly continuous functions φ on $\mathcal{X}$ with bounded support?

Such a function φ extends to a continuous function $\bar{\varphi}$ on $\mathcal{Z}$ with compact support, and

$$\int \varphi dN_n \Rightarrow \int \bar{\varphi} dN = \sum_k \bar{\varphi}(X_k, 0).$$

If there is a limit point process M on $\mathcal{X}$, then $\int \varphi dM = \int \bar{\varphi} dN$. Hence M and N have the same distribution on $\mathcal{Z}$. In particular M is a Poisson point process whose mean measure μ satisfies $\mu[a,b] \times (0, 1/n] = b - a$ for $b > a$, $n \geq 1$. Such a measure does not exist on $\mathcal{X}$. $\Diamond$

Exercise 4.3. In the example above $\mathcal{X}$ itself is lcscH. Show that there is a point process M on $\mathcal{X}$ such that $\int \varphi dN_n \Rightarrow \int \varphi dM$ for all continuous functions $\varphi \colon \mathcal{X} \to \mathbb{R}$ with compact support. $\Diamond$

In order to avoid such paradoxes – N_n converges if we regard the open upper halfplane as an lcscH space, but fails to converge if we regard it as a separable metric space in the Euclidean metric – we henceforth regard point processes as random integer valued Radon measures on lcscH spaces, with convergence defined in terms of continuous functions with compact support.

4.3 Weak convergence of probability measures on metric spaces. In this subsection we recall some results about convergence in distribution of random elements in a separable metric space. A number of basic results are given without proof. In addition we state and prove a few more technical results which will be needed later.

Let $\mathcal{P} = \mathcal{P}(\mathcal{X})$ denote the set of probability distributions on the separable metric space $\mathcal{X}$ with the weakest topology which makes the maps

$$\pi \mapsto \int \varphi d\pi$$

continuous for each bounded continuous function $\varphi \colon \mathcal{X} \to \mathbb{R}$.

Theorem 4.4. *Let $\mathcal{X}$ be a separable metric space. Then $\mathcal{P}$ is metrizable. If $\mathcal{X}$ is compact then so is $\mathcal{P}$. If $\mathcal{X}$ is locally compact or Polish then $\mathcal{P}$ is Polish.*

It is useful to have a number of equivalent criteria for weak convergence. The first three conditions below are equivalent by definition. The implications 5)$\Rightarrow$3) and 5)$\Rightarrow$6) are obvious. (The function 1_O is π_0-a.s. continuous precisely if $\pi_0(\partial O) = 0$.) Proposition 4.11 gives 4)$\Rightarrow$3).

Theorem and Definition 4.5. *Let $\mathcal{X}$ be a separable metric space. Let X_n be random elements in $\mathcal{X}$ with distribution π_n for $n \geq 0$. The following are equivalent:*

1) $\pi_n \to \pi_0$ *weakly;*

2) $X_n \Rightarrow X_0$*;*

3) $\mathbb{E}\varphi(X_n) \to \mathbb{E}\varphi(X_0)$ *for every bounded continuous function φ on $\mathcal{X}$;*

4) $\mathbb{E}\varphi(X_n) \to \mathbb{E}\varphi(X_0)$ *for bounded uniformly continuous functions φ;*

5) $\mathbb{E}\varphi(X_n) \to \mathbb{E}\varphi(X_0)$ *for every bounded Borel function φ which is π_0-a. s. continuous;*

6) $\pi_n(O) \to \pi_0(O)$ *for open sets O with $\pi_0(\partial O) = 0$;*

7) *If $\pi_0(O) > c$ for an open set O and $c > 0$, then $\pi_n(O) > c$ eventually.*

Definition. A family of random elements $X_t, t \in T$ (or their distributions $\pi_t, t \in T$) is *tight* if for each $\varepsilon > 0$ there exists a compact set Kß$\mathcal{X}$ such that $\pi_t(\mathcal{X} \setminus K) < \varepsilon$ for all $t \in T$.

Theorem 4.6 (Prohorov). *A tight family of probability measures is relatively compact. If the underlying space is complete, then relatively compact sets in $\mathcal{P}$ are tight.*

Theorem 4.7 (Continuous Mapping Theorem). *Let X_n, $n \geq 0$, be random elements of the separable metric space $\mathcal{X}$, and $f : \mathcal{X} \to \mathcal{Y}$ a continuous map into the separable metric space $\mathcal{Y}$. If $X_n \Rightarrow X_0$, then $f(X_n) \Rightarrow f(X_0)$. The same result holds if f is continuous X_0-a.e. on $\mathcal{X}$.*

Proof. Let $\varphi : \mathcal{Y} \to \mathbb{R}$ be bounded and continuous. Then so is $\psi = \varphi \circ f : \mathcal{X} \to \mathbb{R}$ if f is continuous. If $X_n \Rightarrow X_0$, then $\mathbb{E}\varphi(f(X_n)) = \mathbb{E}\psi(X_n) \to \mathbb{E}\psi(X_0) = \mathbb{E}\varphi(f(X_0))$. If f is X_0-a.e. continuous we use criterium 4) in Theorem 4.5. □

Both almost-sure convergence and convergence in probability imply convergence in distribution. For complete separable metric spaces there is a converse:

Theorem 4.8 (Skorohod's Representation Theorem). *Let $\mathcal{X}$ be a complete separable metric space (or a Borel set of such a space) and let $X_n \Rightarrow X_0$. There exist random elements X'_n, $n \geq 0$, in $\mathcal{X}$, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that X'_n is distributed like X_n for $n \geq 0$, and such that $X'_n \to X'_0$ almost surely.*

Proof (Outline). For $\mathcal{X} = \mathbb{R}$ the result is simple. Let X_n have df F_n with inverse function $\varphi_n = F_n^{\leftarrow} : (0,1) \to \mathbb{R}$. Choose Ω to be the unit interval $(0,1)$ with Lebesgue measure λ on the Borel sets. Then weak convergence, $F_n \to F_0$, is equivalent to weak convergence of the inverse functions, $\varphi_n \to \varphi_0$, and equivalent to almost-sure convergence, $X'_n \to X'_0$, where $X'_n = \varphi_n$ on the probability space Ω. This result also holds if the X_n take values in a Borel subset E of $\mathbb{R}$.

If the X_n assume values in the countable product $\mathbb{R}^I$ there exists a countable dense set Sß$\mathbb{R}$ such that $\mathbb{P}\{X_n^{(i)} \in S\} = 0$ for each $i \in I$ and $n \geq 0$. So we may regard the

X_n as elements of T^I where $T = \mathbb{R} \setminus S$ is homeomorphic to the set of irrationals, as is T^I. Now apply the above with $E = \mathbb{R} \setminus \mathbb{Q}$.

Finally note that a complete separable metric space is homeomorphic to a Borel set in $\mathbb{R}^\infty$. $\square$

We also need some more technical results.

Theorem 4.9 (Skorohod's Representation Theorem, Extension). *Let $Z_n = (X_n, Y_n)$, $n \geq 0$, be random elements in the product $\mathcal{X} \times \mathcal{Y}$ of two complete separable metric spaces. Suppose $X_n \Rightarrow X_0$. There exist $Z'_n = (X'_n, Y'_n)$ distributed like Z_n, and defined on a common probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, such that $X'_n(\omega) \to X'_0(\omega)$ for each $\omega \in \Omega'$.*

Proof. The products $\mathcal{X}' = \mathcal{X}^{\mathbb{N}_0}$ and $\mathcal{Y}' = \mathcal{Y}^{\mathbb{N}_0}$ with $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ are Polish. Hence there exists a regular conditional probability $\pi_x(dy)$ for $Y = (Y_0, Y_1, \dots)$ given $X = (x_0, x_1, \dots)$. In the previous theorem one may choose $\Omega = \mathcal{X}'$, and one may then restrict Ω to the set of ω for which $X_n(\omega) \to X_0(\omega)$. Set $\Omega' = \Omega \times \mathcal{Y}'$ß$\mathcal{X}' \times \mathcal{Y}'$, with the probability measure $d\,\mathbb{P}'(\omega, y) = \pi_\omega(dy) d\,\mathbb{P}(\omega)$. The canonical variable $Z' = (Z'_0, Z'_1, \dots)$ is distributed like $(Z_0, Z_1, \dots)$, and $X'_n(\omega') \to X'_0(\omega')$ for $\omega' = (\omega, y)$, $\omega \in \Omega$. $\square$

Lemma 4.10. *Let $X_0, X_1, \dots$ be random elements in $\mathcal{X}$. Suppose for each ε there exist $Y_0, Y_1, \dots$ such that $\mathbb{P}\{d(X_n, Y_n) \geq \varepsilon\} < \varepsilon$ for $n \geq n_\varepsilon$ and $n = 0$, and $Y_n \Rightarrow Y_0$. Then $X_n \Rightarrow X_0$.*

Proof. Let $\varphi \colon \mathcal{X} \to [0, 1]$ be uniformly continuous. Let $\varepsilon > 0$. Choose $\delta \in (0, \varepsilon]$ so that $d(x, y) < \delta \Rightarrow |\varphi(x) - \varphi(y)| < \varepsilon$. Then

$$\mathbb{E}|\varphi(X_n) - \varphi(Y_n)| \leq \mathbb{P}\{d(X_n, Y_n) \geq \delta\} + \mathbb{E}|\varphi(X_n) - \varphi(Y_n)| 1_{\{d(X_n, Y_n) < \delta\}} \leq \varepsilon + \delta$$

for $n \geq n_\delta$ and $n = 0$. The triangle inequality gives an upper bound $2\varepsilon + 2\delta + \varepsilon_n$ with $\varepsilon_n \to 0$ for $\mathbb{E}|\varphi(X_n) - \varphi(X_0)|$. $\square$

Definition. Let $\mu_0, \mu_1, \dots$ be finite measures on the separable metric space $\mathcal{X}$. Then $\mu_n \to \mu_0$ *weakly* if $\int \varphi d\mu_n \to \int \varphi d\mu_0$ for all bounded continuous functions $\varphi \colon \mathcal{X} \to \mathbb{R}$.

Results about weak convergence of finite measures $\mu_n \to \mu_0$ follow from the corresponding results for the probability measures $\pi_n = \mu_n / \mu_n(\mathcal{X})$, using convergence $\mu_n(\mathcal{X}) \to \mu_0(\mathcal{X})$. If $\mu_0 = 0$ it suffices that $\mu_n(\mathcal{X}) \to 0$. We give three results on weak convergence of finite measures.

Bounded continuous functions on $\mathbb{R}^d$ need not have a limit for $\|x\| \to \infty$. For weak convergence it suffices to consider functions which have a finite limit.

Proposition 4.11. *Let $\mathcal{X}$ be a subspace of the compact metrizable space $\mathcal{Z}$, and let $\mu_0, \mu_1, \dots$ be finite measures on $\mathcal{X}$. If $\int \varphi d\mu_n \to \int \varphi d\mu_0$ for each continuous function $\varphi\colon \mathcal{Z} \to \mathbb{R}$, then $\int \varphi d\mu_n \to \int \varphi d\mu_0$ for all bounded continuous functions φ on $\mathcal{X}$.*

Proof. Any measure μ on the subspace $\mathcal{X}$ is automatically a measure $\bar{\mu}$ on $\mathcal{Z}$ by $\bar{\mu}(E) = \mu(E \cap \mathcal{X})$. By assumption $\bar{\mu}_n \to \bar{\mu}_0$ weakly on $\mathcal{Z}$. Let $c_n = \mu_n(\mathcal{X}) = \bar{\mu}_n(\mathcal{Z})$ for $n \geq 0$. Weak convergence $\bar{\mu}_n \to \bar{\mu}_n$ implies $c_n \to c_0$. Hence $\mu_n \to \mu_0$ weakly if $c_0 = 0$. Assume $c_0 > 0$. We may assume that all c_n are positive. Hence the probability measures $\pi_n = \mu_n/c_n$ and $\bar{\pi}_n = \bar{\mu}_n/c_n$ are well defined. Weak convergence $\bar{\pi}_n \to \bar{\pi}_0$ on $\mathcal{Z}$ is given. We shall prove weak convergence $\pi_n \to \pi_0$ on $\mathcal{X}$, using condition 4) in Theorem 4.5. Let $O ß \mathcal{X}$ be open, and $\pi(O) > c$. We have to show that $\pi_n(O) > c$ eventually. By definition of subspace $O = \mathcal{X} \cap U$ for an open set U in $\mathcal{Z}$. Since $\bar{\pi}_n(U) = \pi_n(O)$ it suffices to show that $\bar{\pi}_0(U) > c$ implies $\bar{\pi}_n(U) > c$ eventually. This follows by weak convergence $\bar{\pi}_n \to \bar{\pi}_0$. □

For convergence of real-valued random variables there does not exist a special theory for positive random variables, or for random variables with values in the interval $[-1, 1]$. Weak convergence does not depend on the space on which the random variables live!

Theorem 4.12 (Subspace Theorem). *If $\mathcal{X}$ is a topological subspace of the separable metric space $\mathcal{Y}$, then $\mathcal{P}(\mathcal{X})$ is a topological subspace of $\mathcal{P}(\mathcal{Y})$.*

Proof. Let $\mathcal{Z} \supset \mathcal{Y}$ be a compact metric space. Random elements X_n, $n \geq 0$, in $\mathcal{X}$ are automatically random elements of $\mathcal{Y}$ and $\mathcal{Z}$. By the proposition above $\mathbb{E}\varphi(X_n) \to \mathbb{E}\varphi(X_0)$ holds for continuous $\varphi\colon \mathcal{Z} \to \mathbb{R}$ if and only if this holds for continuous bounded $\varphi\colon \mathcal{X} \to \mathbb{R}$, or $\varphi\colon \mathcal{Y} \to \mathbb{R}$. □

What can one say about convergence of the integrals $\int \varphi_n d\mu_n$ when the integrands and the measures both vary with n?

Theorem 4.13 (Continuity Theorem). *Let $\mu_n \to \mu_0$ weakly on $\mathcal{X}$. Let $\varphi_n\colon \mathcal{X} \to \mathbb{R}$, $n \geq 0$, be uniformly bounded Borel functions. Then $\int \varphi_n d\mu_n \to \int \varphi_0 d\mu_0$ if there exists a μ_0-null set E such that $\varphi_n(x_n) \to \varphi_0(x_0)$ for all $x_0 \in E^c$ and each sequence $x_n \to x_0$.*

Proof. Set $T = \{t_0, t_1, \dots\}$ with $t_n = 1/n$ for $n \geq 1$ and $t_0 = 0$. Define φ on $\mathcal{X}' = \mathcal{X} \times T$ by $\varphi(x, t_n) = \varphi_n(x)$. Let $\bar{\mu}_n$ on $\mathcal{X} \times \{t_n\}$ correspond to μ_n on $\mathcal{X}$. Then $\bar{\mu}_n \to \bar{\mu}_0$ weakly on $\mathcal{X}'$, and

$$\int \varphi_n d\mu_n = \int \varphi d\bar{\mu}_n \to \int \varphi d\bar{\mu}_0 = \int \varphi_0 d\mu_0,$$

since φ is $\bar{\mu}_0$-a.e. continuous. □

4.4 Radon measures and vague convergence. In this subsection $\mathcal{X}$ is an lcscH space. Define $\mathcal{R} = \mathcal{R}(\mathcal{X})$ as the set of all *Radon measures* ρ on $\mathcal{X}$. Such measures are finite on compact sets. A *counting measure* on $\mathcal{X}$ is a Radon measure ρ such that $\rho(C)$ is an integer for all compact sets Cß$\mathcal{X}$. Convergence of measures is determined by convergence of certain integrals

$$\int \varphi d\rho_n \to \int \varphi d\rho_0. \tag{4.1}$$

A function $\varphi\colon \mathcal{X} \to \mathbb{R}$ has *compact support* if the set $\{f \neq 0\}$ is *relatively compact*, i.e. if the closure of this set is compact.

Definition. Let $\rho_0, \rho_1, \dots$ be Radon measures on the lcscH space $\mathcal{X}$. Then $\rho_n \to \rho_0$ *vaguely* if (4.1) holds for all continuous functions φ with compact support.

Example 4.14. For vague convergence it is sufficient that densities converge. Let $f_0, f_1, \dots$ be continuous non-negative functions on the open set Oß$\mathbb{R}^d$. Suppose $f_n(x_n) \to f_0(x_0)$ whenever $x_n \to x_0$. Then $f_n(x)dx \to f_0(x)dx$ vaguely on O.$\Diamond$

Theorem 4.15. *The space $\mathcal{R}$ of Radon measures on the lcscH space $\mathcal{X}$ with the weakest topology which makes the maps $\rho \mapsto \int \varphi d\rho$ continuous for each continuous function φ on $\mathcal{X}$ with compact support, is Polish. If $\mathcal{X}$ is compact then $\mathcal{R}$ is lcscH, and the sets $\{\rho \in \mathcal{R} \mid \rho(\mathcal{X}) \leq c\}$ are compact.*

Proof. See Parthasarathy [1967] or Kallenberg [2002]. □

On compact metric spaces vague and weak convergence coincide. On locally compact spaces vague convergence is weaker than weak convergence.

Proposition 4.16. *Suppose $\mu_0, \mu_1, \dots$ are finite measures on the lcscH space $\mathcal{X}$, and $\mu_n \to \mu_0$ vaguely. Then $\mu_n \to \mu_0$ weakly if $\mu_n(\mathcal{X}) \to \mu_0(\mathcal{X})$.*

Proof. By Proposition 4.11 it suffices to prove $\int \varphi d\mu_n \to \int \varphi d\mu_0$ for functions φ which are continuous on the one-point compactification $\mathcal{X} \cup \{\infty\}$. Let $\varphi(\infty) = c$. Then $\varphi = c + \varphi_0$ where φ_0 vanishes in ∞. In particular $\varphi_n = (\varphi_0 - 1/n)_+$ is continuous on $\mathcal{X}$ and has compact support. Hence (4.1) holds for $c + \varphi_m$ for each m. Since $c + \varphi_m \to \varphi$ uniformly on $\mathcal{X}$ it also holds for φ. □

Let χ be a continuous positive function on $\mathcal{X}$, as in Section 4.2, such that the sets $\{\chi \leq c\}$ are compact. One may create continuous functions with compact support, $0 \leq \chi_n \uparrow 1$, by setting

$$\chi_n = 0 \vee (n - \chi) \wedge 1. \tag{4.2}$$

Vague convergence $\rho_n \to \rho_0$ is equivalent to weak convergence of $\chi_m d\rho_n \to \chi_m d\rho_0$ for each of these functions χ_m. This yields a number of equivalent formulations of vague convergence.

Proposition 4.17. *Let $\rho_0, \rho_1, \ldots$ be Radon measures on the lcscH space $\mathcal{X}$. Equivalent are:*

1) $\rho_n \to \rho_0$ *vaguely;*

2) (4.1) *holds for continuous $\varphi \geq 0$ with compact support;*

3) (4.1) *holds for bounded Borel functions which vanish outside a compact set, and which are ρ_0-a.e. continuous;*

4) $\rho_n(O) \to \rho_0(O)$ *for relatively compact open sets with $\rho_0(\partial O) = 0$.*

Proof. The equivalence of these four criteria follows from Theorem 4.5 by the remark above. □

Proposition 4.18. *Let $\rho_n \to \rho_0$ on the lcscH space $\mathcal{X}$, and let Oß$\mathcal{X}$ be open, and $\rho_0(O) > c$. Then $\rho_n(O) > c$ eventually.*

Proof. Since O itself is lcscH there exist continuous functions χ_n with compact support, as in (4.2), such that $0 \leq \chi_n \uparrow 1_O$. Choose m so large that $\int \chi_m d\rho_0 > c$, and use $\int \chi_m d\rho_n \to \int \chi_m d\rho_0$. □

Relatively compact sets in $\mathcal{R}$ have a simple characterization.

Proposition 4.19. *A set Eß$\mathcal{R}$ is relatively compact if* $\sup\{\rho(C) \mid \rho \in E\}$ *is finite for Cß$\mathcal{X}$ compact.*

Proof. The measures $\chi_m d\rho$, $\rho \in E$, with χ_m as in (4.2), are bounded, and live on the compact set $\mathcal{X}_m = \{\chi \leq m\}$. By a diagonalisation argument there is a sequence ρ_n in E such that for each m the sequence $\chi_m d\rho_n$ converges weakly to a finite measure σ_m on $\mathcal{X}_m$. The measures σ_n, $n > m$, agree on $\mathcal{X}_m$. Hence they define a Radon measure σ on $\mathcal{X}$, and $\int \varphi d\rho_n \to \int \varphi d\sigma$ holds for any continuous φ with compact support, since such a function vanishes off $\mathcal{X}_m$ for $m \geq m_0$. □

Exercise 4.20. Suppose $\rho_n \to \rho$ vaguely on $\mathcal{X}$. Let Oß$\mathcal{X}$ be open, and $\psi : O \to [0, \infty)$ continuous. Then $1_O \psi d\rho_n \to 1_O \psi d\rho$ vaguely on O. ◊

Both open and closed sets in $\mathcal{X}$ are lcscH spaces in their own right. For open sets convergence is simple by the exercise above. For closed sets we have:

Theorem 4.21 (Weak Convergence). *Let $\rho_n \to \rho_0$ vaguely on the locally compact separable metric space $\mathcal{X}$. Let F be a closed set in $\mathcal{X}$, and let $\psi : F \to [1, \infty)$ be continuous. Assume*

$$\rho_0(\partial F) = 0, \qquad \int_F \psi d\rho_n \to \int_F \psi d\rho_0 < \infty. \tag{4.3}$$

Then $1_F \psi d\rho_n \to 1_F \psi d\rho_0$ weakly. In particular

$$\int_F \varphi d\rho_n \to \int_F \varphi d\rho_0 \tag{4.4}$$

for ρ_0-a.e. continuous $\varphi : F \to \mathbb{R}$ for which φ / ψ is bounded.

Proof. Let $d\mu_n = 1_F \psi d\rho_n$. Then $\mu_n \to \mu_0$ vaguely on $\mathcal{X}$ by 4) in Theorem 4.17 since $f 1_F \psi$ is ρ_0-a.e. continuous, and bounded with compact support if f is continuous with compact support. In fact $\mu_n \to \mu_0$ weakly by (4.3) and Proposition 4.16. The relation (4.4) follows from 4) in Theorem 4.5. □

Exercise 4.22. Suppose $\rho_n \to \rho_0$ on $\mathcal{X}$. If ρ_n is a counting measure for $n \geq 1$, then so is ρ_0. ♢

Let us say a few words about the metric on $\mathcal{R}(\mathcal{X})$. The reader may wonder why we do not define metrics on the Polish spaces $\mathcal{P}(\mathcal{X})$ and $\mathcal{R}(\mathcal{X})$. The answer is simple: It is easier to work with pseudometrics. A pseudometric has all the properties of a metric, except that $d(x, y) = 0$ does not imply $y = x$.

The metric d on $\mathcal{R}(\mathcal{X})$ is not very intuitive. One starts out with a sequence of pseudometrics d_n which define the topology, and sets

$$d(\mu, \rho) = \max_n (2^{-n} \wedge d_n(\mu, \rho)).$$

If $\mathcal{X}$ is compact one uses the pseudometrics $d_n(\mu, \rho) = |\int \varphi_n d\mu - \int \varphi_n d\rho|$. Here (φ_n) is a dense sequence in the unit ball of the separable Banach space $C(\mathcal{X})$ of all continuous functions $\varphi \colon \mathcal{X} \to \mathbb{R}$ with the sup-norm. If $\mathcal{X}$ is locally compact but not compact we use the metrics d'_m on $\mathcal{R}(\mathcal{X}_m)$, where $\mathcal{X}_m = \{\chi \leq m\}$ is compact, and define $d_n(\mu, \rho) = d'_n(\chi_n d\mu, \chi_n d\rho)$ with χ_n defined in (4.2).

Metrizability of $\mathcal{R}(\mathcal{X})$ allows us to apply results about weak convergence of random elements on separable metric spaces to point processes.

4.5 Convergence of point processes. The space $\mathcal{N}(\mathcal{X})$ of all counting measures on the locally compact separable metric space $\mathcal{X}$ is a closed subset of the complete separable metric space $\mathcal{R}(\mathcal{X})$, and hence it is itself a complete separable metric space. In particular one may apply the results of Subsection 4.3 to the convergence of point processes $N_n \Rightarrow N_0$ if one restricts attention to point processes on locally compact separable metric spaces whose realizations are integer-valued Radon measures. We shall do so henceforth.

Theorem 4.23 (Tightness). *A family of point processes N_t, $t \in T$, on a locally compact separable metric space is tight if for each compact set C the family of random integers $N_t(C)$, $t \in T$, is tight.*

Proof. Let $\chi > 0$ be continuous such that $\{\chi \leq c\}$ is compact for all $c > 0$. Choose $m_n \uparrow \infty$ such that $\mathbb{P}\{N_t\{\chi \leq n\} > m_n\} < 1/2^n$ for all t. The union E_k of these events for $n > k$ has probability $< 1/2^k$. For any $k \geq 1$ the set of Radon measures $\rho = N_t(\omega)$, $t \in T$, $\omega \in \Omega \setminus E_k$, is relatively compact by Proposition 4.19 since $\rho\{\chi \leq n\} \leq m_n$ for $n > k$. □

We now turn to convergence in distribution for point processes. We employ a technique which is widely used to establish convergence in distribution for a sequence of stochastic processes to a limit X_0. First show that the sequence is *tight*, then check that any limit point of the sequence has to satisfy certain conditions, and finally observe that X_0 is the only process which satisfies these conditions.

Theorem 4.24 (Convergence in Distribution). *Let N_n, $n \geq 0$, be point processes on the lcscH space $\mathcal{X}$ which are finite on compact sets. The following are equivalent:*

1) $N_n \Rightarrow N_0$;

2) $\int \varphi dN_n \Rightarrow \int \varphi dN_0$ *for all continuous $\varphi \geq 0$ on $\mathcal{X}$ which vanish outside a compact set;*

3) $(N_n(O_1), \dots, N_n(O_m)) \Rightarrow (N_0(O_1), \dots, N_0(O_m))$ *for relatively compact open sets $O_1, \dots, O_m$ with $\mathbb{P}\{N_0(\partial O_i) > 0\} = 0$ for $i = 1, \dots, m$.*

Proof. The implication from 1) to 2) holds by the continuity theorem since $\nu \mapsto \int \varphi d\nu$ is continuous on $\mathcal{N}ß\mathcal{R}$ by definition for any continuous function φ with compact support. Similarly 1) implies 3). Conversely each of the two conditions 2) and 3) ensures that the family of distributions π_n, $n \geq 1$, of N_n is tight. By Prohorov's theorem, 4.6, any subsequence contains a subsubsequence which converges weakly to some distribution π on $\mathcal{N}$. By the argument above each limit point π satisfies the conditions 2) and 3). It remains to observe that π_0 is the only distribution which satisfies these conditions: the distributions of $\int \varphi dN$ determine the distribution of N, as do the distributions of the vectors $(N(O_1), \dots, N(O_m))$, where we restrict the open sets O_i to be finite unions of sets in some countable relatively compact open base $U_1, U_2, \dots$, which we choose so that $N(\partial U_i)$ vanishes a.s. for $i = 1, 2, \dots$, see Lemma 2.17. □

Corollary 4.25. *Let the Poisson point processes N_n on $\mathcal{X}$ have mean measure ρ_n for $n \geq 0$. If $\rho_n \to \rho_0$ vaguely, then $N_n \Rightarrow N_0$ vaguely.*

Remark 4.26. For a simple limit point process it suffices for condition 3) to hold for $m = 1$; see Theorem 3.8.

One may regard a point process as

1) a random element of the separable metric space $\mathcal{N}$, or as

2) a random process $N(f)$, $f \in \mathcal{C}_c$ (continuous with compact support), or as

3) a random process $N(U)$, U open and relatively compact.

Each of these interpretations suggests an interpretation of the limit relation $N_n \Rightarrow N_0$. In the first interpretation we speak of convergence of random elements in the metric space $\mathcal{N}$; in the second we speak of finite-dimensional convergence of the marginals $N_n(f) \Rightarrow N_0(f)$, and it turns out that univariate convergence suffices; in the third interpretation we consider finite-dimensional convergence in the continuity points, i.e. in those relatively compact open sets U for which $N_0(\partial U)$ vanishes almost surely. By Theorem 4.24 the three interpretations of $N_n \Rightarrow N_0$ are equivalent.

Point processes were introduced as geometric objects – the trace of a shower on a subset of the plane. How should one envisage convergence $N_n \Rightarrow N_0$ in this context? If the state space $\mathcal{X}$ is an open set in $\mathbb{R}^d$ then the set $\mathcal{N}$ of counting measures is a closed subset of the Polish space $\mathcal{R}$ and we may replace convergence in distribution by almost sure convergence by the Skorohod representation. So assume $N_n \to N_0$ a.s.

First suppose N_0 is a finite point process and $N_n(\omega) \to N_0(\omega)$ weakly. Then $K_n(\omega) = N_n(\omega)(\mathcal{X}) \to k = N_0(\omega)(\mathcal{X})$. So $K_n(\omega) = k$ for $n \geq n_0$. Let $x_1, \dots, x_k$ be the points of $N_0(\omega)$, where multiplicity is taken care of by repetition. The k points of $N_n(\omega)$ for $n \geq n_0$ may be enumerated so that $x_{ni} \to x_i$ for $i = 1, \dots, k$. Indeed, there are $j \leq k$ distinct points $a_1, \dots, a_j$ in the set $\{x_1, \dots, x_k\}$, with multiplicities $m_1, \dots, m_j$. Choose $\delta > 0$ small so that the j balls $B^\delta(a_i)$ of radius δ around a_i are disjoint. Then $N_n(\omega)(B^\delta(a_i)) > m_i - 1/2$ eventually by Proposition 4.18. Since there are only k points, and the balls are disjoint, we conclude that

$$N_n(\omega)(B^\delta(a_i)) = m_i, \quad i = 1, \dots, j,\ n \geq n_1(\omega).$$

To handle vague convergence we need a lemma.

Lemma 4.27. *For any point process N_0 there exists an increasing sequence of open relatively compact sets O_n which cover the lcscH space $\mathcal{X}$ such that $N_0(\partial O_n) = 0$ a.s. for each n.*

Proof. By Exercise 1.4 there exists a finite measure ν^* on $\mathcal{X}$ such that

$$\nu^*(E) = 0 \Rightarrow N_0(E) = 0 \text{ a. s.}, \quad E \text{ a Borel set in } \mathcal{X}. \tag{4.5}$$

Choose $\chi\colon \mathcal{X} \to (0, \infty)$ continuous so that $\{\chi \leq c\}$ is compact for all c. By Lemma 2.17 in each interval $(n, n+1)$ we may choose c_n so that the boundary of $O_n = \{\chi < c_n\}$ has mass $\nu^*(\partial O_n) = 0$. □

So now suppose $N_n(\omega) \to N_0(\omega)$ vaguely, and $N_0(\omega)(\partial O_m) = 0$ for $m \geq 1$. Let $x_1, x_2, \dots$ denote the points of $N_n(\omega)$ (with repetition to take care of multiplicity), arranged so that $x_1, \dots, x_{k_m}$ are the points inside O_m. One may enumerate the points $x_{n1}, x_{n2}, \dots$ of $N_n(\omega)$ similarly. As above $x_{nk} \to x_k$ for $k = 1, 2, \dots$. It may happen that $N_0(\omega)$ has no points, or only finitely many points, even if $N_n(\omega)$ is infinite for each n. If $N_0(\omega)(\mathcal{X}) = k$, then all k points lie inside some set O_m. Eventually $N_n(\omega)$ has k points inside O_m. These k points converge to the k points $x_1, \dots, x_k$. The remaining points diverge uniformly: For any compact set Cß$\mathcal{X}$ containing O_m there exists an index n_0 such that $x_{ni} \in C^c$ for $i > k,\ n \geq n_0$.

Enumerating points is rather artificial. So we give an alternative description. We assume for simplicity that the point processes N_n live on a compact set $F = \{\chi \leq c\}$

in $\mathcal{X}$, and that $N_0(\partial F) = 0$ a.s. For any $\delta > 0$ there is a finite partition $\mathcal{C}$ of F into cells C of diameter less than δ with the additional property $N_0(\partial C) = 0$ a.s. Let $U_1, \dots, U_j$ denote the interiors of these j cells. The vector K_0 of occupancy numbers $K_{0i} = N_0(U_i), i = 1, \dots, j$ describes how the points of N_0 are distributed over the cells. Define $K_n = (K_{n1}, \dots, K_{nj})$ similarly. If $N_n(\omega) \to N_0(\omega)$ then $K_n(\omega) \to K_0(\omega)$ and since the components of these vectors are integers, we even have $K_n(\omega) = K_0(\omega)$ for $n \geq n_0$. One may choose n_0 so large that $N_n(\omega)(F) = N_0(\omega)(F)$ for $n \geq n_0$. The partition $\mathcal{C}$ fails to distinguish the realizations $N_n(\omega)$ and $N_0(\omega)$ for $n \geq n_0$.

Proof of Theorem 2.16 (Grigelionis). First assume $\mathcal{X}$ is lcscH. The limit is a Poisson point process N_0 with mean measure μ_0. We may restrict the point processes to an open relatively compact set O with $\mu_0(\partial O) = 0$. Then μ_0 is finite. It suffices to construct Poisson point processes M_n with mean measure μ_n such that $\mu_n \to \mu_0$ weakly, and such that $\mathbb{P}\{M_n \neq N_n\} \to 0$ by Lemma 4.10. Weak convergence $\mu_n \to \mu_0$ implies weak convergence $M_n \Rightarrow N_0$ by Corollary 4.25. So choose independent 0-1 point processes L_{ni} which agree with N_{ni} for $N_{ni}(\mathcal{X}) \leq 1$, and which are zero for $N_{ni}(\mathcal{X}) > 1$, and let M_{ni} be the corresponding independent Poisson point processes. We compare the sums M_n of M_{ni} and N_n of N_{ni}. The conditions of the theorem then give

$$\mathbb{P}\{M_n \not\equiv N_n\} \leq \sum_i \mathbb{P}\{N_{ni}(\mathcal{X}) > 1\} + \sum_i \mathbb{P}\{N_{ni}(\mathcal{X}) = 1\}^2 \to 0.$$

If $\mathcal{X}$ is not locally compact, we embed $\mathcal{X}$ in an lcscH space $\mathcal{Y}$. By assumption there is a measure μ on $\mathcal{X}$. This allows us to construct a limit Poisson point process M on $\mathcal{X}$ with mean measure μ. □

5 Converging sample clouds

5.1 Introduction. The theory of vague convergence of Radon measures does not quite fit our needs. Our set-up is very limited. We start with the n-point sample cloud N_n from a probability distribution π_n on $\mathbb{R}^d$, and assume vague convergence on an open set $O ß \mathbb{R}^d$ of the mean measures:

$$d\rho_n := 1_O n d\pi_n \to d\rho \quad \text{vaguely on } O, \tag{5.1}$$

where ρ is a Radon measure on O. Both $\mathbb{R}^d$ and O are locally compact separable metric spaces. It follows from the general theory that $1_O dN_n \Rightarrow dN$ vaguely, where N is the Poisson point process on O with mean measure ρ. We may assume that O is maximal, see Proposition 5.16. Compact sets a.s. contain finitely many points of N; bounded sets may contain infinitely many.

Our interest is in halfspaces rather than compact subsets. Closed halfspaces are our basic tool for investigating the edge of the n-point sample clouds and of the limiting Poisson point process. In geometric terms the extreme of a sample cloud is its *convex hull*. To see (feel) the convex hull of a typical realization $E = N(\omega)$ of the limit process, we use halfspaces which do not contain any points of E. For exploring the edge of the sample cloud, we use halfspaces which contain finitely many points. A Radon measure is described in terms of compact sets. We want to describe the mean measure ρ of the limit Poisson point process in terms of its behaviour on closed halfspaces on which it is finite.

This means that we have to consider the class of halfspaces H on which ρ is finite:

$$\mathcal{H}_\rho = \{H \text{ß} O \mid \rho(H) < \infty\}.$$

Vague convergence means that $\rho_n(K) \to \rho(K)$ for compact sets K whose boundary carries no mass, $\rho(\partial K) = 0$. We are interested in the same relation, but for halfspaces.

In the asymptotic theory for coordinatewise maxima the transition from closed halfspaces to compact sets is achieved by a special partial compactification:

$$\mathbb{R}^d \text{ß} \mathcal{X} = [-\infty, \infty]^d \setminus \{(-\infty, \dots, -\infty)\}.$$

Topologically the space $\mathcal{X}$ is just the compact unit cube $[0, 1]^d$ with the lower vertex 0 removed. In the enlarged space the upper coordinate halfspaces $\{\xi_i \geq c\}$ extend to compact sets; halfspaces $\{\xi \geq c\}$ with $\xi \in [0, \infty)^d$, $\xi \neq 0$, are relatively compact. The exponent measure ρ of a max-stable df G is a Radon measure on $\mathcal{X}$. It may charge the faces and edges in $-\infty$. The mass of ρ is infinite on neighbourhoods of the (absent) lower *vertex*. Vague convergence $n\pi_n \to \rho$, with $\pi_n = \alpha_n^{-1}(\pi)$ for suitable affine transformations α_n, holds for π in the domain of G. Section 7 gives details.

We take a different route. Our approach is elementary. Skorohod's Representation Theorem will play an important role. The approach via almost-sure convergence has the advantage that it allows us to handle convergence of convex hulls of sample clouds without having to introduce weak convergence for probability distributions of random convex sets. Geometric operations like peeling off the extreme points are simple to handle in this setting.

Since our halfspaces are not compactified it is possible to consider the behaviour of stochastic integrals $\int_H \varphi dN_n$ for *unbounded* loss functions such as $1 + \|w\|^2$. It is clear that such integrals are of interest in risk analysis. Two questions will occupy us in this section:

1) When do convex hulls converge?

2) When do stochastic loss integrals converge?

Exercise 5.1. The vector $Z \in \mathbb{R}^d$ has density f which is continuous and positive in the origin. Determine the density of Z/c. Choose $c_n \to 0+$ so that nf_n has a non-zero limit where f_n is the density of Z/c_n. Prove that $d\rho_n(w) = nf_n(w)dw$ converges vaguely to a multiple of *Lebesgue measure* on $\mathbb{R}^d$. What does this say about the sample clouds? ♢

5.2 Convergence of convex hulls, an example. The Poisson point process N on $\mathbb{R}^3$ whose mean measure ρ has a *Gauss-exponential* density,

$$g_0(w) = e^{-u^T u/2-v}/2\pi, \quad w = (u, v) \in \mathbb{R}^{2+1}, \tag{5.2}$$

plays a central role in the theory of high risk scenarios. Let us take a typical realization $E = N(\omega)$, and a sequence of finite sets F_n which converge vaguely to E. Under what conditions will the convex hulls $c(F_n)$ converge to $c(E)$?

Since the measure ρ has no atoms the point process N is simple, and we may treat E as a subset of $\mathbb{R}^3$ with points $w_n = (u_n, v_n) \in \mathbb{R}^{2+1}$. The projection of ρ on the vertical axis has density e^{-v}. So we may arrange the points w_n in decreasing order of their last coordinate: $v_1 > v_2 > \cdots$. The points $u_1, u_2, \ldots$ are dense in $\mathbb{R}^2$. Conditional on $V_n = v$, the horizontal component of the nth point W_n of N has a standard normal distribution. The upper halfspace $H_+ = \{v \geq 0\}$ has measure $\rho(H_+) = 1$. So has every halfspace H supported by the paraboloid $V = \{v < -u^T u/2\}$. The mass of a halfspace decreases exponentially when it moves upwards. Eventually the translate of H will contain no points of E. Thus the convex hull of E is a rough piecewise linear approximation to the *paraboloid* V.

Vague convergence $F_n \to E$ implies that each $w \in E$ is limit of a sequence $w_n \in F_n$. Hence interior points of $c(E)$ lie in $c(F_n)$ eventually:

$$\text{int}(c(E))\text{ß} \liminf c(F_n). \tag{5.3}$$

If any halfspace disjoint from E eventually contains no points of F_n then the convex hulls converge.

Since F_n is finite, one may arrange its points $w_{n1}, \ldots, w_{nm_n}$ in decreasing order of the vertical component. Suppose $c(F_n) \to c(E)$. Adding the point $(0, n)$ to F_n does not affect *vague convergence*. It does affect the convex hulls: $c(F_n) \to \mathbb{R}^d$ now. Moreover w_{n1} does not converge to w_1. So assume $F_n \to E$ weakly on all horizontal halfspaces $\{v \geq c\}$, $c \in \mathbb{R}$. Then $w_{nk} \to w_k$ for every k. For horizontal halfspaces H whose boundary contains no points of E the convex hull of $F_n \cap H$ converges to the convex hull of $E \cap H$. That does not yield convergence of the convex hulls $c(F_n) \cap H \to c(E) \cap H$ on such a halfspace. (Add the four points $(\pm n^2, \pm n^2, -n)$ to F_n. Then $c(F_n)$ will converge to the halfspace $\{v \leq v_1\}$.) The shape of $c(E) \cap H$ depends on the behaviour of E in the complement of the halfspace H!

An obstruction to convergence of the convex hulls is a sequence $w_n = (u_n, v_n) \in F_{k_n}$ which diverges away from the parabola: this happens if v_n is bounded below and $\|w_n\| \to \infty$, or if $v_n \to -\infty$, and $|v_n| = O(\|u_n\|)$. Let us call a convex set C a *needle* if the diameter of $C \cap \{v = -n\}$ is $o(n)$ for $n \to \infty$.

Exercise 5.2. The convex hulls converge if $\bigcup F_n$ lies in a needle. $\Diamond$

If $c(F_n) \to c(E)$, vague convergence implies

$$\#(F_n \cap H) \to \#(E \cap H) \tag{5.4}$$

for any halfspace $H = \{v \geq c^T u + c_0\}$ whose boundary contains no points of E. It follows that for each halfspace $\{v \geq c^T u\}$ there exists a constant $c_0 \geq 0$ such that $\{v \geq c^T u + c_0\}$ contains no points of $\bigcup F_n$. Conversely, existence of such translates ensures convergence of the convex hulls. This may be seen as follows: Consider the four halfspaces

$$\{v \geq mu_1\}, \quad \{v \geq -mu_1\}, \quad \{v \geq mu_2\}, \quad \{v \geq -mu_2\}.$$

The intersection of the four complements is an open cone V_m. The condition yields constants $c_m \geq 0$ such that $V_m + (0, c_m)$ contains $E \cup \bigcup F_n$. The intersection of these translated cones is a needle. Now use the exercise above.

The complement of the paraboloid $V = \{v < -u^T u/2\}$ has infinite mass. This also holds for vertical translates of V. Hence no such translate covers E. However the mass of the complement of $W = rV + (0, c)$ for $r > 1$ may be made arbitrarily small by choosing c large. Indeed, $\rho(W^c) = e^{-c}\rho(rV^c)$, and the second factor is finite:

$$\begin{aligned}\rho(rV^c) &= \int_{v/r > u^T u/2r^2} g_0(u, v)\,dv\,du \\ &= \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{-u^T u/2} e^{u^T u/2r}\,du = 1/(1 - 1/r).\end{aligned}$$

For large c the paraboloid W covers E. If W also contains the sets F_n then $c(F_n) \to c(E)$.

Exercise 5.3. Give conditions for $c(F_n) \to c(E)$ where E is a typical realization of the standard Poisson point process on the quadrant $(-\infty, 0)^2$. $\Diamond$

5.3 Halfspaces, convex sets and cones. Halfspaces, convex sets and cones play an important role in geometric extreme value theory. To emphasize the geometric setting we write L for the vector space $\mathbb{R}^d$, and L^* for the dual space of linear functionals.

Definition. $\mathcal{H}$ is the collection of all halfspaces. Halfspaces H are closed by definition: $H = \{\xi \geq c\}$ with $\xi \in L^* \setminus \{0\}$ and $c \in \mathbb{R}$. Define $\mathcal{F}$ as the $d + 1$-dimensional linear space of *affine functions* $\varphi = \xi - c$. Then $H = \{\varphi \geq 0\}$ with $\varphi \in \mathcal{F}$, φ non-constant. The collection of convex sets with non-empty interior is denoted by $\mathcal{C}$. Recall that C is *convex* if it contains the line segment $[p, q] = \{p + \theta(q - p) \mid 0 \leq \theta \leq 1\}$ for $p, q \in C$. A set K is a *cone* if $p, q \in K$ and $s, t > 0$ implies $sp + tq \in K$. A closed cone is *proper* if it contains no lines through the origin.

Halfspaces constitute a basic ingredient of our theory. One may parametrise the set $\mathcal{H}$ of halfspaces by the points of the cylinder, $\partial B \times \mathbb{R}$ in $L^* \times \mathbb{R}$. Here B is the open unit ball. The correspondence

$$\mathcal{H} \ni H = \{\theta \geq c\} \leftrightarrow (\theta, c) \in \partial B \times \mathbb{R} \tag{5.5}$$

is a homeomorphism. The set $\mathcal{H}$ thus is a locally compact separable metric space. If L is an inner product space we may interpret the direction θ as the unit normal on ∂H pointing into H. Halfspaces with the same *direction* determine a vertical line in the cylinder.

Actually we are not so much concerned with $\mathcal{H}$, but rather with

$$\mathcal{H}_O = \{H \in \mathcal{H} \mid H \text{ß} O\} \tag{5.6}$$

where $O\text{ß}L$ is the open set on which the measure ρ lives, or on which the measures ρ_n converge vaguely to ρ_0. The set $\mathcal{H}_O$ need not be open in $\mathcal{H}$.

Example 5.4. Let $O\text{ß}\mathbb{R}^2$ be the complement of $(-\infty, 0]^2$. Then $\mathcal{H}_O$ corresponds to $S_1 \times (0, \infty)$ in the cylinder $S \times \mathbb{R}$ where S is the unit circle, and S_1 a closed quarter circle. This set is not open. $\diamondsuit$

From the point of view of halfspaces, lines are large sets. If the complement of O contains a line M then all $H\text{ß}O$ contain a line parallel to M. Hence the interior of $\mathcal{H}_O$ is empty.

There is an alternative description of halfspaces, in terms of affine functions. Rays, $\{r\varphi \mid r > 0\}$, in $\mathcal{F} \setminus \mathbb{R}\mathbf{1}$ correspond to halfspaces, $H = \{\varphi \geq 0\}$.

To clarify the terminology, write

$$H = \{\varphi \geq 0\}, \quad \varphi = \xi - c, \ \xi = r\theta \in L^*, \ \xi \neq 0, \ r > 0, \ \theta \in \partial B\text{ß}L^*.$$

This gives the commuting diagram of quotient maps below:

$$\begin{array}{ccc} \mathcal{F} \setminus \mathbb{R}\mathbf{1} & \longrightarrow & \mathcal{H} \\ \downarrow & & \downarrow \\ L^* \setminus \{0\} & \longrightarrow & \partial B \end{array} \qquad \begin{array}{ccc} \varphi & \longrightarrow & H \\ \downarrow & & \downarrow \\ \xi & \longrightarrow & \theta\,. \end{array} \tag{5.7}$$

We now turn to convex sets. Recall that sets in $\mathcal{C}$ have non-empty interior.

Definition. A sequence C_n in $\mathcal{C}$ *converges* to $C_0 \in \mathcal{C}$ if $1_{C_n} \to 1_{C_0}$ almost everywhere.

Exercise 5.5. $C_n \to C_0$ in $\mathcal{C}$ if and only if $1_{C_n}(x_n) \to 1_{C_0}(x_0)$ for $x_n \to x_0$, $x_0 \notin \partial C_0$. $\diamondsuit$

A fundamental result on convex sets is the Separation Theorem.

Proposition 5.6. *If* UßL *is a convex open set which does not contain the origin, then there exists a linear functional* $\xi \in L^*$ *such that* $\{\xi \geq 0\}$ *and* U *are disjoint.*

Proof. See Boyd & Vandenberghe [2004], Section 2.5.1. □

Theorem 5.7 (Separation Theorem). *If* C *and* V *are disjoint non-empty convex sets,* V *open, then there exists a halfspace* $H \supset C$ *which is disjoint from* V.

Proof. The set $U = C - V$ is convex and open; it does not contain the origin; and $\xi(U)$ß$(-\infty, 0)$ implies $\xi z < \xi v$ for $z \in C$, $v \in V$. Set $c = \sup \xi C$. Then $H = \{\xi \leq c\}$ contains C, and H is disjoint from V since $\xi(V)$ is open. □

Definition. The closed *convex hull* $c(E)$ of a set EßL is the closure of the convex hull of E. By the Separation Theorem $c(E)$ is the intersection of all halfspaces containing E.

The Separation Theorem gives an alternative description of the convex set $c(N(\omega))$, the closed convex hull of the point set $N(\omega)$ for a point process N on $\mathbb{R}^d$. The complement U of $c(N(\omega))$ is the union of all halfspaces which do not contain a point of $N(\omega)$ as interior point. By the Separation Theorem we may write

$$U = \bigcup \mathcal{H}_U, \quad \mathcal{H}_U = \{H \in \mathcal{H} \mid H \text{ß} U\}. \tag{5.8}$$

We want a stronger result: If the convex set $c(N(\omega))$ has non-empty interior, and contains no lines then U is the union of the halfspaces in the interior of $\mathcal{H}_U$.

We are not only interested in halfspaces which are disjoint from $N(\omega)$, but also in those which intrude into the sample cloud, halfspaces which contain finitely many points of $N(\omega)$. Such halfspaces contain a halfspace disjoint from $N(\omega)$. With a convex set V we therefore associate the cone

$$V^+ = \{\xi \in L^* \mid \sup \xi(V) < \infty\}. \tag{5.9}$$

Remark 5.8. If V is bounded then $V^+ = L^*$. The convex sets $V, V + a, a \in L, rV$, $r > 0$, and the closure $\text{cl}(V)$ all determine the same cone V^+. If V is a closed cone then $V^{++} = V$. If V contains a line then V^+ is contained in a hyperplane in L^*.

For the proof of the stronger *Separation Theorem* announced above we need a geometric result on cones.

Lemma 5.9. *Let* K *be the closed cone generated by a set* $C \in \mathcal{C}$ *whose closure does not contain the origin. If* K *contains a line* $M = \mathbb{R}a$ *for some* $a \neq 0$*, then* C *contains a line parallel to* M.

Proof. We may assume that C is open. First observe that K is a cylinder: K contains the line $p + M$ for each $p \in K$. This follows from plane geometry by intersecting K with the two-dimensional linear subspace through M and p. Similarly one sees that $\operatorname{int}(K)$ is a cylinder. The original problem may now also be reduced to a problem in plane geometry: If the closure of the non-empty open set C in $\mathbb{R}^2$ does not contain the origin, and if the open cone $\bigcup rC$, $r > 0$, contains a horizontal line, then so does C. We leave this to the reader as an exercise. □

Theorem 5.10. *If $C \in \mathcal{C}$ is closed and does not contain the origin, and contains no lines, then there is a linear functional $\xi \in \operatorname{int}(C^+)$. Moreover $\{\xi \geq 0\}$ and C are disjoint.*

Proof. Let K be the closed cone generated by C. If K^+ is contained in a proper linear subspace of L^*, then $K = K^{++}$ contains a line through the origin. The lemma above implies that the interior Λ of C^+ is non-empty. Let $\xi \in \Lambda$, and choose $\zeta_1, \dots, \zeta_d$ in Λ independent, such that $\xi = \zeta_1 + \cdots + \zeta_d$. We regard the ζ_i as coordinates. Then Kß$(-\infty, 0]^d$. (If the linear functional ζ_i is bounded above on the cone K, then Kß$\{\zeta_i \leq 0\}$.) Hence $K \cap \{\xi \geq 0\} = 0$. □

5.4 The intrusion cone. Recall that a halfspace $H_0 = \{\xi_0 \geq c_0\}$ is interior point of a set $\mathcal{H}_0$ß$\mathcal{H}$ if $\mathcal{H}_0$ contains all halfspaces $\{\xi \geq c\}$ with ξ close to ξ_0 and c close to c_0. One may think of ξ_0 as a non-zero linear functional on $L = \mathbb{R}^d$, or, for Euclidean spaces L, as the unit normal on ∂H_0 pointing into the halfspace H_0.

Lemma 5.11. *If $H_0 = \{\xi_0 \geq c_0\}$ is an interior point of $\mathcal{H}_0$ one may choose affine coordinates $\zeta_1, \dots, \zeta_d$ such that $H_0 = \{\zeta_1 + \cdots + \zeta_d \geq 1\}$ and such that $\mathcal{H}_0$ contains all halfspaces*

$$\{q_1\zeta_1 + \cdots + q_d\zeta_d \geq 0\}, \quad q_i \geq 0,\ q_1 + \cdots + q_d > 0.$$

If $\mathcal{H}_0$ is decreasing, i.e.

$$H_1 \text{ß} H_2 \in \mathcal{H}_0 \Rightarrow H_1 \in \mathcal{H}_0,$$

then $\mathcal{H}_0$ contains all halfspaces

$$\{q_1\zeta_1 + \cdots + q_d\zeta_d \geq c\}, \quad q_i \geq 0,\ q_1 + \cdots + q_d > 0,\ c \geq 0.$$

Proof. Suppose $H_0 = \{\xi_0 \geq c_0\}$. Assume $\mathcal{H}_0$ open. Choose $c_1 < c_0$ such that $H_1 = \{\xi_0 \geq c_1\} \in \mathcal{H}_0$. Choose a new origin in the hyperplane $\{\xi_0 = c_1\}$ and Euclidean coordinates such that $H_1 = \{\theta_0 \geq 0\}$ for some unit vector θ_0. There exists $\delta > 0$ such that $\{\theta \geq 0\} \in \mathcal{H}_0$ holds for all unit vectors θ with $\|\theta - \theta_0\| < \delta$. Choose unit vectors $\theta_1, \dots, \theta_d$, linearly independent, with $\|\theta_i - \theta_0\| < \delta$, such that $\theta_0 = c(\theta_1 + \cdots + \theta_d)$ for some $c > 0$. Then $H_0 = \{\theta_0 \geq \varepsilon\}$ for some $\varepsilon > 0$. So we set $\zeta_i = c\theta_i/\varepsilon$. □

Definition. Let ρ be a Radon measure on an open set Oß$\mathbb{R}^d$. A halfspace J_0 is *sturdy* if J_0ßO and $\rho(J_0)$ is finite, and if JßO and $\rho(J)$ is finite for all halfspaces J is a neighbourhood of J_0.

The set $\mathcal{H}(\rho)$ of sturdy halfspaces is open in $\mathcal{H}$. It has a simple structure.

Theorem 5.12. *Let ρ be a Radon measure on the open set Oß$L = \mathbb{R}^d$. There is an open cone ΔßL^* such that a halfspace in O is sturdy if and only if its direction lies in Δ. If O contains a sturdy halfspace, then Δ is non-empty, and the direction of any halfspace HßO of finite mass lies in the closure of Δ.*

Proof. If two halfspaces in O have finite mass, this also holds for all halfspaces in their union. As a result the set Δ_1 of all $\xi \neq 0$ for which there exists a halfspace $H = \{\xi \geq c\}$ßO, with $\rho(H) < \infty$, together with the zero functional, forms a cone in L^*. If $\xi_1, \dots, \xi_d$ in Δ_1 are independent, then, by treating these functionals as coordinates, we see that $S = \mathbb{R}^d \setminus O$ is contained in a translate $(-\infty, c)$ of the open negative orthant, and that the complement of $(-\infty, c)$ has finite mass. Hence for ξ in the interior of the cone generated by $\xi_1, \dots, \xi_d$ all halfspaces $H = \{\xi \geq c\}$ in O are sturdy. So Δ is the interior of Δ_1. If there is a sturdy halfspace J_0ßO the lemma above with $\mathcal{H}_0 = \mathcal{H}(\rho)$ shows that Δ is non-empty: it contains the interior of the cone generated by the coordinates $\zeta_1, \dots, \zeta_d$. □

Definition. The open cone ΔßL^* in Theorem 5.12 is the *intrusion cone*.

Let N be the standard Poisson point process on $(-\infty, 0)^3$. The mean measure ρ is *Lebesgue measure* on the negative octant. Let $O = \mathbb{R}^3$, and let $\zeta_1, \zeta_2, \zeta_3$ denote the standard coordinates. The halfspace $\{\zeta_3 \geq c\}$ contains no points for $c \geq 0$, and infinitely many points for $c < 0$. Similarly for the halfspace $\{\zeta_1 + \zeta_2 \geq c\}$. Only halfspaces $H = \{\xi \geq c\}$ with direction $\xi \in \Delta = (0, \infty)^3$ are able to *intrude* arbitrarily deeply into the point set N. For any $c < 0$ such a halfspace contains finitely many points a.s.

Proposition 5.13. *The map $H \mapsto \rho(H)$ on $\mathcal{H}(\rho)$ is continuous in H_0 if and only if $\rho(\partial H_0) = 0$.*

Proof. Introduce the coordinates $\zeta_1, \dots, \zeta_d$ of Lemma 5.11. Then the complement of $V = (-\infty, 0)^d$ has finite mass. Hence $H_n \to H_0$ implies $\rho(H_n) \to \rho(H_0)$ by Lebesgue's theorem on dominated convergence applied to the functions 1_{H_n} with the dominating function 1_{V^c}. □

For an open cone Λ in L^* define the set $\mathcal{H}^\Lambda$ß$\mathcal{H}$ by

$$\mathcal{H}^\Lambda = \{H \in \mathcal{H} \mid H = \{\xi \geq c\}\text{ß}O,\ \xi \in \Lambda,\ \xi \neq 0\}. \tag{5.10}$$

The arguments above show that $\mathcal{H}(\rho) = \mathcal{H}^\Delta$.

Proposition 5.14. *For an open cone* $\Lambda ß \Delta$, *the set* $\mathcal{H}^\Lambda$ *is open in* $\mathcal{H}$.

Proof. The map $H = \{\theta \geq c\} \to \theta \in \partial B$ is continuous. □

5.5 The convergence cone. In this section we assume that ρ_n, $n \geq 0$, are Radon measures on the open set $O ß L = \mathbb{R}^d$, and that $\rho_n \to \rho_0$ vaguely on O. We do not assume that $d\rho_n = 1_O n d\pi_n$, since we also want to apply the results to measures $d\rho_n = n\psi 1_O d\pi_n$ where ψ is a continuous positive loss function on O. Here we want to determine the halfspaces on which weak convergence holds. As in the previous section there is an open cone which plays a crucial role. Here it is the convergence cone Γ.

Recall that vague convergence $\rho_n \to \rho_0$ on an open set O implies weak convergence $1_F d\rho_n \to 1_F d\rho_0$ for closed sets $F ß O$ of finite mass $\rho_0(F)$, whose boundary carries no mass, provided $\rho_n(F) \to \rho_0(F)$, see Theorem 4.21.

Example 5.15. The condition $\rho_0(\partial F) = 0$ above is necessary. Let ρ_n on $\mathbb{R}$ be the sum of an exponential distribution with mean n on $(0, \infty)$ and an exponential distribution with mean $-1/n$ on $(-\infty, 0)$. Then $\rho_n \to \rho_0$ vaguely on $\mathbb{R}$, where ρ_0 is the probability measure concentrated in the origin. The halfspace $J_0 = [0, \infty)$ has finite mass, and $\rho_n(J_0) \to \rho_0(J_0)$, but $1_{J_0} d\rho_n$ does not converge (weakly or vaguely) to $1_{J_0} d\rho_0$. ◊

Let us first show that one may take the open set O on which vague convergence holds, maximal. Vague convergence is a local affair.

Proposition 5.16. *If* O *is covered by a family* $\mathcal{U}$ *of open sets and* $\rho_n \to \rho_0$ *vaguely on each* $U \in \mathcal{U}$ *then* $\rho_n \to \rho_0$ *vaguely on* O.

Proof. Let $\varphi\colon O \to \mathbb{R}$ be continuous with compact support K. Cover K by open balls centered in $z \in K$ with radius $r(z)$, so that the ball of radius $2r(z)$ is contained in a set $U \in \mathcal{U}$, and is relatively compact, and let $V_1, \dots, V_m$ be a finite subcover. There exists a continuous function $\eta\colon [0, 2] \to [0, 1]$ which vanishes in 2 and is one on $[0, 1]$. Hence there exist continuous ψ_i which are one on V_i and vanish outside the corresponding ball with the doubled radius. The sum ψ is at least one on K, and the functions $\varphi_i = \psi_i \varphi / \psi$ are continuous on O with sum φ. Since each φ_i is supported by a compact subset of some $U_i \in \mathcal{U}$ we have $\int \varphi_i d\rho_n \to \int \varphi_i d\rho_0$, $i = 1, \dots, m$. Now take the sum. □

Definition. Let $\mathcal{U} ß \mathcal{H}$ be open. Then $\rho_n \to \rho_0$ *weakly* on $\mathcal{U}$ if $\rho_0(H)$ is finite for all $H \in \mathcal{U}$, and if $\rho_n(H) \to \rho_0(H)$ for $H \in \mathcal{U}$ whenever $\rho_0(\partial H) = 0$.

We shall show that there exists a maximal open cone Γ in L^*, such that ρ_0 is finite on $\mathcal{H}^\Gamma$, see (5.10), and such that $\rho_n \to \rho_0$ weakly on $\mathcal{H}^\Gamma$.

Theorem 5.17. *Let $\rho_n \to \rho_0$ vaguely on the open set $O ß L = \mathbb{R}^d$. Assume O is maximal. Suppose there exists a non-empty open set $U ß \mathcal{H}$ on which $\rho_n \to \rho_0$ weakly. There exists a non-empty open cone $\Gamma ß L^*$ such that $\rho_n \to \rho_0$ weakly on $\mathcal{H}^\Gamma$, and such that the two conditions, $\rho_n(H) \to \rho_0(H)$ and $\rho_0(\partial H) = 0$, for a halfspace $H = \{\xi \geq c\} ß O$, imply that ξ lies in the closure of Γ.*

Proof. The proof is similar to that of Theorem 5.12. We here define Γ_1 as the set of ξ for which there exists a halfspace $H = \{\xi \geq c\} ß O$ with $\rho(\partial H) = 0$ and $\rho_n(H) \to \rho(H) < \infty$, and check that Γ_1 is a cone, if we add the zero functional. □

Definition. The open cone $\Gamma ß L^*$ in Theorem 5.17 is called the *convergence cone* associated with the sequence of measures ρ_n, $n \geq 0$, on O.

The three propositions below give conditions which ensure that the cone Γ is non-empty.

Proposition 5.18. *Let $H_i = \{\xi_i \geq c_i\}$ for $i = 1, \dots, d$, with $\xi_1, \dots, \xi_d$ independent. Suppose*

$$\rho_n(H_i) \to \rho_0(H_i) < \infty, \quad \rho_0(\partial H_i) = 0, \ i = 1, \dots, d.$$

Then $\rho_n \to \rho_0$ weakly on $\mathcal{H}^\Lambda$, where Λ is the interior of the cone generated by $\xi_1, \dots, \xi_d$.

Proof. The result follows from the next proposition applied to the open convex set $V = \{\xi_1 < c_1, \dots, \xi_d < c_d\}$. □

Recall that for an open convex set V, the cone V^+ in L^*, defined in (5.9), consists of all linear functionals which are bounded above on V. If V contains no lines then V^+ has non-empty interior by Theorem 5.10.

Proposition 5.19. *Let V be an open convex set which does not contain a line. Let $V^c ß O$, $\rho_0(\partial V) = 0$, and $\rho_n(V^c) \to \rho_0(V^c) < \infty$. Then $\rho_n \to \rho_0$ weakly on $\mathcal{H}^\Lambda$, and $1_{V^c} d\rho_n \to 1_{V^c} d\rho_0$ weakly on $\mathbb{R}^d$.*

Proof. The result follows from the next proposition. □

Choose coordinates so that the origin lies in V. Then the sets nV cover $\mathbb{R}^d$. Weak convergence $1_{V^c} d\rho_n \to 1_{V^c} d\rho_0$ implies tightness: For any $\varepsilon > 0$ there exists $r > 1$ such that $\rho_n(rV^c) < \varepsilon$ for $n \geq n_0$. The convergence condition in the proposition above may be replaced by a *tightness* condition.

Proposition 5.20. *Let V be an open convex set which contains no lines. Let the following tightness condition hold: For each $\varepsilon > 0$ there exist $r \geq 1$ and $p \in \mathbb{R}^d$ such that*

$$W = rV + p \supset S := \mathbb{R}^d \setminus O,$$

and $\rho_n(W) < \varepsilon$ for $n \geq n_0$. Then $\rho_n \to \rho_0$ weakly on $\mathcal{H}^\Lambda$ in (5.10).

Proof. Observe that $W^+ = V^+$. Assume the tightness condition holds. Let $\varepsilon > 0$. By blowing up W slightly from an interior point we may ensure that $\rho_0(\partial W) = 0$. By vague convergence $\rho_0(\mathrm{int}(W^c)) \leq \varepsilon$. Let $H = \{\xi \geq c\}ßO$ with $\xi \in \Lambda$, $\xi \neq 0$ and $\rho_0(\partial H) = 0$. Eventually

$$\rho_n(H) = \int \chi d\rho_n + \rho_n H \setminus W \leq \int \chi d\rho_0 + \varepsilon + \rho_n(W^c) \leq \rho_0(H) + 2\varepsilon$$

since the function $\chi = 1_{H \cap W}$ is ρ_0-a.e. continuous with compact support. Vague convergence implies $\int \chi d\rho_n \to \int \chi d\rho_0$. □

For increasing functions weak convergence $\varphi_n \to \varphi_0$ on an open interval (a, b) implies

$$\varphi_n(x_n) \to \varphi_0(x_0), \quad x_n \to x_0,\ x_0 \in (a, b),\ \varphi_0(x_0 - 0) = \varphi_0(x_0 + 0).$$

A similar result holds for weak convergence $\rho_n \to \rho_0$ on open sets of halfspaces in $\mathcal{H}_O = \{H \in \mathcal{H} \mid HßO\}$.

Proposition 5.21. *Suppose $\rho_n \to \rho_0$ weakly on the open set* $\mathrm{U}ß\mathcal{H}_O$. *If $H_n \to H_0 \in \mathrm{U}$ and $\rho_0(\partial H_0) = 0$, then $\rho_n(H_n) \to \rho_0(H_0)$, and $1_{H_n} d\rho_n \to 1_{H_0} d\rho_0$ weakly on $\mathbb{R}^d$.*

Proof. By Lemma 5.11 one may choose coordinates $\zeta_1, \dots, \zeta_d$ such that $H_0 = \{\zeta_1 + \cdots + \zeta_d \geq 1\}$ and such that $\{\zeta_i \geq 0\} \in \mathrm{U}$. One may choose $\varepsilon \in (0, 1/d)$ so small that

$$J_i = \{\zeta_i \geq \varepsilon\} \in \mathrm{U}, \quad \rho_0(\partial J_i) = 0.$$

Hence $\rho_n \to \rho_0$ weakly on $\mathcal{H}^V$ where $V = \{\zeta_1 < \varepsilon, \dots, \zeta_d < \varepsilon\}$, and $1_{V^c} d\rho_n \to 1_{V^c} d\rho_0$ weakly on $\mathbb{R}^d$. Now apply the Weak Convergence Theorem, Theorem 4.21, to the functions $\varphi_n = \psi 1_{H_n}$ where ψ is a bounded continuous function on $\mathbb{R}^d$. If $x_n \to x_0 \notin \partial H_0$, then convergence $H_n \to H_0$ implies $\varphi_n(x_n) \to \varphi_0(x_0)$. See Exercise 5.5. □

Let us show how these results apply to loss functions.

Example 5.22. Let $Z_1, Z_2, \dots$ be independent observations from the distribution π on $\mathbb{R}^d$. Let μ_0 be a Radon measure on the open set $Oß\mathbb{R}^d$, and suppose

$$d\mu_n = 1_O n d\pi_n \to d\mu_0 \text{ vaguely on } O, \quad \pi_n = \alpha_n^{-1}(\pi),\ \alpha_n \in \mathcal{A}.$$

Let $d\rho_n = \psi d\mu_n$ for $n \geq 0$ where ψ is a continuous positive function on O. If the tightness condition of Proposition 5.20 holds for ρ_n, and if $V^c ßO$, then for any sequence of halfspaces $J_n ßV^c$, $n \geq 0$, with $J_n \to J_0$, and $\mu_0(J_0) = 0$

$$\mathbb{E} \sum_{k=1}^{n} \varphi(\alpha_n^{-1}(Z_k)) 1_{\{Z_k \in H_n\}} \to \int_{J_0} \varphi d\mu_0,$$

where $H_n = \alpha_n(J_n)$, and φ is a ρ_0-a.e. continuous Borel function on O such that φ/ψ is bounded. ◊

5.6* The support function. *Support processes* give an elegant analytic description of the convex hulls of point processes. We prefer a more geometric approach via halfspaces. Hence we restrict ourselves here to two simple results.

For a non-empty convex open set $U \text{ß} L = \mathbb{R}^d$ in $\mathcal{C}$ we introduce a cone $\mathcal{F}_U$ of affine functions.

$$\mathcal{F}_U = \{\varphi \mid \varphi(U) \text{ß} (-\infty, 0]\} \text{ß} \mathcal{F}.$$

This cone is related to various dual convex sets for U in L^*. Recall that $\mathcal{F}$ is the linear space of affine functions on L. The cone U^+ in (5.9) is the image of $\mathcal{F}_U$ under the projection $\varphi = \xi - c \mapsto \xi$. If $\xi - c \in \mathcal{F}_U$ then $\xi - t \in \mathcal{F}_U$ for $t > c$. It follows that there is a unique convex *positive-homogeneous* function $\tau \colon U^+ \to \mathbb{R}$ such that $\mathcal{F}_U$ is the epigraph of τ:

$$\varphi = \xi - c \in \mathcal{F}_U \iff c \geq \tau(\xi).$$

Definition. The function τ is the *support function* of the convex set U, and U^+ is its domain. The halfspace $H = \{\xi \geq c\}$ *supports* the convex open set U if $c = \tau(\xi)$ is the minimal value of t for which $\{\xi \geq t\}$ is disjoint from U. For a counting measure ν on $\mathbb{R}^d$, we define τ_ν as the support function of the interior of the convex hull $c(\nu)$ of the support of ν. Its domain is $c(\nu)^+$.

Proposition 5.23. *Let $\nu_0, \nu_1, \ldots$ be counting measures on L such that $\nu_n \to \nu_0$ vaguely, and $c(\nu_0) \in \mathcal{C}$. Let τ_n be the support function of ν_n for $n \geq m_0$. Let the elements $\zeta_1, \ldots, \zeta_d \in L^*$ be linearly independent, and let Λ be the open cone of all positive linear combinations $r_1\zeta_1 + \cdots + r_d\zeta_d$ with $r_i > 0$. If $\tau_n(\zeta_i)$ is bounded for $n \geq n_i$ for $i = 1, \ldots, d$, then $\tau_n(\xi_n) \to \tau_0(\xi_0)$ for $\xi_n \to \xi_0 \in \Lambda$.*

Proof. Regard the $\zeta_1, \ldots, \zeta_d$ as coordinates. Then $\Lambda = (0, \infty)^d$. Let $\tau_n(\zeta_i) \leq c$ for $n \geq n_0$. Suppose the halfspace H has direction $\theta \in (0, \infty)^d$, and $\nu_0(H) = 0$. By vague convergence $\nu_n(H) = \nu_n(H \cap (-\infty, c]^d) = 0$ eventually, since the closure K of $H \cap (-\infty, c]^d$ is compact and $\nu_0(\partial K) = 0$. So $\tau_n \to \tau_0$ on $(0, \infty)^d$. Support functions are convex. So convergence is uniform on compact subsets of $(0, \infty)^d$. □

Given $\xi_0 \neq 0$ in L^*, we say that the support functions τ_n *converge in the direction* ξ_0 if $\tau_n(\xi_0) \to \tau_0(\xi_0)$. Now assume more:

$$\xi_n \to \xi_0 \Rightarrow \tau_n(\xi_n) \to \tau_0(\xi_0). \tag{5.11}$$

This limit relation says that the direction ξ_0 is robust: if H_0 is a horizontal supporting halfspace to ν_0, then $H_n \to H_0$ for any sequence of supporting halfspaces to ν_n which are asymptotically horizontal. The assumption (5.11) has an unexpected implication.

Proposition 5.24. *Suppose* (5.11) *holds. Then there exists a neighbourhood U of ξ_0 such that τ_0 is finite on U, and $\tau_n \to \tau_0$ uniformly on U.*

Proof. The conditions imply that τ_n, $n \geq n_0$, is uniformly bounded on a neighbourhood V of ξ_0 for some n_0. (Otherwise there is a sequence of points ξ_n with $\|\xi_n - \xi_0\| < 1/n$, and indices $k_1 < k_2 < \cdots$ such that $\tau_{k_n}(\xi_n) > n$.) Choose $\zeta_1, \ldots, \zeta_d$ in V, linearly independent, such that ξ_0 lies in the interior V_0 of the cone generated by these d vectors. Choose a neighbourhood U of ξ_0 whose closure lies in V_0, and apply the previous proposition. □

5.7 Almost-sure convergence of the convex hulls. We are now ready to prove the two main results, convergence of the convex hulls, and of the stochastic loss integrals. We start with a preliminary result to show how *Skorohod's Representation Theorem* is applied.

The limit process N_0 lives on $O = \mathbb{R}^d \setminus S$. If S is non-empty one needs a weak boundary condition to ensure that the convex hulls $N_n(\omega)$ contain the hole S in the limit:

(S) There is a closed set S_0ßS such that Sß$c(S_0)$, and

$$n\pi_n(U) \to \infty, \quad U \text{ open}, \ U \cap S_0 \neq \emptyset. \tag{5.12}$$

Proposition 5.25. *Let ρ_0 be a Radon measure on the open set Oß$\mathbb{R}^d$. Let π_n be probability measures on $\mathbb{R}^d$ such that* (5.1) *holds. Let V_i, $i \in I$ß$\{1, 2, \ldots\}$ be a decreasing family of convex open sets containing $S = \mathbb{R}^d \setminus O$ such that*

$$\rho_0(\partial V_i) = 0, \quad \rho_n(V_i^c) \to \rho_0(V_i^c) < \infty, \quad i \in I.$$

Let $\mathcal{H}_0$ be the collection of halfspaces in O which intersect one of the sets V_i in a bounded set. There exist n-point sample clouds N_n from the distribution π_n, and a Poisson point process N_0 on O with mean measure ρ_0, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that for every $\omega \in \Omega$

$$N_0(\omega) \text{ is a Radon measure on } O,$$

$$1_O dN_n(\omega) \to dN_0(\omega) \text{ vaguely on } O,$$

$$N_n(\omega)(V_i) \to N_0(\omega)(V_i), \quad i \in I,$$

$$N_n(\omega)(H) \to N_0(\omega)(H), \quad H \in \mathcal{H}_0, \ N_0(\omega)(\partial H) = 0.$$

If $O = \mathbb{R}^d$ or if the condition (S) *in* (5.12) *holds, one may choose the point processes to satisfy:*

$$N_n(\omega)(U) \to \infty, \quad U \cap S_0 \neq \emptyset, \ U \text{ open},$$

$$\mathrm{cl}(\liminf c(N_n(\omega))) \supset c(N_0(\omega) \cup S),$$

$$T_n(\omega)(\xi_n) \to T_0(\omega)(\xi_0), \quad \xi_0 \in \Lambda, \ \xi_n \to \xi_0 \neq 0,$$

where cl(A) *is the closure of* A*,* $c(A)$ *the closed convex hull,* T_n *the support process of* N_n*, and* Λ *the interior of the cone* $\bigcup V_i^+$.

Proof. Let U_i, $i \in I$, be a countable open basis for $\mathbb{R}^d$, and U_j, $j \in J$, those sets which intersect S_0. It suffices to show that $N_n(\omega)(U_j) \to \infty$ for $j \in J$. Let M_n be an n-point sample cloud from the distribution π_n, and M_0 the Poisson point process on ρ. Then $1_O dM_n \Rightarrow dM_0$ vaguely on O. The conditions on the convex sets V_i imply that $M_n(V_i^c) \Rightarrow M_0(V_i^c) \in \{0, 1, 2, \dots\}$. We apply Skorohod's Representation Theorem, Theorem 4.9, to the sequence (X_n, Y_n) in $\mathcal{X} \times \mathcal{Y}$ with

$$\mathcal{X} = \mathcal{N}(O) \times \{0, 1, \dots\}^{\mathbb{N}} \times \{0, 1, \dots, \infty\}^J, \quad \mathcal{Y} = \mathcal{N}(\mathbb{R}^d),$$

with $\mathbb{N} = \{1, 2, \dots\}$, and $Y_n = M_n$ for $n \geq 1$, $Y_0 = 0$, and

$$X_n = (1_O dM_n, (M_n(V_i^c), i \in I), (M_n(U_j), j \in J)),$$
$$X_0 = (M_0, (M_0(V_i^c), i \in I), (\infty, \infty, \dots)).$$

Let $N_n, n \geq 0$, be representatives which yield convergence for all ω. We may assume that $N_0(\omega)(\partial V_i) = 0$ for $i = 1, 2, \dots$ and all ω, and $N_0(\omega)(V_i^c) < \infty$. Vague convergence $N_n(\omega) \to N_0(\omega)$ on O together with $N_n(\omega)(V_i^c) \to N_0(\omega)(V_i^c) < \infty$ and $N_0(\omega)(\partial V_i) = 0$ gives weak convergence $1_{V_i^c} dN_n(\omega) \to 1_{V_i^c} dN_0(\omega)$. This in turn gives weak convergence on every halfspace H which is disjoint from one of the convex sets V_i, and whose boundary does not contain a point of $N_0(\omega)$, and on every halfspace $H = \{\xi \geq c\}$ßO with $\xi \in \Lambda$.

The remaining relations are proved similarly. The closure of $\liminf c(N_n(\omega))$ contains $N_0(\omega)$ by vague convergence, and contains S since $N_n(\omega)(U_j) > 0$ eventually for each $j \in J$. □

Now assume the *intrusion cone* Δ and *convergence cone* Γ are non-empty and coincide.

Theorem 5.26 (Convergence of Convex Hulls). *Let* π_n *be probability distributions on* $\mathbb{R}^d$*, and* ρ_0 *a Radon measure on the open set* $O = \mathbb{R}^d \setminus S$*. Assume* $O = \mathbb{R}^d$ *or condition* (S) *holds, see* (5.12)*,* $1_O n d\pi_n \to d\rho_0$ *vaguely on* O*, and* $\Gamma = \Delta \neq \emptyset$*. Then there exist* n*-point sample clouds* N_n *from the distribution* π_n*, and a Poisson point process* N_0 *on* O *with mean measure* ρ_0*, defined on a common probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, such that for* $\omega \in \Omega$ *the convex hulls converge:*

$$c(N_n(\omega)) \to c(N_0(\omega) \cup S).$$

Moreover the support processes converge: for $\xi_n \to \xi_0 \in \Delta$*,* $\xi_0 \neq 0$*,*

$$T_n(\omega)(\xi_n) \to T_0(\omega)(\xi_0).$$

If H_nßO *for* $n \geq 0$, $H_n \to H_0 = \{\xi_0 \geq c_0\}$, *with* $\xi_0 \in \Delta$, *and if* $N_0(\omega)(\partial H_0) = 0$, *then*

$$1_{H_n} dN_n(\omega) \to 1_{H_0} dN_0(\omega) \ \textit{weakly}.$$

Proof. Let $\theta_1, \theta_2, \dots$ be dense in $\partial B \setminus \mathrm{cl}(\Delta)$. Then $\rho_0\{\theta_i \geq m\} = \infty$ for $m > s(\theta_i) = \sup \theta_i(S)$, and we may choose Ω so that

$$N_0(\omega)\{\theta_i \geq m\} = \infty, \quad i \geq 1,\ m > s(\theta_i),\ \omega \in \Omega. \tag{5.13}$$

Choose $\xi_i \in \partial B \cap \Delta$, $i = 1, 2, \dots$ dense and $c_i > 0$ so that $H_i = \{\xi_i \geq c_i\}$ßO, $\rho_0(\partial H_i) = 0$. Set $V_i = (H_1 \cup \dots \cup H_i)^c$. Then $\rho_0(H_i) < \infty$, and $\xi_i \in \Gamma$ implies $\rho_n(V_i^c) \to \rho_0(V_i^c)$. We may choose N_n so that $M_n = 1_O dN_n$ satisfies

$$M_n(\omega)(V_i^c) \to M_0(\omega)(V_i^c), \quad i \geq 1,\ \omega \in \Omega. \tag{5.14}$$

Let $\Gamma(\omega)$ and $\Delta(\omega)$ be defined in terms of the Radon measures $M_n(\omega)$ on O. By assumption $M_n(\omega) \to M_0(\omega) = N_0(\omega)$ vaguely on O for each $\omega \in \Omega$, and $N_n(\omega)(H) \to \infty$ for $H = \{\theta_i \geq c\}$ß$\mathbb{R}^d$, $i = 1, 2, \dots$ and $c > 0$. Now observe

$$\Gamma ß \Gamma(\omega) ß \Delta(\omega) ß \Delta, \quad \omega \in \Omega$$

by (5.14), by definition, and by (5.13) (if Δ is dense then $\Delta = L^*$). Therefore $\Gamma = \Delta$ implies $\Gamma(\omega) = \Delta(\omega) = \Delta$ for all ω. The Separation Theorem 5.10 gives $c(N_n(\omega)) \to c(N_0(\omega) \cup S)$, since any $w \notin c(N_0(\omega) \cup S)$ is contained in a halfspace $H = \{\xi \geq c\}$ßO with $\xi \in \Delta$, and we may choose c so that $N_0(\omega)(H) = 0$. Then $N_n(\omega)(H) = 0$ eventually. □

Let H_nßO, $n \geq 0$, be closed halfspaces, $H_n \to H_0$, and let $\varphi\colon O \to \mathbb{R}$ be a Borel function. We shall formulate conditions which ensure that H_0 is *steady* and

$$\int_{H_n} \varphi dN_n \to \int_{H_0} \varphi dN_0 \quad \text{a.s. and in } \mathbf{L}^1 \tag{5.15}$$

for appropriately chosen n-point sample clouds N_n from π_n and a Poisson point process N_0 with mean measure ρ_0.

Theorem 5.27. *Let* (5.1) *hold, and let* $\psi \geq 1$ *be continuous on* O. *Let* Λ *be the interior of the cone generated by* d *independent linear functionals* $\zeta_1, \dots, \zeta_d$ *which satisfy the tightness condition:*

For each $\varepsilon > 0$ *there exists* $c \in \mathbb{R}$ *and an index* n_0 *such that*

$$\int_{\{\zeta_i \geq c\}} \psi d\rho_n < \varepsilon, \quad i = 1, \dots, d,\ n \geq n_0.$$

Then (5.15) *holds for* $H_0 = \{\xi_0 \geq c_0\}$ßO *provided* $\rho_0(\partial H_0) = 0$, φ/ψ *is bounded on* O, φ *is* ρ_0*-a.e. continuous on* O, *and* $\xi_0 \in \Lambda$.

Proof. Since Λ is open there exist independent linear functionals $\xi_1, \dots, \xi_d$ in Λ such that $\xi_0 = \xi_1 + \cdots + \xi_d$. Proposition 5.20 applied to $d\mu_n = \psi d\rho_n$ rather than $d\rho_n$ implies that $\mu_n \to \mu_0$ weakly on $\mathscr{H}^{\Lambda_0}$ where Λ_0 is the interior of the cone generated by $\xi_1, \dots, \xi_d$. Regard the $\xi_1, \dots, \xi_d$ as coordinates. Choose a point a such that $(-\infty, a) = \bigcap\{\xi_i < a_i\}$ contains the complement of O, and $\mu_0(\partial(-\infty, a)) = 0$. In Proposition 5.25 take $V_i = V = (-\infty, a) \cap \{\xi < c\}$ with $c < c_0$ chosen so that $\mu_0\{\xi_0 = c\} = 0$ and $\{\xi_0 \geq c\}\ss O$. Then $N_n \to N_0$ holds a.s. on V^c, and by the ρ_0-a.e. continuity (5.15) holds a.s. As in Proposition 5.25 one may use Skorohod's theorem to obtain

$$X_n(\omega) := \int_{V^c} \psi dN_n(\omega) \to \int_{V^c} \psi dN_0(\omega).$$

Since the expectations converge, Lebesgue's theorem on dominated convergence, with dominating sequence X_n, yields $\mathbf{L}^1$-convergence in (5.15). □

5.8 Convergence to the mean measure. Estimating the mean measure on a half-space H, or the integral of a loss function $\int_H \varphi d\rho_0$, is straightforward provided the sample is large. In order to produce arbitrarily large samples, we introduce a time coordinate and consider the point processes with points $(W_{ni}, i/n)$, where $W_{n1}, W_{n2}, \dots$ is a sequence of independent observations from the distribution π_n. By Theorem 2.21 these converge on $O \times [0, \infty)$ to a Poisson point process with mean measure $d\rho_0(w)dt$ if (5.1) holds. We shall look at sample clouds on $\mathbb{R}^d$ with $m_n \gg n$ points, corresponding to time horizons $t_n = 1/\varepsilon_n \to \infty$. Let $\widetilde{N}_n$ be the m_n-point sample cloud from the distribution π_n. The random approximation R_n to the Radon measure $d\rho_n = 1_O n d\pi_n$ is defined as

$$dR_n(\omega) = \varepsilon_n 1_O d\widetilde{N}_n(\omega), \quad n = 1, 2, \dots. \tag{5.16}$$

The random measure R_n has the same mean measure ρ_n as N_n on O. It has more points, but these carry less weight. As a result the variance of stochastic integrals is smaller. By (2.4)

$$\begin{aligned} \mathbb{E}\Big(\int_H \varphi dR_n\Big) &= \int_H \varphi d\rho_n \\ \operatorname{var}\Big(\int_H \varphi dR_n\Big) &= \varepsilon_n\Big(\int_H \varphi^2 d\rho_n - \frac{1}{n}\Big(\int_H \varphi d\rho_n\Big)^2\Big). \end{aligned} \tag{5.17}$$

Theorem 5.28 (Consistency). *If $\varepsilon_n \to 0$, then $R_n \to \rho_0$ vaguely on O in probability. Let $H\ss O$ be a halfspace such that $\rho_n(H) \to \rho_0(H) < \infty$ and $\rho_0(\partial H) = 0$. Then*

$$\int_H \varphi dR_n \to \int_H \varphi d\rho_0 \quad \textit{weakly in probability}$$

for any bounded ρ_0-a.e. continuous function φ on H.

Proof. Let $\varphi_0 : O \to \mathbb{R}$ be continuous with compact support. We have to show that $\int \varphi dR_n \to \int \varphi d\rho_0$ in probability. In fact convergence holds in $\mathbf{L}^2(\mathbb{P})$ by (5.17). The proof of the second part is similar. □

We now turn to the *loss integral*. The loss integral has the form

$$\int_J \varphi dR_n = \varepsilon_n \sum \{\varphi(W_{ni}) \mid W_{ni} \in J,\ 1 \le i \le m_n\}.$$

In general loss functions need not be bounded. We assume that φ is ρ_0-a.e. continuous, and that φ^2 is ρ_0-integrable over the halfspace J with integral $\int_J \varphi^2 d\rho_0 = \sigma^2 < \infty$.

Define the error of the stochastic loss integral by

$$E_n = \int_J \varphi dR_n - \int_J \varphi d\rho_0. \tag{5.18}$$

Under suitable conditions the error is asymptotically Gaussian $N(0, \varepsilon_n \sigma^2)$.

In order to understand the asymptotic behaviour of the error E_n we need to study $dR^\varepsilon = \varepsilon dM^\varepsilon$ for $\varepsilon \to 0$, where M^ε is the Poisson point process with mean measure ρ/ε on H. It will be convenient to use a different normalization.

Definition. A *Gaussian space* $\mathcal{G}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a separable closed linear subspace of $\mathbf{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ consisting of Gaussian variables (modulo null variables). Typically $\mathcal{G}$ is the set of all elements $\sum c_n U_n$ with $U_1, U_2, \dots$ independent standard normal variables and $\sum c_n^2 < \infty$. Let ρ be a Radon measure on the open set Oß$\mathbb{R}^d$. A *Gaussian field* on O with variance ρ is a linear isometry $W : \mathbf{L}^2(d\rho) \to \mathcal{G}$, where $\mathcal{G}$ is a Gaussian space.

For $\varphi \in \mathbf{L}^2(d\rho)$ define the *compensated integral* $S_\varepsilon(\varphi) = \sqrt{\varepsilon} \int \varphi(dN^\varepsilon - d\rho/\varepsilon)$. The random variable $S_\varepsilon(\varphi)$ is centered with variance $\int \varphi^2 d\rho$, see Section 2.5. So the linear map $S_\varepsilon : \mathbf{L}^2(d\rho) \to \mathbf{L}^2(d\mathbb{P})$ also is an isometry.

Theorem 5.29. *With the notation above $S_\varepsilon(\varphi) \Rightarrow W(\varphi)$ for $\varphi \in \mathbf{L}^2(d\rho)$, $\varepsilon \to 0+$.*

Proof. If φ is the indicator function of a Borel set E, then one may write $S_\varepsilon(\varphi) = \sqrt{\varepsilon}(N^\varepsilon(E) - \rho(E)/\varepsilon)$ with $N^\varepsilon(E)$ a Poisson variable with expectation $\rho(E)/\varepsilon$. In this case the theorem follows by the asymptotic normality of Poisson variables. By independence $(S_\varepsilon(E_1), \dots, S_\varepsilon(E_m)) \Rightarrow (W(E_1), \dots, W(E_m))$ for disjoint Borel sets E_i with finite measure. Hence the theorem holds for simple functions φ. These are dense in $\mathbf{L}^2(d\rho)$. Now apply Lemma 4.10. □

In a similar way one proves convergence of the finite-dimensional distributions. Here we only want to show that under appropriate conditions one may conclude that $\int_H \varphi dR_n$ is close to $\int_H \varphi d\rho_0$.

Proposition 5.30. *Let r_0 be a continuous positive function on the open set Oß$\mathbb{R}^d$, and let π_n be probability measures on $\mathbb{R}^d$ with densities f_n such that*

$$nf_n(w_n) \to r_0(w_0), \quad w_n \to w_0,\ w_0 \in O.$$

Let JßO be a halfspace such that

$$\int_J |nf_n(w) - r_0(w)|dw = o(\delta_n),$$

where $\delta_n \to 0$ so slowly that $n\delta_n \to \infty$. Let $m_n/n = 1/\varepsilon_n$ with $\varepsilon_n \asymp \delta_n$, and define the random measures R_n by (5.16). Let φ be a Borel function on O such that the integrals

$$\int_J \varphi(w)r_0(w)dw = \mu, \quad \int_J \varphi^2(w)r_0(w)dw = \sigma^2 < \infty$$

converge. Let U be standard normal. Then

$$\frac{1}{\sqrt{\varepsilon_n}} E_n \Rightarrow \sigma U, \quad E_n = \int_J \varphi dR_n - \mu.$$

Proof. The convergence $nf_n \to r_0$ above is a simple sufficient condition for vague convergence, (5.1). Let $\widetilde{N}_n$ be the m_n-point sample cloud from the distribution π_n. There exist Poisson point processes on J, $\widetilde{M}_n$ with mean measure ρ_n/ε_n, and $\widetilde{M}^{\varepsilon_n}$ with mean measure ρ_0/ε_n, such that

$$\mathbb{P}\{1_J d\widetilde{N}_n \neq d\widetilde{M}_n\} \le m_n \pi_n(J)^2 = (m_n/n^2)\rho_n(J) = o(1),$$
$$\mathbb{P}\{\widetilde{M}_n \neq \widetilde{M}^{\varepsilon_n}\} \le \int_J |nf_n(w) - r_0(w)|dw/\varepsilon_n = o(1).$$

By Theorem 5.29 $S_\varepsilon(\varphi) \Rightarrow W(\varphi)$ holds with $d\rho = r_0 dw$; equivalently

$$\frac{1}{\sqrt{\varepsilon_n}}\Big(\int_J \varphi dR_n^{\varepsilon_n} - \int_J \varphi d\rho_0\Big) \Rightarrow \sigma U,$$

where $R_n^{\varepsilon_n} = \varepsilon_n \tilde{M}^{\varepsilon_n}$. The inequalities above now yield the desired result. □

Corollary 5.31. *If $\sigma_n^2 = \int_J \varphi^2 d\rho_n$ is bounded, then $\mathrm{var}(E_n) = \varepsilon_n(\sigma_n^2 + o(1))$ and $\mathbb{E}(E_n) \to \mu$.*

Proof. The result follows from the lemma below. □

Lemma 5.32. *Let ρ_n, $n \ge 0$, be finite measures on a halfspace H such that $\rho_n \to \rho_0$ weakly. Let $\psi \ge 1$ be continuous on H. If $\int \psi^2 d\rho_n$ is bounded for $n \ge 1$ then $\int \psi^2 d\rho_0$ is finite and*

$$\int \varphi d\rho_n \to \int \varphi d\rho_0$$

for every ρ_0-a.e. continuous function φ for which φ/ψ is bounded.

Proof. First assume φ_0 is continuous and $0 \le \varphi_0 \le \psi$. Let $m \ge 1$. By weak convergence $\int \varphi_0 \wedge m d\rho_n \to \int \varphi_0 \wedge m d\rho_0$, and

$$(\varphi_0 - m)_+ d\rho_n \le \frac{1}{2m} \int_{\{\psi \ge m\}} \psi^2 - m^2 d\rho_n \le \int \psi^2 d\rho_n / 2m < \varepsilon$$

if m is large. Hence $d\mu_n = \psi d\rho_n \overset{\mathrm{d}}{\to} \mu_0 = \psi d\rho_0$ weakly, and the result follows since $\varphi d\rho_n = (\varphi/\psi) d\mu_n$ and φ/ψ is bounded and ρ_0-a.e. continuous. □

So far our setting has been quite general. We start with a sequence of probability measures π_n and assume that the measures $n\pi_n$ converge vaguely to a Radon measure on an open set Oß$\mathbb{R}^d$. In actual fact the probability measures π_n are related. We start with an iid sequence of observations $Z_1, Z_2, \ldots$ from an unknown probability measure π on $\mathbb{R}^d$. The probability measure π_n is the distribution of the normalized vector $\alpha_n^{-1}(Z_1)$ where $\alpha_n(w) = A_n w + b_n$ is an invertible affine transformation. The limit measure will satisfy certain symmetry relations. These make possible a reduction of the variance.

A simple example (for light tails) is the limit measure

$$d\rho(w) = d\rho^*(u) e^{-v} dv, \quad w = (u, v) \in \mathbb{R}^{h+1}.$$

Here ρ^* is a finite measure on $\mathbb{R}^h$, the *spectral measure*. The symmetries are vertical translations $\gamma^t(u, v) = (u, v + t)$, and $\gamma^t(\rho) = e^t \rho$ for $t \in \mathbb{R}$. The symmetry relations, together with the spectral measure, determine the measure ρ. So it suffices to estimate ρ^*.

A typical loss function is $\varphi(u, v) = \|u\| v^m$ on the upper halfspace $H_+ = \{v \ge 0\}$, with $m > 1$. (The loss indexed by the vertical coordinate v is compounded by large deviations of the horizontal coordinate.) The *integral* of this *loss* function with respect to the limiting Poisson point process N over the upper halfspace has variance

$$\mathrm{var}\left(\int_{H_+} \varphi dN \right) = \int_{H_+} \varphi^2 d\rho = (2m)! \int u^T u d\rho^*(u).$$

In general the spectral measure ρ^* is not known, but one assumes that the symmetry relations hold. In that case the loss may be estimated by $\int \|u\| dN^* \int_0^\infty v^m e^{-v} dv$, where N^*ß$\mathbb{R}^h$ is the projection of $1_{H_+} dN$ on the horizontal coordinate space. The variance of this estimate is smaller: $(m!)^2 \int u^T u d\rho^*(u)$.

In the remaining chapters we consider probability distributions π on $\mathbb{R}^d$ which may be normalized so that $n\alpha_n^{-1}(\pi) \to \rho$ vaguely on an open set Oß$\mathbb{R}^d$, where ρ is an excess measure.

II Maxima

This chapter shows how the limit theories for maxima and exceedances relate in the univariate case and contains an overview of multivariate max-stability. We assume some familiarity with extreme value theory. There are several good expositions. Our aim here is to highlight a number of issues which play a role in the multivariate geometric theory which we develop in Chapter III and IV. See Embrechts, Klüppelberg & Mikosch [1997] (EKM) or Resnick [1987] for details, or de Haan & Ferreira [2006]. We concentrate on the probability theory. For the statistical side see Beirlant et al. [2004], Coles [2001], Falk, Hüsler & Reiss [2004] and Joe [1997]. For examples of extreme value distributions, see Kotz & Nadarajah [2000]. For applications to finance and risk management see McNeil, Frey & Embrechts [2005], Embrechts [2004], Malevergne & Sornette [2006] and Moix [2001]. For applications to extreme natural events see Salvadori et al. [2007] and Finkenstädt & Rootzén [2004].

6 The univariate theory: maxima and exceedances

In the univariate setting the limit theories for maxima and for exceedances are two sides of the same coin.

6.1 Maxima. Sums of iid random variables lead to the Central Limit Theorem, infinitely divisible (id) and stable distributions, Brownian motion and Lévy processes. For the sequence of partial maxima

$$Y^{\vee n} := Y_1 \vee \cdots \vee Y_n = \max\{Y_1, \ldots, Y_n\} \tag{6.1}$$

of an iid sequence of random variables similar concepts have been developed. Here distribution functions (dfs) play a central role, since for independent variables X and Y with dfs F and G the maximum $X \vee Y$ has df FG:

$$\mathbb{P}\{X \vee Y \le t\} = \mathbb{P}\{X \le t, Y \le t\} = \mathbb{P}\{X \le t\}\mathbb{P}\{Y \le t\} = F(t)G(t) \tag{6.2}$$

by independence. Every random variable is *max-id* since $F^{1/n}$ is a df for any df F, and $(F^{1/n})^n = F$. In the limit theory for partial maxima there is no analog of the Gaussian distribution which, like a black hole, attracts all but the extremely heavy tailed distributions.

Fisher & Tippett [1928] determined the possible limit laws for the sequence of maxima in (6.1). There are three classes of limit laws. These limit laws are called the *max-stable* or *extreme value* distributions. They may be standardized to form a continuous one-parameter family of distributions on $\mathbb{R}$. For the heavy tailed and the bounded limit laws Gnedenko [1943] gave a simple characterization of the domains of attraction in terms of *regular variation* of the distribution tail; for the *double-exponential* or *Gumbel* law the domain of attraction was determined twenty years later in de Haan [1970]. See also Marcus & Pinsky [1969].

6.2 Exceedances. The corresponding univariate limit theory for *exceedances* was developed independently in Pickands [1975] and Balkema & de Haan [1974] in terms of *residual life times*. A df F *lies in the domain* of the df G on $[0,\infty)$ *for exceedances*, and we write $F \in \mathcal{D}^+(G)$, if there exist $a_t > 0$ such that for $v \geq 0$

$$\frac{1-F(t+a_t v)}{1-F(t)} \to 1-G(v), \quad t \to y_\infty = \sup\{F < 1\}. \tag{6.3}$$

Properly scaled, the limit laws are precisely the standard *generalized Pareto distributions* GPDs G_τ on $[0,\infty)$ defined in (5) in the Preview. Note that G_τ has finite endpoint for $\tau < 0$. Discrete limit laws are possible if one allows shifts $b_t \neq t$ in (6.3); see Balkema & de Haan [1974].

The domain of attraction $\mathcal{D}^+(\tau) = \mathcal{D}^+(G_\tau)$ of the limit law G_τ is precisely the domain of attraction of the corresponding max-stable distribution. For the *Pareto* and the *power laws* the domain is determined by *regular variation* of the tail. The domain $\mathcal{D}^+(0)$ of the exponential law has more variety.

6.3 The domain of the exponential law

Theorem 6.1. *The df F lies in $\mathcal{D}^+(0)$, the domain of Λ for maxima, if and only if*

$$1-F(y) \sim e^{-\psi(y)}, \quad y \to y_\infty = \sup\{F < 1\},$$

where ψ is a C^2 function on a left neighbourhood of y_∞ with the properties:

$$\psi'(y) > 0, \quad \psi(y) \to \infty, \quad (1/\psi')'(y) \to 0, \quad y \to y_\infty. \tag{6.4}$$

Proof. Set $a(t) = 1/\psi'(t)$. By the Mean Value Theorem for $s_n \to s$, $t_n \to y_\infty$

$$\frac{a(t_n + a(t_n)s_n)}{a(t_n)} \to 1, \quad \frac{\psi(t_n + a(t_n)s_n) - \psi(t_n)}{a(t_n)} \to s. \tag{6.5}$$

It follows that the tail function $T = 1 - F \sim e^{-\psi}$ satisfies

$$\frac{T(t+a_t v)}{T(t)} \sim e^{\psi(t)-\psi(t+a_t v)} \to e^{-v}, \quad t \to y_\infty,\ v \in \mathbb{R}. \tag{6.6}$$

To prove necessity, observe that $\varphi = -\log T$ by (6.3) satisfies

$$\varphi(t + a_t v) - \varphi(t) \to v, \quad t \to y_\infty \tag{6.7}$$

for $v \geq 0$. Choose t_n so that $\varphi(t_n) = n + o(1)$. This is possible by asymptotic continuity of T. Relation (6.7) implies that $\varphi(t) - \varphi_0(t) \to 0$ for $t \to y_\infty$, where φ_0 is the piecewise linear function with $\varphi_0(t_n) = n$. Hence $\varphi_0' = 1/c_n$ on (t_{n-1}, t_n), where $c_n = t_n - t_{n-1}$ and $c_{n+1} \sim c_n$ by (6.7). It remains to replace the piecewise constant function φ_0' by a C^1 function $\psi' = \varphi_0' + \theta$ such that the integral of θ over any interval $[t_{n-1}, t_n]$ vanishes, $\theta(t) = o(\varphi_0'(t))$ and $\theta'(t)/\varphi_0'(t)^2 \to 0$.

This may be done as follows: Choose a C^∞ function $\beta\colon [0,1] \to [-1,1]$ so that $\beta \equiv 0$ on a right neighbourhood of 0 and $\beta \equiv 1$ on a left neighbourhood of 1, and so that the integral of β over $[0,1]$ vanishes. Use affine copies of β to interpolate between successive points $(t_n, 1/c_n)$. In this way we replace φ_0' by a C^∞ function which we call ψ'. The function ψ agrees with the piecewise linear function φ_0 in the points t_n, $n \geq m_0$, if it agrees in t_{m_0}. The other conditions hold since $|\theta|$ is bounded by $\|\beta\|_\infty |1/c_{n+1} - 1/c_n|$ on $[t_n, t_{n+1}]$ and $c_{n+1}/c_n \to 1$ implies $|1/c_{n+1} - 1/c_n| = o(1/c_n)$, and

$$|\psi''(t)| = |\theta'(t)| \leq \|\beta'\|_\infty |1/c_{n+1} - 1/c_n|/c_n, \quad t \in [t_n, t_{n+1}].$$

Hence $\psi(t) - \varphi(t)$ and $(1/\psi')'(t) = -\psi''(t)/\psi'(t)^2$ vanish in y_∞. □

6.4 The Poisson point process associated with the limit law. For exceedances, as for maxima, one may distinguish three classes of limit laws. In a suitable normalization they have densities

$$\begin{aligned} s/v^{s+1}, &\quad v \in [1,\infty),\ s = -1/\tau > 0;\\ e^{-v}, &\quad v \in [0,\infty);\\ s|v|^{s-1}, &\quad v \in [-1,0),\ s = 1/\tau > 0. \end{aligned} \tag{6.8}$$

The extensions s/v^{s+1} on $(0,\infty)$, e^{-v} on $\mathbb{R}$, and $s|v|^{s-1}$ on $(-\infty,0)$ are densities of infinite measures, Radon measures ρ on unbounded open intervals. These measures are univariate *excess measures*. For any halfline $H = [y,\infty)$ for which $\rho(H)$ is finite the probability measure $d\rho^H = 1_H d\rho/\rho(H)$ is a GPD, which may be normalized to have a df G_τ as in (5) in the Preview.

There are not many one-parameter groups in $\mathcal{A}(1)$. One-parameter groups of affine transformations are either translations, $\tau^t(v) = v + qt$, or expansions (or contractions) from a fixed center c, $\gamma^t(v) = e^{\lambda t}(v-c)+c$, with $\lambda > 0$ for expansions (and $\lambda < 0$ for contractions). This proves

Theorem 6.2. *Apart from a change in coordinates, and a positive weight factor, the list above contains all univariate excess measures on an interval in* $\mathbb{R}$.

We shall focus on the excess measure $d\rho(v) = e^{-v}dv$. The translation $v \mapsto v+t$ maps ρ onto $e^t\rho$. The Poisson point process M on $\mathbb{R}$ with mean measure ρ has points

$$V_1 > V_2 > \cdots .$$

The rv V_1 has a max-stable distribution, the *Gumbel distribution*,

$$\mathbb{P}\{V_1 \le v\} = \mathbb{P}\{M(v,\infty) = 0\} = e^{-\rho(v,\infty)} = \exp(-e^{-v}).$$

Now consider the behaviour of sample clouds. Let $Y_1, Y_2, \ldots$ be a sequence of independent observations from the df $F \in \mathcal{D}^+(G_0)$. Choose $t_n \uparrow y_\infty$ such that $nT(t_n) \to 1$, where $T = 1 - F$, and let $a_n = 1/\psi'(t_n)$ with $e^{-\psi} \sim T$ as in Theorem 6.1. Then

$$nT(t_n + a_n v) \to e^{-v}, \quad v \in \mathbb{R} \tag{6.9}$$

by (6.6). Let $\alpha_n(v) = t_n + a_n v$. Then $v \mapsto nT(t_n + a_n v)$ is the tail function of the finite measure $d\rho_n = \alpha_n^{-1}(ndF)$. So ρ_n is the mean measure of the normalized sample cloud M_n consisting of the n points $(Y_i - t_n)/a_n$, $i = 1, \ldots, n$. The limit relation (6.9) implies that $d\rho_n(v) \to d\rho(v) = e^{-v}dv$ weakly on any halfline $[c,\infty)$. By the theory of Chapter I

$$M_n \Rightarrow M \ \text{ weakly on } [c,\infty), \quad c \in \mathbb{R}. \tag{6.10}$$

Let $Y_{n1} \ge Y_{n2} \ge \cdots \ge Y_{nn}$ denote the first n observations from the df F arranged in decreasing order. The points $V_{ni} = (Y_{ni} - t_n)/a_n$ form the normalized sample cloud M_n. The limit relation (6.10) implies:

Proposition 6.3. $(V_{n1}, \ldots, V_{nn}) \Rightarrow (V_1, V_2, \ldots)$ *holds in the sense of weak convergence of the finite-dimensional distributions: For each* $k \ge 1$

$$(V_{n1}, \ldots, V_{nk}) \Rightarrow (V_1, \ldots, V_k), \quad n \to \infty.$$

Proof. For any reals $a_1, \ldots, a_k$,

$$\{V_{n1} > a_1, \ldots, V_{nk} > a_k\} = \{M_n(a_1,\infty) \ge 1, \ldots, M_n(a_k,\infty) \ge k\}.$$

A similar relation holds for the variables $V_1, \ldots, V_k$ and the Poisson point process M. The probabilities on the right converge for $n \to \infty$, since ρ does not charge the boundaries $\partial B_i = \{a_i\}$ of the Borel sets $B_i = (a_i, \infty)$ in $\mathbb{R}$. Weak convergence on halflines $[c,\infty)$ prevents points from disappearing at ∞. □

Similar results hold for distributions in the domain of the GPD G_τ for $\tau \ne 0$. In each case the measure ρ and the df H of the maximal point are connected by $H(v) = e^{-R(v)}$ where $R(v) = \rho(v,\infty)$. We call ρ the *exponent measure* of the max-stable df H.

In the univariate setting the limit theory for exceedances (and some point process theory) allows us to handle the asymptotics of partial maxima, and not only the maxima, but the whole sequence of upper *order statistics*. See EKM pp. 244–247 for details.

6.5* Monotone transformations. So far we have used affine transformations to normalize the maxima. Affine transformations are appropriate for normalizing the partial sums $Y^{+n} = Y_1 + \cdots + Y_n$ since

$$\frac{Y^{+n} - nb_n}{a_n} = \frac{Y_1 - b_n}{a_n} + \cdots + \frac{Y_n - b_n}{a_n}.$$

For maxima any monotone transformation φ has the same property

$$\varphi^{-1}(Y^{\vee n}) = \varphi^{-1}(Y_1) \vee \cdots \vee \varphi^{-1}(Y_n),$$

and hence one may consider limits for maxima normalized by monotone transformations.

Definition. A *monotone transformation* is a strictly increasing continuous function from an interval containing an interior point onto another such interval. The set of all monotone transformations will be denoted by $\mathcal{M}^{\uparrow}$.

Continuous dfs G which are strictly increasing on the open interval $\{0 < G < 1\}$ all are of the same type under monotone transformations. Any limit theory for maxima needs asymptotically continuous tail functions $T = 1 - F$. These are characterized by the basic tail condition: there is an increasing sequence t_n such that

$$T(t_n) > 0, \quad T(t_n) \to 0, \quad T(t_{n+1})/T(t_n) \to 1. \tag{6.11}$$

See Section 3.1 in EKM for a good discussion of *asymptotic tail continuity*.

The limit theory for maxima (and exceedances) under monotone transformations is simple.

Theorem 6.4. *A df F lies in the domain $\mathcal{D}^{\uparrow}(E)$ of the standard exponential distribution E on* $(-\infty, 0)$ *for maxima with respect to monotone transformations if and only if the tail function $T = 1 - F$ satisfies the basic tail condition* (6.11).

Proof. We may assume that $T(t_n)$ is strictly decreasing. Replace T by the piecewise linear function T_0 which agrees with T in the points t_n for $n \geq 1$ and equals one in t_0. There is a monotone transformation φ which maps $[-1, 0)$ onto $[t_0, y_\infty)$, such that $T_0 \circ \varphi(v) = -v$ for $v \in [-1, 0)$. (Take $\varphi^{-1} = -T_0$.) Then $T(\varphi(v)) \sim -v$ for $v \uparrow 0$. Set $\varphi_n(v) = \varphi(v/n)$. Then

$$-\log F^n(\varphi_n(v)) \sim nT(\varphi(v/n)) \to -v, \quad v < 0$$

shows that $F^n \circ \varphi_n \to E$ on $(-\infty, 0)$. □

6.6* The von Mises condition. This subsection and the next treat the domain $\mathcal{D}^+(0)$. In this subsection we look at densities. We shall establish a representation

$$d\pi = f d\mu$$

for distributions $\pi \in \mathcal{D}^+(0)$, where f is a density which satisfies the von Mises conditions, and μ is a roughening of Lebesgue measure, a measure which behaves asymptotically like Lebesgue measure. In this representation the function f describes the global behaviour of the upper tail, and the measure μ describes the local variations. We shall also elucidate in how far the scaling function $a(t)$ in (6.3) determines the tail behaviour of a df F in $\mathcal{D}^+(0)$.

The *scale function* $a(t)$ in the basic limit relation (6.3) is the key to describing the domains of attraction. In the multivariate setting it determines a metric, as we shall see later. Suppose $Y \in \mathcal{D}^+(\tau)$. For $\tau < 0$ the upper endpoint y_∞ is finite and one may choose $a(y) = y_\infty - y$; for $\tau > 0$ (heavy tails) one may take $a(y) = y$. For $\tau = 0$ the situation is more delicate. The scale function is self-neglecting. Self-neglecting functions form the subject of the next subsection.

Definition. A df F satisfies the *von Mises condition* if $1 - F(y) = e^{-\psi}$ where ψ satisfies (6.4) on a left neighbourhood of the upper endpoint $y_\infty = \sup\{F < 1\}$. This may be expressed as

$$(1 - F(y))F''(y)/(F'(y))^2 \to -1, \quad y \to y_\infty.$$

In general we say that $f = e^{-\psi}$ or ψ is a *von Mises function* and satisfies (6.4).

Any df in $\mathcal{D}^+(0)$ is tail asymptotic to a df F with tail $1-F = e^{-\psi}$ which satisfies the von Mises condition by Theorem 6.1. We claim that one may choose F to have a density which satisfies the von Mises condition.

The role which distribution functions play in the theory of maxima and of exceedances is exemplified in the relations (6.2) and (6.3). This should not blind us to the fact that in practice it is densities, not dfs, which are the basic concern. Sample clouds suggest densities rather than dfs.

Let us first show that $Y \in \mathcal{D}^+(0)$ if its density satisfies the von Mises condition.

The natural scale function for exceedances is $a(y) = 1/\psi'(y)$. By (6.4) it satisfies

$$a(y) = o(y) \text{ if } y_\infty = \infty, \quad a(y) = o(y_\infty - y) \text{ if } y_\infty < \infty. \tag{6.12}$$

Hence $a(y + va(y))$ is well defined eventually for any $v \in \mathbb{R}$, and $y_n + v_n a(y_n) \to y_\infty$ if $y_n \to y_\infty$ and $v_n \to v \in \mathbb{R}$. The conditions on ψ imply

$$a(y_n + v_n a(y_n))/a(y_n) \to 1, \quad y_n \to y_\infty, v_n \to v, \ v \in \mathbb{R}, \tag{6.13}$$

since $(a(y_n + v_n a(y_n)) - a(y_n))/a(y_n) = a'(y_n + \omega_n a(y_n))v_n \to 0$ by the Mean Value Theorem. This implies

$$\psi(y_n + v_n a(y_n)) - \psi(y_n) = \psi'(y_n + \omega_n a(y_n))v_n a(y_n) \to v$$

$y_n \to y_\infty$, $v_n \to v \in \mathbb{R}$. The reader may check that the two conditions on ψ ensure that $e^{-\psi}$ is integrable over $[0, y_\infty)$. The last relation gives

$$f(y_n + v_n a(y_n))/f(y_n) \to e^{-v}.$$

This implies $Y \in \mathcal{D}^+(0)$ and

$$(1 - F(y + va(y)))/(1 - F(y)) \to e^{-v}, \quad y \to y_\infty, \ v \in \mathbb{R}. \tag{6.14}$$

We leave the details to the reader. In Chapter III we give a multivariate version of this result.

Proposition 6.5. *Let Y have density f. If f satisfies the von Mises condition in its upper endpoint then $Y \in \mathcal{D}^+(0)$.*

We now first turn to a question about the normalization. By the Convergence of Types Theorem the scale function $a(y)$ is unique up to asymptotic equality. Now suppose the dfs F_1 and F_2 have the same upper endpoint and the same scale function. Does this imply that the tails are asymptotic?

The relation (6.14) holds for both F_1 and F_2. It follows that the quotient $L(y) = (1 - F_2(y))/(1 - F_1(y))$ satisfies the relation

$$L(y_n + u_n a(y_n))/L(y_n) \to 1, \quad y_n \to y_\infty, \ u_n \to u, \ u \in \mathbb{R}. \tag{6.15}$$

Definition. Functions L defined and positive on a left neighbourhood of y_∞ are called *flat* function for a if they satisfy (6.15).

If F lies in $\mathcal{D}^+(0)$ with upper endpoint y_∞ and scaling $a(y)$, then so does G if $1 - G = (1 - F)L$ where L is flat for a. Flat functions describe the variation that is possible in $\mathcal{D}^+(0)$ if the upper endpoint y_∞ and scale function a are given. The product of flat functions is flat. Functions asymptotic to flat functions are flat. The scale function a itself is flat by (6.13), and so is its inverse, $\psi' = 1/a$.

In order to construct, for a given df $F \in \mathcal{D}^+(0)$, a df tail asymptotic to F, with a density which satisfies the von Mises condition, we need flat functions which are smooth.

Lemma 6.6. *Suppose $L = e^\lambda$ is flat for a. There exists a function $L_0 = e^{\lambda_0} \sim L$ which is C^∞ and which satisfies*

$$a^n(y)\lambda_0^{(n)}(y) \to 0, \quad y \to y_\infty. \tag{6.16}$$

Proof. Choose $y_n \uparrow y_\infty$ such that $y_{n+1} = y_n + a(y_n)$. Then $\lambda(y_{n+1}) - \lambda(y_n) \to 0$, and one may construct a C^∞ function λ_0 which agrees with λ in the points y_n such that $\lambda_0' = o(1/a(y))$ in y_∞, and such that (6.16) holds. Construct λ_0 by interpolating with affine copies of an increasing C^∞ function χ mapping [0, 1] onto itself, where we choose χ so that the derivatives of all orders vanish in the two endpoints. Compare the construction in the proof of Theorem 6.1. Set $a_n = a(y_n)$ and $c_n = \lambda(y_{n+1}) - \lambda(y_n)$. Then $c_n \to 0$ and

$$\lambda_0(y) = \lambda_0(y_n + ua_n) = \lambda(y_n) + c_n\chi(u), \quad a_n^m \lambda_0^{(m)}(y) = c_n\chi^{(m)}(u). \quad (6.17)$$

The function $L_0 = e^{\lambda_0}$ is flat, and asymptotic to L in y_∞. It is C^∞ and satisfies (6.16). □

Proposition 6.7. *Any df F in $\mathcal{D}^+(0)$ is tail asymptotic to a df with density $e^{-\psi}$ where ψ satisfies (6.4). Moreover $a = 1/\psi'$ may be used as scale function in the limit relation (6.14).*

Proof. Suppose $F \in \mathcal{D}^+(0)$. Then F is tail asymptotic to a df $F_0 = 1 - e^{-\psi}$ which satisfies the von Mises condition. The df F_0 has a continuously differentiable density $f_0(y) = \psi'(y)e^{-\psi(y)}$. The function ψ' is flat and may be replaced by a flat function $L = e^\lambda \sim \psi'$ where λ is C^2, and $\lambda' a$ and $\lambda'' a^2$ vanish in y_∞ by (6.17). Hence F_0 is tail asymptotic to a df F_1 with density $e^{-\psi_1}$ where $\psi_1' = 1/a - \lambda' \sim 1/a$, and $\psi_1''/(j_1')^2 \sim (\psi'' - \lambda'')/a^2 \to 0$. □

We shall now show that any df in $\mathcal{D}^+(0)$ has a representation $dF = f d\mu$ in terms of such a density $e^{-\psi}$. The advantage of this representation over Proposition 6.7 will become apparent in Chapters III and IV when we work in a multivariate setting.

Definition. A Radon measure μ on $(-\infty, y_\infty)$ is a *roughening of Lebesgue measure for a scale function a* if

$$s_{n+1} - s_n = o(a_n) \quad \text{and} \quad \mu[s_n, s_{n+1})/(s_{n+1} - s_n) \to 1$$

for some strictly increasing sequence $s_n \to y_\infty$.

Theorem 6.8. *A df F with upper endpoint y_∞ lies in $\mathcal{D}^+(0)$ with scale function a if and only if $dF = e^{-\psi} d\mu$ on a left neighbourhood of y_∞. Here μ is a roughening of Lebesgue measure with respect to the scale function a, and ψ is a C^2 function which satisfies (6.4). Moreover $\psi'(y)a(y) \to 1$ for $y \to y_\infty$.*

Proof. Define $d\mu = e^\psi dF$ on a left neighbourhood of the upper endpoint y_∞ of F. Then

$$\mu[t, t + va(t))/a(t) \to v, \quad t \to t_\infty, \ v > 0.$$

Indeed, set $V_t = (Y^t - t)/a(t)$. Then

$$\begin{aligned}\mu[t, t + va(t)) &= \mathbb{E}e^{\psi(Y)}[t \le Y < t + va(t)] \\ &= \mathbb{E}e^{\psi(Y^t)}[Y_t - t < va(t)]/\mathbb{P}\{Y \ge t\} \\ &= \mathbb{E}e^{\psi(t+a(t)V_t)}[V_t \le v]/(1 - F(t)).\end{aligned}$$

Observe that $1-F(t) \sim a(t)e^{-\psi(t)}$ by definition of ψ, and that $\psi(t+va(t))-\psi(t) = \varphi_t(v) \to v$ uniformly on bounded intervals. Hence

$$\mathbb{E}e^{\varphi_t(V_t)}[V_t \le v] = \int_0^v e^{\varphi_t(s)} dG_t(s) \to \int_0^v e^s(e^{-s}ds) = v,$$

since the df G_t converges weakly to the standard exponential distribution $dG(s) = e^{-s}ds$. □

6.7* Self-neglecting functions. The scale function satisfies $a(t + sa(t)) \sim a(t)$ for $t \to y_\infty$. This relation plays an important role in asymptotics.

Definition. A function f is *self-neglecting* in $t_\infty \le \infty$ or *Beurling slowly varying* if it is defined and positive on a left neighbourhood $[t_0, t_\infty)$ of t_∞, and satisfies the limit relation

$$f(t + xf(t))/f(t) \to 1, \quad t \to t_\infty,\ x \in \mathbb{R}. \tag{6.18}$$

Implicit in the definition is the assumption that $f(t) = o(t)$ if $t_\infty = \infty$ and $f(t) = o(t_\infty - t)$ if t_∞ is finite. This ensures that $f(t + xf(t))$ is defined eventually for any $x \in \mathbb{R}$. Limits are from below: $t_n \to t_\infty$ implies $t_n < t_\infty$ since f is only defined on a left neighbourhood of t_∞.

A positive C^1 function whose derivative vanishes in $t_\infty = \infty$ is self-neglecting. So too for t_∞ finite if the function itself also vanishes in t_∞. Bloom's theorem, see Bloom [1976] or Geluk & de Haan [1987], states that a continuous self-neglecting function is asymptotic to a positive C^1 function whose derivative vanishes in t_∞. It is not known whether one may replace continuity by measurability. If one replaces (6.18) by the condition (6.13),

$$f(t_n + x_n f(t_n))/f(t_n) \to 1, \quad t_n \to t_\infty,\ x_n \to x,\ x \in \mathbb{R}, \tag{6.19}$$

then f is asymptotic to a C^1 function whose derivative vanishes in t_∞. See below.

Basic to our analysis of the self-neglecting function f on $[t_0, t_\infty)$ is the concept of a *master sequence* (t_n) and the associated *model function* φ. A master sequence is a sequence $t_1, t_2, \ldots$ in $[t_0, t_\infty)$, determined by t_1 and the recursion $t_{n+1} = t_n + f(t_n)$, such that $t_n \to t_\infty$. (The latter condition is satisfied automatically if f is bounded below on $[t_0, t_\infty)$ by a positive continuous function.) The model function φ is the continuous function on $[t_1, t_\infty)$ which agrees with f in the points t_n and is linear between successive points. Since $f(t + f(t)) \sim f(t)$ for $t \to t_\infty$, the slope of a model function tends to zero. For any $\varepsilon > 0$ there exists $t_\varepsilon \in [t_0, t_\infty)$ such that

$$|\varphi(s_2) - \varphi(s_1)| < \varepsilon(s_2 - s_1), \quad t_\varepsilon \le s_1 < s_2 < t_\infty \tag{6.20}$$

for all model functions φ which are defined on $[s_1, s_2]$. In order to prove that f is asymptotic to a C^1 function whose derivative vanishes in t_∞, it suffices to show that f is asymptotic to a model function.

Proposition 6.9. *A function f which satisfies (6.19) is asymptotic to a model function.*

Proof. There is no sequence $s_n \to t_\infty$ such that $(s_n, 0)$ lies in the closure of the graph of f for all n. Otherwise one could choose $x_n \to 0$ such that $f(s_n) > 2f(s_n + x_n f(s_n))$, contradicting (6.19). So model functions exist. Let φ be a model function on $[t_1, t_\infty)$. If $f(t)/\varphi(t)$ does not tend to 1 for $t \to t_\infty$, there is a sequence $s_n = t_{k_n} + x_n f(t_{k_n})$ with $x_n \in [0, 1]$ such that $f(s_n)/\varphi(s_n) \to A \in [0, \infty] \setminus \{1\}$. Take a subsequence $x_{m_n} \to x \in [0, 1]$ to obtain a contradiction with (6.19). □

We now formulate and prove the main result.

Theorem 6.10. *Let $t_0 < t_\infty \le \infty$. Let f be a positive measurable function on $[t_0, t_\infty)$. Assume that $1/f$ is bounded on each interval $[t_0, t_1]$ß$[t_0, t_\infty)$, and that there exists a real $q > 0$ such that*

1) *$t + qf(t) < t_\infty$ eventually;*
2) *$f(t + xf(t))/f(t) \to 1$ for $t \to t_\infty$, $x \in (0, q]$.*

Then f is asymptotic in t_∞ to a C^1 function f_0 whose derivative vanishes in t_∞.

Proof. We may assume that $q = 1$ (replace f by qf) and that $t + f(t) < t_\infty$ for all $t \in [t_0, t_\infty)$ (increase t_0). We shall show that f is asymptotic to the model function φ_0 with master sequence starting in t_0.

First observe that $f(s_n + xf(s_n))/f(s_n) \to 1$ for each $x \in [0, 1]$ if $s_n \to t_\infty$. It follows that

$$f(t + Xf(t))/f(t) \to 1 \text{ in probability for } t \to t_\infty,$$

where X is uniformly distributed on $[0, 1]$. For $\varepsilon > 0$ there exists t_ε such that by (6.20)

$$\mathbb{P}\{e^{-\varepsilon} < (f/\varphi)(t_n + Xf(t_n)) < e^{\varepsilon}\} > 0.9, \quad t_n \ge t_\varepsilon$$

for any model function φ and associated master sequence (t_n). If the interval $[s_1, s_2]$ of length $s_2 - s_1 = r > 0$ in $[t_\varepsilon, t_\infty)$ contains two successive points of the sequence (t_n), then by averaging,

$$\mathbb{P}\{e^{-\varepsilon} < (f/\varphi)(s_1 + rX) < e^{\varepsilon}\} > 9/14,$$

where we assume t_ε so large that $1/2 < f(t + f(t))/f(t) < 2$ for $t > t_\varepsilon$.

With each model function φ we associate the band between $e^{-\varepsilon}\varphi$ and $e^{\varepsilon}\varphi$. The last inequality shows that the bands for any two model functions φ and $\tilde{\varphi}$ intersect somewhere above any interval $[s_1, s_2]$ which contains two successive points of both master sequences (t_n) and $(\tilde{t}_n)$. Let φ_0 be the model function with starting point $(t_0, f(t_0))$. Let φ be the model function with starting point $(t, f(t))$, and assume $f(t) = A\varphi_0(t)$. Let $t' = t + 3(f(t) \vee \varphi_0(t))$. For t close to t_∞, the interval $[t, t']$ will contain two successive points of the master sequences of the model functions φ_0 and φ. Since these functions are asymptotically horizontal, by (6.20) the bands will be disjoint over the interval $[t, t']$ unless $A - 1 = O(\varepsilon)$. This proves that f is asymptotic to φ_0. □

Example 6.11. The condition that $1/f$ is locally bounded cannot be dropped. Define $f > 0$ on $[0, \infty)$ by $f(\sqrt{n} - u) = u^2$ for $0 < u \leq \sqrt{n} - \sqrt{n-1}$, $n = 1, 2, \ldots$. ◊

7 Componentwise maxima

The multivariate limit theory for max-stable distributions is strong, versatile, elegant and rich. One reason for its strength is its ability to incorporate the univariate asymptotics for the d marginals into the multivariate theory. This is done by introducing a dependency. Copulas form the natural tool to describe the dependency structure. Max-stable copulas are copulas which satisfy a simple functional equation. These copulas together with d univariate extreme value distributions uniquely determine the multivariate max-stable distributions. Moreover there is a characterization of the domain of attraction of the max-stable distributions in terms of the univariate marginals and a simple condition on the asymptotic behaviour of the copula in its upper endpoint. The sample copula gives a discrete approximation to the exponent measure of max-stable limit distributions with exponential marginals on $(-\infty, 0]$.

Economic and financial situations are governed by complex stochastic dynamical processes. In order to analyse the *risk* of the situation one introduces a number of variates. These may quantify exchange rates, labour unrest, fire hazards, commodity prices, etc. Suppose we have d such variables, $Z_1, \ldots, Z_d$. For each, large values are associated with high risk. The variates may express losses in euro, the logarithm of such losses, or any other monotone transformations of the losses. For simplicity assume we have a sample of n independent observations of the vector Z in $\mathbb{R}^d$. In first instance we are concerned with the asymptotic behaviour of the univariate maxima, $Z_i^{\vee n}$, for $n \to \infty$. We assume that for each $i = 1, \ldots, d$, the univariate distribution F_i of the ith component Z_i lies in the domain $\mathcal{D}^+(\tau_i)$ of an extreme value distribution. In order to get a better grip on the risk described by the vector $Z = (Z_1, \ldots, Z_d)$ one has to study the dependency between the various components. That is the subject of this section. The concept of max-stable dfs and the associated limit theory of *coordinatewise extremes* were developed independently in de Haan & Resnick [1977] and Galambos [1978]. See Joe [1997] or McNeil, Frey & Embrechts [2005] for a good overview of the theory of copulas and its relevance to multivariate extremes. Two recent theses in this area are Alink [2007] and Demarta [2007]. The relation of componentwise maxima to point processes is treated in Resnick [2004]. There exists a rich literature on the statistical analysis of componentwise extremes; see Huang [1992], de Haan & Ferreira [2006], Drees [2003] or Hall & Tajvidi [2000].

7.1 Max-id vectors. All operations are componentwise. Vectors in $\mathbb{R}^d$ are regarded as real valued functions on the finite set $\{1,\dots,d\}$. The *lower endpoint* a, and *upper endpoint* b, of a df F on $\mathbb{R}^d$ are defined in terms of the marginals

$$a_i = \inf\{F_i > 0\} \geq -\infty, \quad b_i = \sup\{F_i < 1\} \leq \infty, \quad i = 1,\dots,d.$$

As in the univariate case, the df of the maximum of two independent vectors X and Y in $\mathbb{R}^d$ is the product of the dfs of X and Y. Relation (6.2) also holds when we read X, Y and t as vectors.

A df F is *max-id* if F^t is a df for all $t > 0$.

Example 7.1. In the multivariate case not every df is max-id. Suppose

$$F(0,0) = 0, \quad F(1,0) > 0, \quad F(0,1) > 0.$$

Then F is not max-id. Indeed, let π have df F^ε for $\varepsilon > 0$ so small that $F^\varepsilon(1,0) > 1/2$ and $F^\varepsilon(0,1) > 1/2$. Then

$$\pi((0,1] \times (-\infty,0]) + \pi((-\infty,0] \times (0,1]) > 1.$$

In $\mathbb{R}^d$ with the disjoint sets $E_i = \{0 < z_i \leq 1\} \cap \bigcap_{j \neq i}\{z_j \leq 0\}$ the same argument applies. ◊

Now assume F is max-id with lower endpoint $a = 0$. Then F is positive on $(0,\infty)^d$ by the argument above. Define $R = -\log F$ on $(0,\infty)^d$. Observe that F^ε is a df, and so is $F^\varepsilon - 1$, but now normalized to vanish in $(\infty,\dots,\infty)$. The function $(F^\varepsilon - 1)/\varepsilon$ is the df of a measure on $[0,\infty)^d$ of total mass $1/\varepsilon$. The weak limit for $\varepsilon \downarrow 0$ is $\log F = -R$. Thus R is the df of a measure ρ on $[0,\infty)^d \setminus \{0\}$ defined by

$$\rho([0,\infty)^d \setminus [0,z]) = R(z), \quad z \in (0,\infty)^d, \ R = -\log F. \tag{7.1}$$

Definition. The measure ρ in (7.1) is the *exponent measure* of the max-id df F.

Note that ρ is a Radon measure on $[0,\infty]^d \setminus \{0\}$ which lives on $[0,\infty)^d \setminus \{0\}$ and is finite on $[0,\infty)^d \setminus [0,z]$ for any $z \in (0,\infty)^d$. Introduce the Poisson point process M with mean measure ρ, and with points $Z_1, Z_2, \dots$. (If $F(0) > 0$ then ρ is finite, and M will have only finitely many points.) Let Z be the maximum of the points of M:

$$Z := \sup M = 0 \vee Z_1 \vee Z_2 \vee \cdots .$$

For $z \in (0,\infty)^d$, by (7.1)

$$\mathbb{P}\{Z \le z\} = \mathbb{P}\{M([0,\infty)^d \setminus [0,z]) = 0\} = e^{-R(z)}.$$

So Z has df e^{-R}. We thus have a simple correspondence between measures ρ on $[0,\infty)^d \setminus \{0\}$ which are finite on halfspaces $\{z_i > c\}$, $c > 0$, $i = 1,\dots,d$, and max-id dfs with lower endpoint in the origin.

One may define the *max-Lévy process* $W : [0,\infty) \to [0,\infty)^d$ by

$$W(t) = 0 \vee \sup\{W_k \mid T_k \le t\},$$

where $(W_k, T_k) \in [0,\infty)^d \times [0,\infty)$ are points of the Poisson point process with mean measure $d\rho \times dt$. By construction $W(t+s)$ is distributed like the maximum of $W(t)$ and a copy of $W(s)$ independent of $W(t)$. (The Poisson point process on the time slice $(t, t+s]$ is independent of $W(t)$, and distributed like the Poisson point process on the time slice $(0,s]$.) Since $W(1)$ has df F it follows that $W(t)$ has df F^t for all $t > 0$.

The same results hold for any max-id df F. If the lower endpoint is $(-\infty,\dots,-\infty)$, the exponent measure ρ is a Radon measure on $[-\infty,\infty]^d \setminus \{(-\infty,\dots,-\infty)\}$. It may charge some of the coordinate hyperplanes in $-\infty$. Monotone transformations of the coordinates do not affect max-infinite divisibility since such transformations commute with maxima. So in principle there is no difference between the space $[0,\infty]^d$ and $[-\infty,\infty]^d$.

Example 7.2. Suppose $Z = (X,Y)$ has density $f(x,y) = e^{x+y}$ on $(-\infty,0)^2$. Then $F(x,y) = e^{x+y}$ and $R(x,y) = x+y$. The function R is C^2, so we obtain the density $r(x,y)$ of the exponent measure by differentiation: $r \equiv 0$ on $(-\infty,0)^2$. Since F is continuous, the exponent measure ρ is infinite. A little thought will show that ρ is standard one-dimensional *Lebesgue measure* on the two halflines $\{-\infty\} \times (-\infty,0)$ and $(-\infty,0) \times \{-\infty\}$. $\diamondsuit$

7.2 Max-stable vectors, the stability relations. We normalize multivariate maxima by *coordinatewise affine transformations* CATs. These affine transformations α of $\mathbb{R}^d$ have the form

$$\alpha(z) = (a_1 z_1 + b_1, \dots, a_d z_d + b_d), \quad a_i > 0,\ i = 1,\dots,d.$$

For max-stable vectors we need to define powers of CATs. The product (composition) of CATs is a CAT, and so is the inverse. Positive affine transformation $\alpha : y \mapsto ay + b$ on $\mathbb{R}$ define a group of transformations α^t, as in Section 6.4, such that

$$\alpha^{t+s} = \alpha^t \circ \alpha^s, \quad \alpha^1 = \alpha.$$

Since CATs act coordinatewise, one has

$$\alpha^t(z) = (\alpha_1^t(z_1), \dots, \alpha_d^t(z_d)), \quad t \in \mathbb{R},\ z = (z_1, \dots, z_d) \in \mathbb{R}^d.$$

Definition. A random vector Z with df F lies in the *domain of attraction* of the *max-stable* vector W with df G if the components of W are non-degenerate, and if there exist CATs α_n such that one of the equivalent relations holds:

$$F^n \circ \alpha_n \to G, \quad \alpha_n^{-1}(Z^{\vee n}) \Rightarrow W.$$

In this case we write $F \in \mathcal{D}^\vee(G)$ or $Z \in \mathcal{D}^\vee(W)$.

Max-stable dfs are max-id. The max-stable dfs are precisely those max-id dfs which satisfy a *stability relation*: All powers F^n are of the same type with regard to CATs. In terms of the max-Lévy process $W(t), t > 0$, one may describe the stability relation explicitly:

$$W(t) \stackrel{\text{d}}{=} \gamma^{\log t}(W(1)),$$

where γ is a CAT. See EKM or Resnick [2004] for details. Comparing the exponent measures of the max-id vectors on either side of the equation we find: the exponent measure ρ is max-stable if and only if there exists a CAT γ such that

$$\gamma^t(\rho) = e^t \rho, \quad t \in \mathbb{R}. \tag{7.2}$$

So ρ is an excess measure if it lives on $\mathbb{R}^d$.

Since all operations are coordinatewise, it follows that if a vector Z lies in the domain of the max-stable vector with normalizations α_n^{-1}, then the components satisfy: $\alpha_{ni}^{-1}(Z_i^{\vee n}) \Rightarrow W_i$ for $i = 1, \dots, d$. Thus the marginals determine the normalizations. Similarly the stability relations for the univariate extreme value distributions determine the CAT γ in (7.2).

The next result is crucial. It links convergence of coordinatewise maxima to convergence of normalized sample clouds.

Theorem 7.3 (de Haan–Resnick). *Let Z be a random vector in $[0, \infty)^d$ with distribution π and df F, and let α_n be CATs. Let H be a max-id df on $[0, \infty)^d$ with exponent measure ρ on $[0, \infty)^d \setminus \{0\}$. The following are equivalent:*

$$F^n \circ \alpha_n \to H \quad \textit{weakly},$$

$$n\alpha_n^{-1}(\pi) \to \rho \quad \textit{vaguely on } [0, \infty]^d \setminus \{(0, \dots, 0)\}.$$

Proof. The logarithm has derivative $1/x$ on $(0, \infty)$. In particular $(\log x)/(1-x) \to 1$ for $x \uparrow 1$. So the first limit relation is equivalent to

$$1 - F \circ \alpha_n \to R = \log H \quad \text{weakly on } (0, \infty)^d, \tag{7.3}$$

which is equivalent to the second relation. □

If one assumes convergence of the normalized coordinate maxima, the sample cloud may be used to estimate the exponent measure. In fact from our point of view one may forget about convergence of the maxima. The first limit relation is convenient if one likes to work with multivariate dfs, but what is of real interest is convergence of the sample cloud, and that is expressed in terms of the measures $n\alpha_n^{-1}(\pi)$ in the second limit relation. The symmetry of the exponent measure allows us to make estimates on the far tails of the df F using the recipe in the Preview.

The condition $Z \geq 0$ in the theorem above may be relaxed, replacing Z by Z^+ with components $Z_i \vee 0$. The analog of Theorem 7.3 holds for max-stable distributions with lower endpoint $a \in [-\infty, \infty)^d$.

The sequence of partial maxima $Z^{\vee n} = Z_1 \vee \cdots \vee Z_n$ of independent observations from a df F on $\mathbb{R}^d$ is monotone. This does *not* imply convergence of the normalized maxima. If the Z_n themselves are daily or yearly maxima from a stationary continuous-time process where the time interval is long enough to ensure independence of successive terms, then it is not unreasonable to assume that the vectors Z_n are max-id. (One may use the Superposition Theorem for point processes.) So let the vectors Z_n have exponent measure $\hat{\rho}$. The only condition on $\hat{\rho}$ is that it is finite on halfspaces $\{z_i \geq c_i\}$, $i = 1, \ldots, d$, for which $F_i(c_i) > 0$, where F_i is the ith marginal of the df F. (Indeed $\hat{\rho}\{z_i > c_i\} = -\log F_i(c_i)$.) Convergence of the properly normalized partial maxima, $F^n \circ \alpha_n \to H$, is equivalent to convergence $n\alpha_n^{-1}(\hat{\rho}) \to \rho$, where ρ is the exponent measure of H. The measure ρ will satisfy a one-parameter group of symmetries $\gamma^t(\rho) = e^t \rho$, $t \in \mathbb{R}$. These symmetries of ρ have to be present in $\hat{\rho}$ asymptotically. Only then will there be a limit for the partial maxima. The marginal conditions $F_i \in \mathcal{D}^+(\tau_i)$ are necessary for convergence but not sufficient. For convergence of coordinatewise maxima one needs our Ansatz that distributions above certain thresholds asymptotically have the same shape.

The df $\hat{R}$ of the exponent measure $\hat{\rho}$ is asymptotic to $1 - F$ in the upper endpoint of the distribution. It makes no sense to argue whether $\hat{R}$ or $1-F$ gives more accurate information on the exponent measure ρ of the max-stable limit distribution (assuming that it exists).

Example 7.4. The *uniform distribution* on the square $(0, 1)^2$, the uniform distribution on the union of the squares $(0, 1/2)\times(1/2, 1)$ and $(1/2, 1)\times(0, 1/2)$, and the uniform distribution on the line segment with endpoints $(0, 1)$ and $(1, 0)$ lie in the domain of the max-stable distribution H with independent standard exponential marginals on $(-\infty, 0)$. In terms of maxima the uniform distribution on the square $(0, 1)^2$ is the natural distribution in $\mathcal{D}^\vee(H)$; in terms of exceedances it is more natural to pick a distribution which has no mass on a neighbourhood of $(1, 1)$. ◊

7.3 Max-stable vectors, dependence. If the random variable T has a *Gumbel distribution* Λ, then for $c > 0$ the random variable e^{cT} has a heavy tailed extreme value distribution on $(0, \infty)$, and $-e^{-cT}$ has an extreme value distribution on $(-\infty, 0)$. If we also apply a positive affine transformation to change the location and scale, we obtain the whole class of univariate extreme value limit variables indexed by (a, b) and c:

$$V = \alpha^T(c), \quad \alpha(c) = ac + b > c, \; a > 0.$$

The group α^t, $t \in \mathbb{R}$, is defined in Section 6.4. A similar representation exists in the multivariate setting:

Theorem 7.5 (Representation Theorem). *Any max-stable vector W has the representation*

$$W = (\gamma_1^{T_1}(c_1), \ldots, \gamma_d^{T_d}(c_d)), \quad \gamma_i(c_i) = a_i c_i + b_i > c_i, \; a_i > 0,$$

where $T = (T_1, \ldots, T_d)$ is a max-stable vector with univariate Gumbel marginals $\mathbb{P}\{T_i \le t\} = \exp(-e^{-t})$.

Proof. Monotone transformations of the form $\varphi(z) = (\varphi_1(z_1), \ldots, \varphi_d(z_d))$ with $\varphi_i \in \mathcal{M}^\uparrow$ map max-id vectors into max-id vectors. So T is max-id if and only if W is max-id (since the functions $t \mapsto \gamma_i^t(y_i)$ are in $\mathcal{M}^\uparrow$). The stability relation, $T(t)$ is distributed like $T + (\log t, \ldots, \log t)$, gives the stability relation, $W(t)$ is distributed like $\gamma^{\log t}(W)$, and vice versa. □

Remark 7.6. We see that the distribution of any max-stable vector W is determined by

1) the d marginal distributions;

2) the dependency structure.

These may be chosen independently. The marginals are determined by (c_i, γ_i), or by three real parameters, which may be associated with the shape, scale and location. The dependency structure of the multivariate distribution is determined by a copula. (In fact some authors prefer to use the term dependency rather than copula.) Since the components of T and W are linked by monotone transformations, the vectors W and T in the representation above have the same copula.

Definition. A *copula* is a multivariate df C with uniform marginals on $(0, 1)$.

Copulas are an efficient way for describing the dependency structure of random vectors by reducing to the case where the marginals are uniformly distributed. In first instance one should think of the case where the marginal distributions F_i are continuous on $\mathbb{R}$ and strictly increasing on the interval $\{0 < F_i < 1\}$. Then $F_i \in \mathcal{M}^\uparrow$, and so is the inverse function F_i^{-1}. In general, for arbitrary dfs on $\mathbb{R}$, one may define the *generalized inverse* $F^\leftarrow\colon (0, 1) \to \mathbb{R}$ by reflecting the graph of F, augmented with vertical line segments in the jumps to form a connected curve, in the diagonal by interchanging the axes. For an effective demonstration, use an overhead projector, and flip over the sheet. We shall adhere to common practice and choose $F^{\leftarrow}$ to be left-continuous.

In the univariate case any right-continuous increasing function F with limits zero and one in $-\infty$ and $+\infty$ is a df, since it is the df of the random variable $F^{\leftarrow}(U)$ where U is uniformly distributed on $(0,1)$. The multivariate analog describes a random vector as the image of a random vector with uniform marginals under the map

$$u \mapsto (F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)), \quad u = (u_1, \dots, u_d) \in (0,1)^d.$$

Definition. Let F be a multivariate df and C a copula. Then C is a *copula for F* if F is the df of the vector $(F_1^{\leftarrow}(U_1), \dots, F_d^{\leftarrow}(U_d))$, where $U = (U_1, \dots, U_d)$ has df C.

Theorem 7.7. *Every df F on $\mathbb{R}^d$ has a copula. The copula is determined by F in all points of the set $K = K_1 \times \dots \times K_d$, where K_i is the closure of the set $F_i(\mathbb{R})$ in $[0,1]$.*

Proof. See Joe [1997] or McNeil, Frey & Embrechts [2005]. □

In a statistical analysis one may use the univariate samples to estimate the tail behaviour of the d components. To determine the dependency, one then uses the relative ranks of the sample points. Assume $d = 2$. Given a sample of n points (x_i, y_i), draw n horizontal lines at levels y_i and n vertical lines through the x_i. If the marginals of the df are continuous, there are no ties and the n sample points yield n points in the lattice $\{1, \dots, n\}^2$, which retain the order of the original sample. Each row and each column contains exactly one point. This n-point subset of $\{1, \dots, n\}^2$ will be called a *sample copula*. Sample copulas also exist in dimension $d > 2$. If the df is not continuous, toss a coin to eliminate ties. See Deheuvels [1981], where the term *dependogram* is used for a similar concept, and Einmahl, de Haan & Li [2006].

Example 7.8. Let $Z_1, \dots, Z_n$ be a sample of n points from the max-stable df $G(x,y) = e^{x+y}$ on $(-\infty, 0)^2$. In the sample copula the points will be roughly uniformly distributed over $\{1, \dots, n\}^2$. Consider a square in the upper right hand corner with sides of length $k_n \sim c\sqrt{n}$. The number of points in the square is of the order of c^2, the number of points on the margins is $k_n \sim c\sqrt{n}$. For $n \to \infty$ the density of points in the subsquare vanishes, and we conclude that the exponent measure vanishes on $(-\infty, 0)^2$, which implies independent components. ◊

For max-stable dfs G the marginals G_i are monotone transformations, $G_i \in \mathcal{M}^{\uparrow}$, the copula C is unique, and $G = C(G_1, \dots, G_d)$. Since $G^t = G \circ \gamma^{\log t}$ has the same copula as G, it follows that $G^t = C(G_1^t, \dots, G_d^t)$. One may show, see Galambos [1978], that the copulas which describe the dependency in max-stable distributions are precisely the copulas C which satisfy

$$C^t(u) = C(u_1^t, \dots, u_d^t), \quad t \in \mathbb{R}.$$

If one normalizes by power-scaling, $x \mapsto a|x|^c \operatorname{sign}(x)$, see Mohan & Ravi [1992], the *uniform distribution* on $(0,1)$ is a limit law for maxima, and the relation above has a probabilistic interpretation.

7.4 Max-stable distributions with exponential marginals on $(-\infty, 0)$. Let $\mathcal{MSE}$ denote the set of all max-stable dfs G with standard *exponential marginals* on $(-\infty, 0)$. This set of limit laws does not score high on applications. There are two properties which make it interesting for the theory though. The exponent measure may have mass on the coordinate planes and axes in $-\infty$; so it enables us to understand this important phenomenon. There is a close link to copulas. The sample copula introduced above is closely related to the *Poisson copula*, a Poisson point process with standard Poisson marginals. The exponent measures of dfs in $\mathcal{MSE}$ generate Poisson copulas.

To study the dependency relation for max-stable distributions, we are free to choose the marginals as we like. The marginals determine the form of the stability relation (7.2). For $G \in \mathcal{MSE}$ the relation has a simple form. Since the standard exponential df F on $(-\infty, 0)$ has the property $F^t(y) = F(ty)$ for $y < 0, t > 0$, the df G satisfies

$$G^t(w) = G(tw), \quad w \in (-\infty, 0]^2, \ t > 0. \tag{7.4}$$

So the df R of the exponent measure ρ is *positive-homogeneous* of degree one:

$$R(tw) = tR(w), \quad w \in (-\infty, 0)^d, \ t > 0,$$
$$\rho(tB) = t\rho(B), \quad t > 0, \ B \text{ a Borel set in } \mathcal{X} = [-\infty, 0)^d \setminus \{(-\infty, \ldots, -\infty)\}.$$

We want to give an idea of what these measures ρ look like.

Take a finite set of points $a \in \mathcal{X}$ and for each such a let $\rho(a)$ be a multiple of one-dimensional *Lebesgue measure* on the ray

$$L(a) = \{ta \mid t \in (0, \infty)\}.$$

The measures $\rho(a)$ satisfy the stability relation, and so does any positive combination ρ of these measures. We may always choose the points a in the compact set

$$S = \{w \in \mathcal{X} \mid \max_i w_i = -1\},$$

since any ray $L(a)$ will intersect S in precisely one point. Normalize $\rho(a)$ so that $\{ta \mid 0 < t \leq 1\}$ has mass one. Then any finite measure μ on S determines a measure $\rho = \int_S \rho(a) d\mu(a)$ (mixture) which satisfies the stability relation and which is finite on $[-\infty, 0)^d \setminus [-\infty, z]$ for any $z \in (-\infty, 0)^d$. The univariate marginals of ρ are positive multiples of standard one-dimensional *Lebesgue measures* on $(-\infty, 0)$. Proper scaling makes them standard.

The points a above may have certain coordinates equal to $-\infty$. In that case the halfline $L(a)$ lies in the corresponding coordinate planes in $-\infty$. One may avoid working with mass in infinity by taking into account the marginal distributions. This approach is particularly well suited for handling min-stable distributions with exponential marginals on the positive halfline, see Section 15.3.

Theorem 7.9 (Decomposition). *For each non-empty set* Kß$\{1,\dots,d\}$*, let* ρ^K *be a measure on* $(-\infty,0)^K$ *which satisfies the stability relation* $\rho^K(tE) = t\rho^K(E)$ *for* $t > 0$*,* E *a Borel set in* $(0,\infty)^K$*. Suppose the Möbius relations hold: for any non-empty* Iß$\{1,\dots,d\}$

$$\sum_{\{K|K\supset I\}} (-1)^{|K\setminus I|} p_I(\rho^K) \le \rho^I,$$

where p_I *is the natural projection onto* $(-\infty,0)^I$ *for* IßK*. Suppose moreover* $\rho^{\{k\}}$ *is Lebesgue measure on* $(-\infty,0)$ *for* $k = 1,\dots,d$*. There is a one-one correspondence between such families of measures* ρ^K *and dfs in* $\mathcal{MSE}$*.*

Proof. The representation is similar to the decomposition of a probability distribution on $[0,\infty)^d$ into a part on $(0,\infty)^d$ and 2^{d-1} parts on the faces, edges and vertices, or the relation between the df of Z and of $-Z$ for vectors. For dimension $d = 2$ or $d = 3$ it is obvious how to write the measure ρ on $\mathcal{X}$ in terms of the measures ρ^K on $(-\infty,0)^K$. In the general case a *Moebiust@Möbius transform* is involved. □

In dimension $d = 2$ the df $G \in \mathcal{MSE}$ is determined by ρ^{12}.

Example 7.10. A df $G \in \mathcal{MSE}$ has *independent* components if and only if ρ^K vanishes for all Kß$\{1,\dots,d\}$ with more than one element. See Example 7.2. ◊

Example 7.11. In dimension $d = 2$ a triangle with vertices $q = (1,1), (a,0), (0,b)$ with $a, b \in (0,1)$ determines a unique copula C whose mass is concentrated on the sides of this triangle, uniformly on each side. The copula C lies in $\mathcal{D}^\vee(G)$, where $G \in \mathcal{MSE}$ has exponent measure ρ concentrated on two lines. The df $R = -\log G$ satisfies $R(w) = C(q + w)$ for $w_1 + w_2 > -1$.

The sample cloud N_n from C translated to the square $(-1,0)^2$, and then blown up by a factor n, will be close to the Poisson point process M with mean measure ρ if we restrict points to $w_1 + w_2 > -\varepsilon n$, $0 < \varepsilon \le 1$. Both are sample point process mixtures from the same distribution. For the one the mixing distribution is binomial-(n,ε), for the other Poisson-εn. ◊

What can one say about the extreme points of the Poisson point process N whose mean measure ρ is the exponent measure of a df $G \in \mathcal{MSE}$?

Example 7.12. For simplicity take $d = 2$, and assume ρ has a continuous strictly positive density on $(-\infty,0)^2$. We also allow mass on the lines in $-\infty$. The boundary of the *convex hull* of N is a decreasing piecewise linear curve Γ in the third quadrant.

The curve is linear between successive vertices w_i and w_{i+1} with $w_i = (u_i, v_i)$ and $u_i < u_{i+1}$. We claim that there are only finitely many vertices. Hence Γ is the union of a finite number of edges $[w_i, w_{i+1}]$, $i = 1, \dots, m-1$, together with a horizontal halfline $(-\infty, u_1] \times \{v_1\}$ and a vertical halfline $\{u_m\} \times (-\infty, v_m]$.

To see this, first assume ρ lives on $(-\infty, 0)^2$. The coordinate projection $\zeta_1(N)$ is a standard Poisson process on $(-\infty, 0)$, and so is $\zeta_2(N)$. Let V_1 be the maximal point of $\zeta_2(N)$. There is a point $(U_1, V_1) \in N$. This is the first point in the finite sequence of vertices. The condition on the density of ρ ensures that $\rho(H)$ is infinite for all halfplanes $\{v \geq \varepsilon u + c\}$ with $\varepsilon > 0$. Hence the convex hull of N is indeed bounded by this curve Γ.

If ρ also charges the vertical halfline $L = \{-\infty\} \times (-\infty, 0)$, then on L the Poisson point process has intensity $\lambda = \rho(\{-\infty\} \times (-1, 0))$, and the maximal point on L is $(-\infty, Y_1)$, where $-Y_1$ is exponential with parameter λ. By independence of $1_L dN$ and the restriction N_0 of N to $(-\infty, 0)^2$, there is a positive probability that Y_1 exceeds V_1, the maximum of $\zeta_2(N_0)$. This has no effect on the convex hull. The point in infinity does not contribute. In the extreme case where $\rho(\mathbb{R}^2) = 0$ the convex hull is empty. ◊

Exercise 7.13. Let $Z_1, Z_2, \dots$ be independent observations from the distribution π on the square $(-1, 0)^2$. Assume π is a mixture of the *uniform distribution* on the diagonal and on the counterdiagonal of the square. Discuss the collapse of the convex hull of the normalized sample cloud $N_n = \{nZ_1, \dots, nZ_n\}$ as $n \to \infty$. ◊

7.5* Max-stable distributions under monotone transformations. In the literature it is often assumed that all marginals of the df belong to the domain of attraction of the same univariate extreme value distribution, and that they have the same univariate normalizations. See Mikosch [2006]. Grounds for these assumptions are:

1) Simplicity of exposition. In the example above the measures $\rho(a)$ live on the halflines $L(a)$. With marginal distributions of different types one has to replace halflines by rather complicated curves.

2) Applicability. If the maxima in the different coordinates have different rates of growth then, as the sample size increases, the lighter tailed coordinate maxima will become negligible.

3) Invariance of maxima under monotone transformations. If the df is continuous, then one may use a monotone transformation on each of the coordinates to obtain a new df whose marginals are well-behaved at the upper endpoint.

We shall now briefly describe the effects of such *multivariate monotone transformations*

$$\varphi(z_1, \dots, z_d) = (\varphi_1(z_1), \dots, \varphi_d(z_d)), \quad \varphi_i \in \mathcal{M}^{\uparrow},\ i = 1, \dots, d. \tag{7.5}$$

Definition. A vector Z or its df F lies in the *max-stable domain for monotone transformations* if there exist multivariate monotone transformations φ_n, and a vector W with non-constant components such that $\varphi_n^{-1}(Z) \Rightarrow W$. We write $Z \in \mathcal{D}^{\uparrow}(W)$.

Since CATs are multivariate monotone transformations, $Z \in \mathcal{D}^{\uparrow}(W)$ if $Z \in \mathcal{D}^{\vee}(W)$.

If $Z \in \mathcal{D}^{\uparrow}(W)$, then this also holds for the d components of Z. So by Theorem 6.4 we may assume that the components of W have a standard exponential distribution on $(-\infty, 0)$. Moreover the marginals of the df of Z satisfy the basic tail conditions (6.11).

For $F \in \mathcal{D}^{\uparrow}(W)$ there is a monotone transformation $\psi(z) = (\psi_1(z_1), \dots, \psi_d(z_d))$ such that $H = F \circ \psi$ has upper endpoint 0 and

$$1 - H_i(-s) \sim s, \quad s \to 0+ .$$

See the proof of Theorem 6.4. It follows that

$$n(1 - H_i(v/n)) \sim -\log H_i^n(v/n) \to -v, \quad v \in (-\infty, 0).$$

This determines the monotone normalizing transformations for H

$$n(1 - H(u/n)) \sim -\log H^n(u/n) \to R(u), \quad u \in (-\infty, 0)^d. \tag{7.6}$$

The limit df $G = e^{-R}$ has exponential marginals on $(-\infty, 0)$. The normalizations are CATs. Hence $G \in \mathcal{MSE}$. Moreover H shifted over $q = (1, \dots, 1)$ is tail asymptotic to C for any copula C of F by the lemma below.

Lemma 7.14. *Let C be a copula of H, where the marginals H_i satisfy the basic tail conditions (6.11). Suppose that H and C have the same upper endpoint $q = (1, \dots, 1)$, and that the marginals satisfy $1 - H_i(1 - s) \sim s$ for $s \to 0+$. Then H and C are tail asymptotic:*

$$(1 - H(w_n))/(1 - C(w_n)) \to 1, \quad w_n \in [0, 1]^d \setminus \{q\}, \ w_n \to q. \tag{7.7}$$

Proof. Let $w_n \in [0, 1]^d \setminus \{q\}$ and $w_n \to q$. Define $u_{ni} = H_i(w_{ni})$. By assumption $(1-u_{ni})/(1-w_{ni}) \to 1$. Since $H(w_n) = C(u_n)$ by definition of copula, we need that $(1-C(u_n))/(1-C(w_n)) \to 1$. This holds since $|C(w)-C(u)| \le \sum_i |w_i - u_i||$ (as is seen by covering the difference $[0, w] \setminus [0, u]$ by coordinate slices), and $1-C(w) \ge \max_i(1 - w_i)$ (by the same argument). □

There is a simple characterization of $\mathcal{D}^{\uparrow}$ in terms of copulas. The d marginal upper tails of $F \in \mathcal{D}^{\uparrow}$ are asymptotically continuous, and the left hand derivative of the copula at its upper endpoint $q = (1, \dots, 1)$ exists along each line with slope $w \in (0, \infty)^d$.

Theorem 7.15. *Suppose the marginals F_i of F satisfy the minimal tail conditions in* (6.11). *Let C be a copula for F. For each $w \in (-\infty, 0)^d$ let there exist a constant $R(w) \in [0, \infty)$ such that*

$$\frac{1 - C(q + sw)}{s} \to R(w), \quad s \to 0+, \tag{7.8}$$

where $q = (1, \dots, 1)$ is the upper endpoint of C. Then $F \in \mathcal{D}^{\uparrow}(G)$ for $G = e^{-R}$.

Conversely, if $F \in \mathcal{D}^{\uparrow}(G)$ for a df G with non-degenerate marginals, then (7.8) *holds for any copula C of F, and $G = e^{-R \circ \varphi}$ for a multivariate monotone transformation φ.*

Proof. Given F, choose a multivariate monotone transformation φ such that $H = F \circ \varphi^{-1}$ satisfies (7.7). If (7.8) holds, we may replace C by H. Hence $H \in \mathcal{D}^{\uparrow}(e^{-R})$ and $F \in \mathcal{D}^{\uparrow}(G)$ with $G = e^{-R \circ \varphi}$. Conversely, $F \in \mathcal{D}^{\uparrow}(G)$ implies that the marginals satisfy the minimal tail conditions by Theorem 6.4. So $H = F \circ \varphi^{-1}$ satisfies (7.7) for any copula C for F. The transformed df H lies in $\mathcal{D}^{\uparrow}(G \circ \varphi^{-1})$. By the argument above (7.6) holds, and hence (7.8). $\square$

For minima there is a similar result. The differentiability condition should now hold at the lower endpoint: $C(sw)/s \to R(w)$ for $w \in (0, \infty)^d$. It should also hold for all $2^d - d - 2$ marginal copulas of dimension $k = 2, \dots, h$. The differentiability conditions at the upper and lower endpoints of the copula are not as innocent as they look.

Multivariate non-linear normalizations were introduced in Pancheva [1988]. The approach there is in terms of the group of max-automorphisms $\varphi_1 \otimes \cdots \otimes \varphi_d$, where each φ_i is an increasing homeomorphism of the real line onto itself, and *Gumbel marginals*. The approach via exponential marginals has the advantage that it uses the well-developed theory of copulas.

7.6 Componentwise maxima and copulas. For random vectors whose marginal dfs lie in $\mathcal{D}^{+}(0)$, one may use both linear and non-linear normalizations.

Theorem 7.16. *Suppose Z has df F with marginals which satisfy $1 - F_i \sim e^{-\psi_i}$, where ψ_i satisfies* (6.4) *for $i = 1, \dots, d$. Set $X = (X_1, \dots, X_d)$ with $X_i = \psi_i(Z_i)$. Define $b_n = (b_{n1}, \dots, b_{nd})$ and $A_n = \operatorname{diag}(a_{n1}, \dots, a_{nd})$ by*

$$\psi_i(b_{ni}) = \log n, \quad a_{ni} = 1/\psi'(b_{ni}), \quad i = 1, \dots, d.$$

Set $q = (1, \dots, 1) \in \mathbb{R}^d$. The following are equivalent:

$$X^{\vee n} - (\log n) q \Rightarrow W,$$
$$A_n^{-1}(Z^{\vee n} - b_n) \Rightarrow W.$$

Proof. The df H of X satisfies $1 - H \circ \psi \sim 1 - F$ with $\psi = \psi_1 \otimes \cdots \otimes \psi_d$, and

$$\psi(b_n + A_n u_n) - (\log n)q \to u, \quad u_n \to u \in \mathbb{R}^d,$$

since $\psi_i(b_{ni} + a_{ni} v_n) - \psi_i(b_{ni}) \to v$ for $v_n \to v$. Hence the limit relations

$$n(1 - F(b_n + A_n u_n)) \to R(u),$$
$$n(1 - H((\log n)q + u_n) \to R(u)$$

are equivalent. Here R is the df of the exponent measure of W. □

A similar result holds if the marginals belong to different domains. Distribution functions F in $\mathcal{D}^{\uparrow}$ lie in $\mathcal{D}^{\vee}$ if the univariate marginals F_i exhibit the correct asymptotic behaviour.

Theorem 7.17 (Galambos). *Let C be a copula for the random vector Z in $\mathbb{R}^d$. Suppose $Z_i \in \mathcal{D}^+(\tau_i)$ for $i = 1, \dots, d$. Then $Z \in \mathcal{D}^{\vee}(W)$ for a max-stable vector W if and only if C satisfies (7.8) for all $w \in (0, \infty)^d$.*

The result in Galambos [1978], Theorem 5.2.3, is formulated in terms of *dependency* functions, which are slightly more general then copulas. The proof follows from:

Theorem 7.18 (Copula Convergence). *Let the df F_0 on $\mathbb{R}^d$ have continuous marginals F_{0i} and copula C_0. Let C_n be a copula for F_n with marginals F_{ni} which converge weakly to F_{0i}, $i = 1, \dots, d$. Then the dfs F_n converge weakly to F_0 if and only if the copulas C_n converge weakly to C_0.*

Proof. The relation $F_n = C_n(F_{n1}, \dots, F_{nd})$ together with uniform convergence $C_n \to C_0$ by continuity of C_0 and pointwise convergence of the marginals by continuity of the F_{0i} yields weak convergence $F_n \to F_0$. Conversely if the dfs F_n converge, tightness of (C_n) and uniqueness of C_0 yield weak convergence $C_n \to C_0$. □

The weak left differentiability condition (7.8) in the upper right endpoint of the copula may seem innocuous for dfs with a continuous density and marginals in the domains of univariate attraction. The example below paints a different picture.

Example 7.19. Let $X \in [0, \infty)^d$ have continuous density f with marginals $f_i(t) \sim 1/t^2$ for $t \to \infty$, $i = 1, \dots, d$. In polar coordinates

$$f(e^t \theta) = f_0(\theta, t)/e^{dt+t} \quad t \in \mathbb{R},\ \theta \in \Theta = [0, \infty)^d \cap \partial B.$$

Assume f_0 is bounded and uniformly continuous on $\Theta \times [0, \infty)$. If $X \in \mathcal{D}^{\vee}(\rho)$ then ρ has density $h(e^t \theta) = h_0(\theta)/e^{dt+t}$ and $f_0(\theta, t) \to h_0(\theta)$ uniformly on Θ for $t \to \infty$. Indeed by the Arzelà–Ascoli Theorem any sequence $t_n \to \infty$ has a subsequence $s_n = t_{k_n}$ such that $f_0(\theta, s_n + s)$ converges uniformly on compact subsets of $\Theta \times \mathbb{R}$. By weak convergence the limit has the form above. ◊

III High Risk Limit Laws

8 High risk scenarios

8.1 Introduction. In the remaining lectures we shall develop a mathematical theory for handling the asymptotic behaviour of the tails of multivariate distributions. Our treatment will be geometric. The motivation for our interest in this subject comes from *risk theory* in finance. The material of this chapter is adapted from Balkema & Embrechts [2004].

A major concern in finance is that the market moves off in an undesirable direction. In first instance one may think of the market as a vector of asset prices. The components are the prices of the individual assets, or the logarithms of these prices. In such a situation the dimension d is the number of assets, and may be in the hundreds for a large investor. One could also consider a vector whose components describe different sectors of the market, energy, food, banking, heavy industry, etc., or a vector whose components describe various parameters of interest to finance – price, volatility, exchange rate, and interest rate for instance. The present position is known, $z_0 \in \mathbb{R}^d$. We are interested in the position at some fixed future date, say ten trading days ahead, as is customary in market risk management in banking. The future position is a random vector Z in $\mathbb{R}^d$ whose distribution is concentrated around the value z_0. In risk analysis one is interested in the probability that the vector Z will lie in some risky region far out. The direction and the precise form of the region will depend on the build up of one's portfolio.

Due to options and other derivatives, the profit loss distribution need not be a linear functional of the market position Z. Assume that the functional relation between the *loss* and the position Z is known. Then one may write the loss as $f_\sigma(Z)$, where the parameter σ denotes the build up of the portfolio. We assume that for every parameter value σ the function $f_\sigma : \mathbb{R}^d \to \mathbb{R}$ is continuous and unbounded, and vanishes in z_0. Given a good estimate of the distribution of Z, one may simulate a large sample from this distribution, and use this, or a subset, $\{z_1, \dots, z_n\}$, to determine a corresponding empirical distribution for the loss. Given a *risk level*, $\alpha = 0.05$ or $\alpha = 0.01$, the $[n\alpha]$ largest values $f_\sigma(z_k)$ will describe the distribution of the loss exceeding the VaR for the given risk level α. For α small the VaR is large, we speak of *high risk*, and the corresponding large loss values $f_\sigma(z_k)$ will correspond to points z_k far out (with respect to the base value z_0). In a more advanced model, the relation between the position z of the vector Z at the time horizon and the corresponding loss is not given by a function, but for each z and σ by a probability distribution concentrated around the value $f_\sigma(z)$. The random variable describing the loss at the (ten day say) time horizon then has the form $f_\sigma(Z, U)$, where without loss of generality one may take U

to be a random variable which is independent of Z, and with a simple distribution. The simulation of a sample from the loss distribution is as described above, except that one now also needs a sample $\{u_1, \dots, u_n\}$ from the distribution of U.

The situation is particularly simple if loss is a linear function of the position. This is the situation for a *portfolio* consisting of d assets with prices $Z_1, \dots, Z_d$ at the time horizon. We then introduce new coordinates and write $Z = (X, Y)$ where Y, the vertical coordinate, is minus the value of the portfolio at the time horizon, and $X = (X_1, \dots, X_h)$, with $h = d - 1$, is the horizontal part of the vector Z. In this situation Y is a linear functional of Z which determines hyperplanes on which the value of the portfolio is constant. If the VaR at risk level $\alpha = 0.05$ equals q then q is the upper $1 - \alpha$*-quantile* of the distribution of Y. In the univariate case there is a well developed theory for the asymptotic distribution of Y conditional on $Y \geq s$ for $P\{Y \geq s\} \to 0+$ in terms of generalized Pareto distributions on $[0, \infty)$. We are interested in the distribution of the vector $Z = (X, Y)$ on the halfspace $H = \{y \geq s\}$. The horizontal part X of the vector may be regarded as noise. It does not affect the value of the portfolio. On the other hand, one may argue that X contains extra information. A (small) change in the constitution of the portfolio corresponds to a (small) change in the direction of the halfspace H. The effect of such a change on the profit-loss distribution depends on the distribution of the vector Z, not only on the distribution of the vertical coordinate.

Definition. Given a random vector Z in $\mathbb{R}^d$, for any closed halfspace H in $\mathbb{R}^d$ with $\mathbb{P}\{Z \in H\} > 0$ the *high risk scenario* Z^H is defined as the vector Z conditioned to lie in the halfspace H. If Z has distribution π then Z^H has the *high risk distribution* π^H, where

$$d\pi^H(z) = 1_H(z) d\pi(z)/\pi(H).$$

In general we use the term scenario to describe a change in the underlying probability distribution. By conditioning on the region of interest, which for simplicity we assume to be a closed halfspace, we are able to develop a truly multivariate theory of exceedances. In this theory we study the asymptotic behaviour of high risk scenarios Z^H for $\mathbb{P}\{Z \in H\} \to 0$. In the basic theory, developed in this chapter, no assumptions are made about the direction in which the halfspaces H diverge.

The theory applies in situations where there is a strong degree of directional homogeneity. It is helpful to envisage in the first instance a vector Z with a non-degenerate multivariate Gaussian distribution, but with heavier tails, or more generally a vector with a unimodal density having convex level sets. For such a distribution, for risk level $\alpha \in (0, 1)$, and for any direction there is a unique halfspace H with this direction such that $\mathbb{P}\{Z \in H\} = \alpha$. John Tukey has used the term *bland* to describe such distributions. There are no marked characteristics in any particular direction. Sample clouds will consist of a black convex central region surrounded by a halo of isolated points thinning out as one moves away from the center. We are interested in an asymptotic description of the local behaviour at the edge of the sample cloud, by restricting the cloud to halfspaces tangent to the convex central region.

The asymptotic theory of high risk scenarios applies to any data set which fits the above description. In life sciences the data might describe a population in terms of d characteristics. The central part denotes the normal population; the viability decreases as one moves away from the center in any direction. In quality control the halo will consist of products which are unacceptable (with respect to a given set of d characteristics). As in finance, the loss function associated with such points need not be linear, but it will increase as one moves outwards.

The mathematical theory for high risk scenarios is based on the following:

Ansatz. For halfspaces H and J with positive (small) probabilities of being hit by Z, and relatively large overlap, the distribution of the high risk scenarios Z^H and Z^J have approximately the same shape.

8.2 The limit relation. In order to compare high risk scenarios on different halfspaces one needs a fixed reference halfspace J_0. Usually J_0 will be the *upper halfspace*

$$H_+ = \{w = (u, v) \in \mathbb{R}^{h+1} \mid v \geq 0\} = \mathbb{R}^h \times [0, \infty), \quad h = d - 1. \tag{8.1}$$

Any halfspace H is the image of J_0 under an *affine transformation*

$$\alpha \colon w \mapsto z = \alpha(w) = a + Aw, \quad a \in \mathbb{R}^d, \ \det(A) \neq 0.$$

The inverse α^{-1} maps H onto J_0 and maps the high risk scenario Z^H into a vector on J_0. We shall assume that there is a limit vector W with a non-degenerate distribution ρ_0 on the halfspace J_0: For every halfspace H with positive mass $\pi(H) = \mathbb{P}\{Z \in H\}$, one may choose an affine transformation α_H mapping J_0 onto H such that

$$\alpha_H^{-1}(Z^H) \Rightarrow W, \quad \pi(H) \to 0+. \tag{8.2}$$

In order to make sense of this definition, we assume that $\pi(H)$ tends to zero when H diverges in any direction. Hence we impose the following regularity *condition* on the *boundary* of the distribution π:

$$\pi(H) > 0 \Rightarrow \pi(\partial H) < \pi(H), \tag{8.3}$$

where ∂H denotes the topological boundary of the set H, the bounding hyperplane. An equivalent formulation of this condition is: For any non-zero linear functional ξ, the distribution of the random variable ξZ is continuous in its upper endpoint.

Definition. Let Z be a random vector with distribution π on $\mathbb{R}^d$, and W a random vector with distribution ρ_0 on the halfspace J_0ß$\mathbb{R}^d$. Then W is called a *high risk limit* *vector*, and ρ_0 a *high risk limit distribution* for the vector Z (or for the distribution π), if the distribution ρ_0 is *non-degenerate* (it does not live on a hyperplane in $\mathbb{R}^d$), if π satisfies the regularity condition (8.3), and if there exist affine transformations α_H mapping J_0 onto H such that (8.2) holds. In this case we write

$$Z \in \mathcal{D}(W), \quad \pi \in \mathcal{D}(\rho_0),$$

and say that Z (or π) lies in the *domain of attraction* of W (or ρ_0).

Two natural questions are:

1) What are the possible limit laws?

2) What do the domains of attraction look like?

These questions have not yet been answered in full generality. We shall present a continuous one-parameter family of high risk limit laws, the multivariate generalized Pareto distributions, which generalize the standard univariate GPDs on $[0, \infty)$, and for each of these laws we shall exhibit a rich class of probability densities contained in their domain of attraction. These densities contain the Gaussian, the spherical Student and the hyperbolic densities.

Before proceeding with the theory, we present three examples.

8.3 The multivariate Gaussian distribution. Assume Z has a multivariate Gaussian distribution. Since our theory is geometric and does not depend on coordinates, we may assume that Z has density $e^{-z^T z/2}/(2\pi)^{d/2}$. Let H be a halfspace. By *spherical symmetry* of the distribution, we may assume that H is the horizontal half-space $H^t = \{y \geq t\}$, where we write $z = (x, y) \in \mathbb{R}^{h+1}$. Set $Z = (X, Y)$, and let Y^t denote the random variable Y conditioned to exceed the level t. Then Y^t has density

$$f_t(y) = e^{-y^2/2} 1_{[t,\infty)}(y)/c_t,$$

where c_t is the integral of $e^{-y^2/2}$ over $[t, \infty)$. (This ensures that f_t is a probability density.) It is convenient to write

$$Y^t = t + V_t/t,$$

where V_t is a non-negative random variable. The factor $1/t$ is an extra scaling. The variable V_t has density

$$g_t(v) = f_t(t + v/t)/t = h_t(v) 1_{[0,\infty)}(v)/c'_t, \quad h_t(v) = e^{-v} e^{-v^2/2t^2} \to e^{-v}, \tag{8.4}$$

for $t \to \infty$ since $(t + v/t)^2/2 = C_t + v + v^2/2t^2$. Convergence holds in $\mathbf{L}^1([0, \infty))$ by the monotone convergence theorem. So $V_t \Rightarrow V$ for $t \to \infty$, where V is standard exponential; see EKM for more details.

Introduce the affine transformations

$$\alpha_t(u,v) = (u, t + v/t).$$

Then $\alpha_t^{-1}(X, Y^t) = (X, V_t)$, and we see that $\alpha_t^{-1}(Z^{H^t}) \Rightarrow W = (U, V)$, where U is standard normal on $\mathbb{R}^h$ and V is standard exponential and independent of the vector U. By spherical symmetry this limit relation holds in every direction: We may choose affine transformations α_H mapping H_+ onto H such that

$$\alpha_H^{-1}(Z^H) \Rightarrow W, \quad \mathbb{P}\{Z \in H\} \to 0,$$

where W has the *standard Gauss-exponential distribution* ρ_0 on H_+ with density

$$g_0(w) = e^{-v} 1_{[0,\infty)}(v) e^{-u^T u/2}/(2\pi)^{h/2}, \quad w = (u,v) \in \mathbb{R}^{h+1}.$$

Here we have a non-trivial multivariate high risk limit distribution!

Observe that the Gauss-exponential density has a natural extension to $\mathbb{R}^d$. The Radon measure ρ on $\mathbb{R}^d$ with density $e^{-(v+u^T u/2)}/(2\pi)^{h/2}$ has a large group of *symmetries*. We include the vertical translations:

$$\tau^t(\rho) = e^t \rho, \quad \tau^t(u,v) = (u, v+t).$$

The dimension of the symmetry group of ρ is larger than that of the standard Gaussian distribution on $\mathbb{R}^d$, as one easily checks for dimensions two and three. A parabola is more symmetric than a circle! For a finite measure a symmetry is mass-preserving; for infinite measures the *symmetries* satisfy $\sigma(\rho) = c\rho$ for a positive constant c.

Lemma 8.1. *For any halfspace*

$$H = \{v \geq c_0 + c_1 u_1 + \cdots + c_h u_h\}, \quad c_0, \ldots, c_h \in \mathbb{R}, \tag{8.5}$$

there is a symmetry σ of ρ mapping H_+ onto H.

Proof. On the parabola any point becomes the top in an appropriate system of coordinates. Here are the details: Let $\gamma \in \mathcal{A}$ have the form

$$\gamma(u,v) = (u + b, v - b^T u - b^T b/2). \tag{8.6}$$

Write $w = (u,v)$. Then $\gamma^{-1}(\rho)$ has density

$$\begin{aligned} g(\gamma(w))|\det\gamma| &= g(u+b, v - b^T u - b^T b/2) \\ &= e^{-(u+b)^T(u+b)/2} e^{-v + b^T u + b^T b/2}/(2\pi)^{h/2} = g(w). \end{aligned}$$

So $\gamma^{-1}(\rho) = \rho$, and γ^{-1} and γ are symmetries. Now observe that

$$w \in \gamma^{-1}(H^t) \iff \gamma(w) \in H^t \iff v - b^T b/2 - b^T u \geq t.$$

The halfspace in (8.5) has the form $H = \gamma^{-1}(\tau^t(H_+))$ if we choose $b = (c_1, \ldots, c_h)$ and $t = c_0 - b^T b/2$. □

The probability distribution obtained by conditioning ρ to $H = \sigma(H_+)$ is Gauss-exponential:

$$1_H d\rho/\rho(H) = \sigma(\rho_0).$$

Observe that the convergence in (8.4),

$$e^{-v-v^2/2t^2} \to e^{-v}, \quad t \to \infty,$$

holds for all $v \in \mathbb{R}$, uniformly on compact sets, and in $\mathbf{L}^1([c,\infty))$ for any $c \in \mathbb{R}$. Choose t_n such that $\mathbb{P}\{Y \geq t_n\} = 1/n$. Then

$$\rho_n = n\pi_n \to \rho, \quad \pi_n = \alpha_{t_n}^{-1}(\pi)$$

vaguely on $\mathbb{R}^d$, and weakly on each horizontal halfspace $\mathbb{R}^h \times [c,\infty)$, $c \in \mathbb{R}$, and also on any halfspace H of the form (8.5), by spherical symmetry of the distribution π. Now observe that ρ_n is the mean measure of the n-point sample cloud N_n consisting of n independent normalized observations $W_k = \alpha_{t_n}^{-1}(Z_k) = (X_k, t_n(Y_k - t_n))$. Vague convergence $\rho_n \to \rho$ implies

$$N_n \Rightarrow N \text{ vaguely on } \mathbb{R}^d,$$

where N is a Poisson point process with mean measure ρ. Weak convergence holds on all halfspaces H of the form (8.5), since $\rho_n(H) \to \rho(H)$ and $\rho(\partial H) = 0$.

For the Gaussian vector Z the normalizations α_H^{-1} which yield a limit distribution for the high risk scenarios Z^H also yield a limit for the sample clouds.

8.4 The uniform distribution on a ball. The second example concerns a probability distribution with bounded support. It is of interest because of its simplicity.

Start with a vector Z which is uniformly distributed on an open ball D in $\mathbb{R}^d$. Let H be a closed halfspace which intersects the ball. Then Z^H is uniformly distributed on the cap $D \cap H$. By symmetry we may take the halfspace to be horizontal. We assume that D is the unit ball centered in $(0,-1)$ in $\mathbb{R}^{h+1}$. This ball supports the upper halfspace in the origin.

The caps $D \cap H_t$ with $H_t = \{y \geq -t\}$ for $t \to 0+$ may be normalized to converge to a parabolic cap. Take the basic halfspace to be $J_0 = \{v \geq -1\}$, and set

$$\alpha_t(u,v) = (\sqrt{t}u, tv), \quad t > 0.$$

Then $H_t = \alpha_t(J_0)$ and the linear diagonal map α_t^{-1} transforms the ball D into the open ellipsoid E_t centered on $(0,-1/t)$ and supporting H_+ in the origin:

$$E_t = \{tu^Tu + (tv+1)^2 < 1\} = \{u^Tu + 2v < -tv^2\}.$$

As $t \to 0+$, the ellipsoids E_t converge to the open *paraboloid* $Q = \{v < -u^T u/2\}$, and the normalized high risk scenarios $W_t = \alpha_t^{-1}(Z^{H_t})$ which are uniformly distributed on the elliptic cap $E_t \cap J_0$ converge in distribution to the random vector W which is uniformly distributed on the parabolic cap

$$Q \cap J_0 = \{-1 \le v < -u^T u/2\}.$$

By symmetry of D, the *uniform distribution* on $Q \cap J_0$ is a high risk limit law, which contains the *uniform distribution* on balls (and open ellipsoids) in its domain of attraction. Since the vertical coordinate V of the limit vector has a power law,

$$\mathbb{P}\{V \ge -t\} \propto \int_0^t s^{h/2} ds \propto t^{(d+1)/2}, \quad 0 \le t \le 1,$$

we shall write $Z \in \mathcal{D}(\tau)$ with $\tau = -2/(d+1)$.

The *paraboloid* Q has a large group of *symmetries*. Rotations around the vertical axis map Q onto itself. So do the linear transformations $\eta(u,v) = (cu, c^2 v)$, and the parabolic translations γ in (8.6). We conclude that all parabolic caps have the same shape, and so too the uniform distribution on these caps.

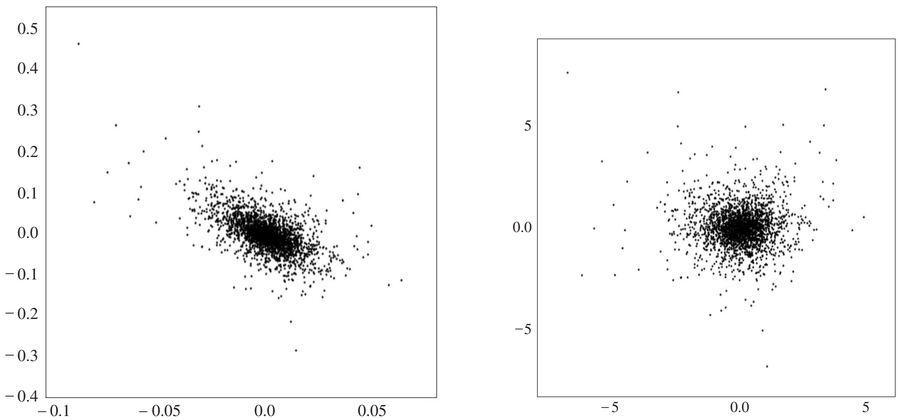
Choose halfspaces H_n so that $n\mathbb{P}\{Z \in H_n\} \to c > 0$. The sample cloud of n independent observations from the uniform distribution on the ball D, normalized by $\alpha_{H_n}^{-1}$, is an n-point sample cloud N_n from the uniform distribution on the (vertical) cigar shaped ellipsoid E_{t_n}, and $N_n \Rightarrow N$ vaguely, where N is the Poisson point process on the paraboloid Q with constant intensity $j = c/|Q \cap J_0|$. By the same arguments as in the previous example, $N_n \Rightarrow N$ weakly on halfspaces H which intersect the open paraboloid Q in a bounded set. The point process N describes the asymptotic behaviour of a large sample cloud from a uniform distribution on a ball at its edge; in particular, the convex hull of the parabolic sample cloud describes the asymptotics of the local behaviour of the convex hull of a uniform sample from the ball.

The map $\Phi\colon (u,v) \mapsto (u, v + u^T u/2)$ maps Q onto the open halfspace $\{v < 0\}$ and preserves *Lebesgue measure*. Hence it maps the Poisson point process N into the Poisson point process M on $\{v < 0\}$ with intensity j. Since Φ maps halfspaces $H = \{v \ge c^T u + c_0\}$ into paraboloids $\{v \ge \|u - u_0\|^2/2 + a_0\}$, the vertices of the convex hull $c(N)$ correspond to points of M supporting such paraboloids. It follows that the projection of the hull $c(N)$ onto the horizontal coordinate plane yields a stationary random triangulation. The same argument applies to the Gauss-exponential point process of the previous example. See Cabo & Groeneboom [1994] for a detailed analysis in dimension $d = 2$, Robinson [1989] for a program to simulate the convex hull of N in $\mathbb{R}^3$, and Baryshnikov [2000] for details on probe processes. For more references on the convex hull, see Finch & Hueter [2004].

The limit paraboloid Q derives from a second-order Taylor expansion of the upper boundary of the ball D in the origin. It will hold for any bounded open convex set whose boundary has continuous positive definite curvature. Such rotund sets will be studied in more detail in Section 9.2.

8.5 Heavy tails, returns and volatility in the DAX. We shall now discuss a bivariate sample which is of some interest in finance. The sample consists of the daily log-returns and the daily changes in the logarithm of the volatility for the German stock exchange (DAX) over the ten year period 1992–2001. It was kindly made available by the Institut für Mathematische Stochastik of Freiburg University. We refer to Zmarrou [2004] for details.

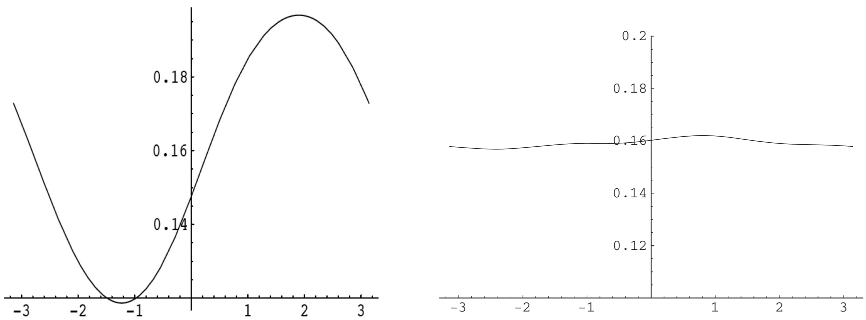
For day traders whose position may consist of both stocks (long or short) and options of various kinds, the relative changes in the price and in the volatility of the underlying stocks are the two variates which determine the profit (or loss). A good estimate of the tail behaviour of these bivariate distributions for the assets in which they are trading would seem to be of interest for determining the build-up of their position, and for evaluating the risk involved.



Log-returns versus changes in the log volatility: data, and the standardized cloud.

The two figures above are from Zmarrou [2004]. The first figure shows the bivariate data: log returns on the horizontal axis and relative changes in the volatility on the vertical axis. The sample cloud is elliptical, showing negative correlation between returns and changes in the volatility. The second figure shows the sample cloud after a simple linear change of coordinates $x = u$, $y = av + bu$, to make the sample *covariance* matrix a multiple of the identity matrix. Since we are interested in the tails, we delete the sample points within a circle of radius $r_0 = 2.5$. The radial coordinate is fitted to a Pareto distribution; for the angular coordinate a density is estimated using a Gaussian kernel on periodically extended data. The data points were divided into six sectors to test independence of the radial and angular coordinate.

The first figure below shows the kernel estimate in Zmarrou [2004] of the angle density on the interval $(-\pi, \pi]$. By a slight change of the origin, in the direction where there is a surplus of points on the circle, one gets rid of the sinusoidal shape of the angle density, as is shown in the second figure below. Since the original data set of 2500 points was reduced to 160 by eliminating the points within the circle of radius r_0, and the model assumption of independence and stationarity are certainly not fulfilled for the data, the tail estimate $g(w) = c/\|w\|^{2+1/\tau}$, with $\tau \approx 0.27$, for the density after the linear transformation and translation, should be looked upon with a certain reserve.



Kernel estimate of the angle density, before and after an additional affine transformation.

For $\tau > 0$ the measure ρ on $O = \mathbb{R}^d \setminus \{0\}$ with density $1/\|w\|^{d+1/\tau}$ is an excess measure. It is symmetric for rotations and scalar expansions. The probability measures $d\rho^H = 1_H d\rho/\rho(H)$ have the same shape for all closed halfspaces HßO. Each probability measure is a high risk limit scenario, the *Euclidean Pareto distribution* . The reader may check that the domain of attraction $\mathcal{D}(\tau)$ contains the spherical Student distribution.

8.6 Some basic theory. Before proceeding to a concrete description of limit laws and their domains of attraction, we present some basic general results.

The limit theory for high risk scenarios is geometric. It does not depend on coordinates.

Theorem 8.2 (Geometric Invariance). *Suppose $Z \in \mathcal{D}(W)$ has normalizations α_H^{-1} mapping H onto J_0. Let β and γ be affine transformations. Then the random vector $Z' = \beta(Z)$ lies in the domain of $W' = \gamma(W)$ with normalizations θ_J^{-1} that map*

$H = \beta^{-1}(J)$ onto $J = \gamma(J_0)$, where

$$\theta_J = \beta\alpha_{\beta^{-1}J}\gamma^{-1}.$$

Proof. First note that W' is non-degenerate and that Z' satisfies condition (8.3). The events $\{Z \in H\}$ and $\{Z' \in J = \beta(H)\}$ are equal. If their common probability is positive, then $(Z')^J = \beta(Z^H)$ in distribution. Hence $\alpha^{-1}_{\beta^{-1}(J)} \circ \beta^{-1}((Z')^J) \Rightarrow W$ for $\mathbb{P}\{Z' \in J\} \to 0+$, and $\theta_J^{-1}((Z')^J) \Rightarrow W' = \gamma(W)$ by continuity of γ. □

The theory is asymptotic. The limit relation depends only on the asymptotic behaviour of the distribution at the edge of its domain.

Let O be the interior of the convex support of the distribution π of Z. The *convex support* is the intersection of all closed halfspaces H with $\pi(H) = 1$. It is possible that $\pi(O) = 0$, even when π satisfies the regularity condition (8.3). (Take Z uniformly distributed on the boundary of a ball.) Altering the distribution on a compact subset K of O does not affect the limit relation. The set O allows us to formulate the condition $\pi(H) \to 0+$ without mentioning the probability measure π.

Exercise 8.3. Under the boundary condition (8.3) convergence $\pi(H_n) \to 0+$ is equivalent to: H_n eventually intersects O and is disjoint from K, for each compact KßO. ◊

If there is a density, it may be replaced by any function which is asymptotic to it. Let us first define asymptotic equality in a multivariate setting.

Definition. Let O be an open set in $\mathbb{R}^d$. A sequence z_n in $\mathbb{R}^d$ *diverges* in O, and we write

$$z_n \to \partial_O, \tag{8.7}$$

if all but finitely many terms of the sequence lie in O, and if any compact subset of O contains only finitely many terms of the sequence. The functions f and g are *asymptotic* on O, and we write $g \sim f$ in ∂_O (or ∂), if for any $\varepsilon > 0$ there exists a compact set KßO such that f and g are defined and positive on $O \setminus K$ and

$$e^{-\varepsilon} < g(z)/f(z) < e^{\varepsilon}, \quad z \in O \setminus K. \tag{8.8}$$

Thus asymptotic equality holds if and only if $g(z_n)/f(z_n) \to 1$ for any divergent sequence (z_n). We shall use the same notation for halfspaces H, and write

$$H \to \partial \tag{8.9}$$

if H intersects the interior O of the convex support of π but is disjoint from K eventually for any compact set KßO; compare Exercise 8.3.

Theorem 8.4 (Asymptotic Invariance). *Let $d\pi_i = f_i d\mu$, $i = 0, 1$, be probability measures on $\mathbb{R}^d$. Let O be the interior of the convex support of μ, and suppose $\mu(\mathbb{R}^d \setminus O) = 0$. Suppose $f_1 \sim cf_0$ in ∂_O for some positive constant c. If $\pi_0 \in \mathcal{D}(\rho_0)$, then $\pi_1 \in \mathcal{D}(\rho_0)$ with the same normalizations.*

Proof. Given $\varepsilon > 0$, there is a compact set KßO such that $e^{-\varepsilon} f_0 < f_1/c < e^{\varepsilon} f_0$ on $O \setminus K$; see (8.8). If $\pi_i(H_n) \to 0+$, then eventually H_n is disjoint from K; see Exercise 8.3. So it suffices to consider high risk scenarios for halfspaces H disjoint from K. Let f_i^H denote the densities with respect to μ. We find $e^{-\varepsilon}\pi_1 H < c\pi_0 H < e^{\varepsilon}\pi_1 H$, and $e^{-2\varepsilon} f_1^H < f_0^H < e^{2\varepsilon} f_1^H$ on H, since the constant c drops out on conditioning. Hence the normalized variables $W_{iH} = \alpha_H^{-1}(Z_i^H)$ satisfy

$$e^{-2\varepsilon}\mathbb{E}\varphi(W_{1H}) \le \mathbb{E}\varphi(W_{0H}) \le e^{2\varepsilon}\mathbb{E}\varphi(W_{1H})$$

for any bounded continuous $\varphi \colon \mathbb{R}^d \to [0, \infty)$. □

Multivariate distributions often are given by their densities. If Z has distribution π and density f, then Z^H has density $f^H(z) = f(z)1_H(z)/\pi(H)$, and the normalized high risk vector $W_H = \alpha_H^{-1}(Z^H)$ has density

$$g_H(w) = f(\alpha_H(w))|\det \alpha_H|/\pi(H), \quad w \in J_0.$$

Determining the factor $\pi(H)$ entails a d-fold integration! In order to prove convergence $\alpha_H^{-1}(Z^H) \Rightarrow W$ it is often simpler to check that

$$h_H(w) = \frac{f(\alpha_H(w))}{f(\alpha_H(w_0))} \to h(w), \quad H \to \partial \tag{8.10}$$

holds almost everywhere on J_0 for some Borel function h on J_0 which is assumed continuous in $w_0 \in J_0$, with $h(w_0) = 1$ (obviously). We also assume that h is integrable.

Let W have density $g \propto h$. Convergence in (8.10) does not ensure $W_H = \alpha_H^{-1}(Z^H) \Rightarrow W$. We need an extra condition.

Proposition 8.5. *In the situation above the following are equivalent:*

1) *$Z \in \mathcal{D}(W)$;*
2) *$(W_H, \pi(H) > 0)$ is tight;*
3) *$(h_H, \pi(H) > 0)$ is uniformly integrable;*
4) *$h_H \to h$ in $\mathbf{L}^1(J_0)$ for $\pi(H) \to 0+$;*
5) *$g_H \to g$ almost everywhere on J_0 for $\pi(H) \to 0+$.*

Proof. Standard. □

Write 5) in terms of h_H and h:

$$g_H(w) = \frac{|\det \alpha_H| f(\alpha_H(w_0))}{\pi(H)} h_H(w) \to g(w).$$

Take $w = w_0$ and use $h_H(w_0) = 1$ to obtain a simple asymptotic expression for high risk probabilities $\pi(H)$ in terms of densities, without integration!

Theorem 8.6 (High Risk Probabilities). *Suppose that W has density g on J_0 and $Z \in \mathfrak{D}(W)$ has density f. Let g be continuous and positive in $w_0 \in J_0$ and suppose (8.10) holds with $h = g/g(w_0)$. Then*

$$\pi(H) \sim |\det \alpha_H| f(\alpha_H(w_0))/g(w_0), \quad H \to \partial. \tag{8.11}$$

Definition. A probability distribution on $\mathbb{R}^d$ is *unimodal* if it has a unimodal density. The function $f : \mathbb{R} \to [0, \infty]$ is *unimodal* or *quasiconvex* if the sets $\{f > c\}$ for $c > 0$ are convex. The sets $\{f > c\}$ will be called *level sets* of the function f.

One may impose the regularity condition that $\{f > c\}$ be open by replacing the density f by its lower semicontinuous version. This regularity condition ensures that the density is uniquely determined by the distribution. Unimodal functions have many pleasant properties.

Lemma 8.7. *Let f_n and f be defined on a convex set in $\mathbb{R}^d$. If the f_n are unimodal, f is continuous, and $f_n \to f$ pointwise, then f is unimodal.*

Proof. Suppose $f(z_i) > c$ for $i = 1, 2$. Then $f(z_i) > c_0$ for some $c_0 > c$, and $f_n(z_i) > c_0$ eventually. By unimodality $f_n > c_0$ on $[z_1, z_2]$ eventually, and $f \geq c_0 > c$ on $[z_1, z_2]$. So $\{f > c\}$ is convex. □

Suppose Z has a bounded unimodal density f and W has a continuous density g on J_0, and $Z \in \mathfrak{D}(W)$. Assume (8.10) holds uniformly on bounded subsets of J_0 with $h = g/g(w_0)$. Then g is unimodal by the lemma, since the h_H are unimodal on J_0. Hence $\{g > \varepsilon\}$ is bounded for all $\varepsilon > 0$. This proves that $h_H \to h$ uniformly on J_0 (and also in $\mathbf{L}^1(J_0)$ by Proposition 8.5). In particular, the functions h_H are uniformly bounded for $\pi(H) < \varepsilon_0$ for some $\varepsilon_0 > 0$. Hence h^t and h_H^t are integrable for $t \geq 1$, and $h_H^t \to h^t$ uniformly on J_0 and in $\mathbf{L}^1(J_0)$. This implies that $g_t = h^t/C_0(t)$ is the density of a high risk limit vector and that the density $f^t/C_1(t)$ lies in its domain. We have the following result.

Theorem 8.8 (Power Families). *Suppose Z with distribution π has a bounded unimodal density f. Let there exist a continuous integrable function h on J_0, affine transformations α_H mapping J_0 onto H, and a point $w_0 \in J_0$ such that (8.10) holds uniformly on bounded subsets of J_0. If $Z \in \mathfrak{D}(W)$, then W has density $g \propto h$, and for each $t \geq 1$ the density $g_t \propto h^t$ is a high risk limit density, whose domain contains the density $f_t \propto f^t$. The normalizations do not depend on t.*

For the standard Gaussian density f and the standard *Gauss-exponential* limit density g, the theorem gives no new high risk limit distributions, since the densities obtained by normalizing f^t and g^t are again Gaussian and Gauss-exponential. We mention two power families:

1) The multivariate *Euclidean Pareto* densities g_s, obtained by restricting the spherical functions $1/\|w\|^{d+s}$, $s > 0$, to the halfspace $J_0 = \{v \geq 1\}$. The domain of g_s contains the *spherical Student* density $f_s(z) \propto 1/(1 + z^T z)^{(d+s)/2}$.

2) The multivariate *parabolic power* densities g_s, obtained by restricting the functions $(v - u^T u)_+^{s-1}$, $s > 0$, on the parabola $\{v > u^T u\}$ to the halfspace $J_0 = \{v \leq 1\}$. The domain of g_s contains the *spherical beta densities* $f_s(z) \propto (1 - \|z\|)_+^{s-1}$ and $h_s(z) \propto (1 - z^T z)_+^{(s-1)/2}$.

Both these power families are asymptotically *Gauss-exponential* for $s \to \infty$. For the multivariate power densities the parameter value $s = 1$ yields the uniform distribution on the parabolic cap $\{u^T u < v \leq 1\}$; the corresponding densities f_1 and h_1 define the uniform distribution on the unit ball. There is a singular high risk limit law for $s \to 0+$. In Section 13 we shall show that these limit laws, suitably normalized, together with the Gauss-exponential distribution, form a continuous one-parameter family of probability distributions on the halfspace H_+, the standardized multivariate GPDs.

9 The Gauss-exponential domain, rotund sets

This section and the next two are devoted to the domain of attraction $\mathcal{D}(0)$ of the Gauss-exponential high risk limit law. The present section introduces a class of structured densities in $\mathcal{D}(0)$. This class extends the one-parameter family of spherical *Weibull densities*

$$e^{-\|x\|^c}/C_c, \quad c > 0. \tag{9.1}$$

We shall formulate two limit theorems and prove pointwise convergence of the quotients associated with high risk densities. The rather technical issue of $\mathbf{L}^1$-convergence has been relegated to the next section. The third section introduces flat functions and roughening of Lebesgue measure.

The subject of the present section is the class $\mathcal{RE}$ of *rotund-exponential* densities. This is a semiparametric class of probability densities in the domain $\mathcal{D}(0)$. The class $\mathcal{RE}$ is broad enough to model sample clouds which have a convex central region surrounded by a rapidly decaying bland halo of isolated points, but which need not fit an elliptic model. A density in $\mathcal{RE}$ is determined by two objects, a convex set which describes the shape of the central black region of the sample cloud, and a decreasing function which describes the tail behaviour of the underlying distribution, the decay rate of the halo of the sample cloud. Precise definitions are given below.

The simple structure of densities in $\mathcal{RE}$ makes it possible to apply standard statistical techniques, for instance by introducing appropriate finite-dimensional subfamilies. Certain results in this section, such as those on the continuity of the normalizations, are typical for spherical distributions. Others are trivial for spherical distributions, by symmetry, for instance uniformity of convergence in different directions. It requires extra effort to handle these in the general setting of $\mathcal{RE}$. The basic issue which we have to face here, and which is peculiar to the asymptotic theory for the domains of the multivariate GPDs, is that of having to handle convergence of multivariate densities, indexed by a multivariate parameter, the halfspace H. This multivariate parameter is specific to the theory in Chapter III. The multivariate parameter H also is the source of the inordinate amount of symmetry of the limiting excess measures: The limit Radon measure has to accommodate d-dimensional families of neighbouring limit distributions, all of the same shape.

9.1 Introduction. In Section 8.3 we saw that $\mathcal{D}(0)$ contains the Gaussian densities: If Z is a non-degenerate Gaussian vector there exist affine normalizations α_H mapping the upper halfspace $H_+ = \mathbb{R}^h \times [0, \infty)$ onto H such that $\alpha_H^{-1}(Z^H) \Rightarrow W$ for $\mathbb{P}\{Z \in H\} \to 0+$. The Gaussian density f satisfies

$$h_H(w) = \frac{f(\alpha_H(w))}{f(\alpha_H(0))} \to h(w) = e^{-(u^T u/2+v)}, \quad w = (u, v) \in \mathbb{R}^{h+1} \tag{9.2}$$

uniformly on bounded sets in $\mathbb{R}^d$, and in $\mathbf{L}^1(H_+)$.

Now consider spherical Weibull densities, $f(z) \propto e^{-\|z\|^c}$. For $c \geq 1$ the density is light tailed, the mgf exists on a neighbourhood of the origin; for $0 < c < 1$ the density has intermediate tails, the mgf is infinite outside the origin but all moments are finite. We shall extend this class of densities in two directions. The Weibull function e^{-r^c} is replaced by a function $e^{-\psi}$ which satisfies the *von Mises conditions*, see (6.4). The Euclidean norm $\|\cdot\|$, based on the unit ball B, is replaced by a gauge function $n = n_D$ based on an *egg-shaped* set D. This yields a *rotund-exponential* density, which by definition has the form

$$f = e^{-\varphi}/C, \quad \varphi(z) = \psi(n(z)). \tag{9.3}$$

One may think of $n(z)$ as the norm of z. The set $D = \{n < 1\}$ is a bounded open convex set containing the origin. We do not assume that it is symmetric, but we do assume that the boundary is C^2 and that the curvature in each boundary point is strictly positive. Such egg-shaped sets will be called *rotund*. Precise definitions and examples are given below.

The function ψ describes the decay rate of the tail; the function n, or rather the set $D = \{n < 1\}$, describes the shape of the level sets. All *level sets* $\{f > c\}, c > 0$, are scaled copies of the basic set D. The conditions on ψ ensure that the level curves

$\{\varphi = i\}$, $i = 1, 2, \ldots$, are asymptotically equidistant; the conditions on the set D ensure that asymptotically these curves locally look like parabolas.

In this chapter we prove two results on rotund-exponential densities.

Theorem 9.1. *Let Z have density $f = e^{-\psi \circ n}/C$, where ψ is an increasing unbounded C^2 function on $[0, t_\infty)$ with $t_\infty \in (0, \infty]$, such that ψ' is positive and $(1/\psi')'$ vanishes in t_∞, and where n is the gauge function of a rotund set D. If t_∞ is finite, f is assumed to vanish outside the set $t_\infty D$. Then Z lies in the domain of the Gauss-exponential law: there exist affine transformations α_H, mapping $H_+ = \mathbb{R}^h \times [0, \infty)$ onto H, such that*

$$W_H = \alpha_H^{-1}(Z^H) \Rightarrow W, \quad \mathbb{P}\{Z \in H\} \to 0+.$$

Moreover all moments converge. Write $W_H = (U_H, V_H) \in \mathbb{R}^{h+1}$. Then

$$\mathbb{E} W_H \to (0, 1), \qquad \operatorname{cov}(W_H) \to I,$$
$$\mathbb{E} V_H^r \to \Gamma(r+1), \quad \mathbb{E}\|U_H\|^r \to 2^{(h+r)/2}\Gamma((h+r)/2)/\Gamma(h/2).$$

Theorem 9.2. *Let the assumptions of the previous theorem hold. Let H_n be half-spaces such that*

$$n\mathbb{P}\{Z \in H_n\} \to c_0 > 0.$$

Let ρ be the Radon measure with density

$$ce^{-(u^T u/2+v)}, \quad c = c_0/(2\pi)^{h/2}.$$

Let $\pi_n = \alpha_{H_n}^{-1}(\pi)$ be the distribution of $\alpha_{H_n}^{-1}(Z)$. Then, for each $m \geq 1$

$$n\pi_n \to \rho \text{ weakly on } \mathbb{R}^d \setminus C_m, \quad C_m = \{v < -m(1 + \|u\|)\}.$$

The theory presented here escapes from the straight jacket of spherically symmetric distributions. The main reason for introducing rotund sets though is that sample clouds in practice often fit a *unimodal distribution*, but may fail to have an elliptic shape.

We shall now first define rotund sets, give some examples, and show how such sets are handled. It will then be shown that the limit relation (9.2) holds uniformly on bounded sets for any density f in (9.3). Theorem 9.2 describes the asymptotic local behaviour of the sample cloud; Section 9.5 describes the global behaviour. The normalizations α_H are determined explicitly in (9.7). We shall see that in certain dimensions it may not be possible to choose α_H to depend continuously on H.

9.2 Rotund sets. One may describe the upper boundary of an open bounded convex set D by a concave function $\partial^+ D \colon D_0 \to \mathbb{R}$, where D_0 is the shadow of D under the vertical projection $(x, y) \mapsto x$. The set D_0ß$\mathbb{R}^h$ is bounded, open and convex. For each $x \in D_0$ the vertical line through $(x, 0)$ intersects D in an open interval with endpoints $a < b$ depending on x. The upper endpoint is $b = \partial^+ D(x)$. If the set D is rotund, then $\partial^+ D$ is C^2 and the second derivative is negative definite in each point $x \in D_0$. If these conditions hold for any choice of the vertical coordinate, then D is rotund. This gives a local description of rotund sets. The boundary ∂D is a C^2 manifold, whose curvature is positive in each point.

There is an alternative global description which is more convenient to use:

Any convex bounded open set Dß$\mathbb{R}^d$ which contains the origin determines a unique function

$$n = n_D \colon \mathbb{R}^d \to [0, \infty)$$

with the properties:

1) $n(tz) = tn(z)$ for $t \geq 0$, $z \in \mathbb{R}^d$;

2) $D = \{n < 1\}$.

Definition. Let Dß$\mathbb{R}^d$ be a bounded open convex set in $\mathbb{R}^d$ which contains the origin. The function $n = n_D$ defined above is the *gauge function* of the set D. The set D is called *rotund* if n is C^2 outside the origin and if

$$n^* = n'' + n' \otimes n' \tag{9.4}$$

is positive definite in each point $z \neq 0$.

In terms of coordinates, one may write $n^*(z)$ as a symmetric matrix with entries

$$c_{ij}(z) = \frac{\partial^2 n(z)}{\partial z_i \partial z_j} + \frac{\partial n(z)}{\partial z_i} \frac{\partial n(z)}{\partial z_j}.$$

The second term $n' \otimes n'$ is needed here since n is linear on rays, and hence n'' cannot be positive definite. Alternatively one could impose the condition that $x \mapsto n(x, 1)$ has a continuous positive definite second derivative in $x \in \mathbb{R}^h$, and similarly for the other $2d - 1$ such functions. By the implicit function theorem, these $2d$ conditions are equivalent to the corresponding conditions in terms of the $2d$ upper boundary functions $\partial^+ D$.

There are many alternative definitions. Observe that $(n^s)' = sn^{s-1}n'$, and hence

$$(n^s)'' = sn^{s-1}n'' + s(s-1)n^{s-2}n' \otimes n'.$$

For rotund sets, n^s has a positive definite second derivative for $s > 1$. In particular, for $s = 2$ the second derivative is homogeneous of degree zero.

Proposition 9.3. *The gauge functions of rotund sets form a cone.*

Proof. Let n_i be the gauge function of the rotund set D_i for $i = 1, 2$, and let c_1 and c_2 be positive constants. Then $n = c_1 n_1 + c_2 n_2$ is the gauge function of a rotund set, since n is positive outside the origin, n is positive-homogeneous of degree one, n is C^2 outside the origin, and $n_{xx}(x, 1)$ and $n_{xx}(x, -1)$ are positive definite, as are the other $2d - 2$ second-order partial derivatives of size h. □

The next result is trivial.

Proposition 9.4. *Linear transformations of rotund sets are rotund.*

We now present some examples of rotund sets.

Example 9.5. Let $D = B^r(p)$ be the open ball of radius $r > 0$ centered in p with $\|p\| < r$. Then D is rotund. For $c \in (-1, 1)$ the following expressions are equivalent:

$$t^2 = x^T x + (y + ct)^2, \qquad t > 0$$
$$t = \sqrt{x^T x + (1 - c^2)y^2} + cy,$$

and $t = n(x, y)$ is the gauge function of the unit ball centered in $(0, c) \in \mathbb{R}^{h+1}$. The function $f = e^{-n}$ has level sets which are balls, but these balls are not concentric. So f is not spherically symmetric. ◊

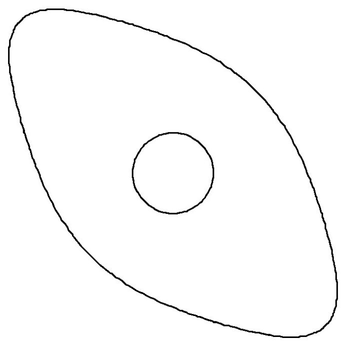
Example 9.6. Let $P(x, y) = x^4 + 4ax^3y + 6bx^2y^2 + 4cxy^3 + y^4$. The function $n = P^{1/4}$ is homogeneous of degree one. The set $\{n < 1\}$ is rotund if P'' is positive definite outside the origin. This is the case if $a^2 < b$, $c^2 < b$, and if

$$Q(x) = (b - a^2)x^4 + 2(c - ab)x^3 + (1 + 2ac - 3b^2)x^2 + 2(a - bc)x + b - c^2$$

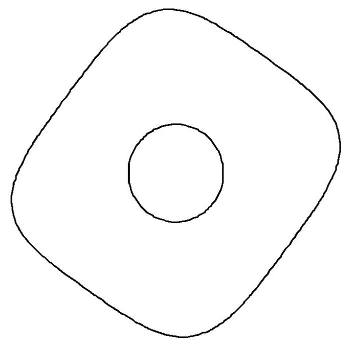
is positive for all x. Below are two solutions of $P(x, y) = 1$, together with a unit circle. ◊

Example 9.7. Let $\varphi\colon O \to [0, \infty)$ be a convex function on an open convex set. Assume φ is C^2, φ'' is positive definite, $\varphi(0) = 0$ and $\varphi(z) \to \infty$ for $z \to \partial_O$. Then all level sets $\{\varphi < c\}$, $c > 0$, are rotund. Cumulant generating functions are convex analytic functions such that $\varphi(0) = 0$. If the mgf exists on a neighbourhood of the origin, then its domain is a convex set. The second derivative φ'' is a variance, and is positive definite if the underlying distribution is non-degenerate, and $\varphi'(0) = 0$ if the distribution is centered. If the domain is open, then all level sets of the mgf are rotund. ◊

Example 9.8. The analytic expression $(y^2 - b^2)((x - e)^2 - x^2y^2) = a^2(y^2 - b^2)$ with $a > b + e$ for a two-dimensional egg was presented in the popular mathematics journal *Pythagoras*, see Mulder [1977]. ◊



The solution with $a = 0.4, b = 0.3, c = 0.4$;



$a = 0.4, b = 0.8, c = 0.1$.

9.3 Initial transformations. The concept of a rotund set does not depend on the coordinates. It may be defined on any vector space. It is only when we want to analyse the asymptotic behaviour of the function $e^{-\psi \circ n}$ that we need coordinates, and then we choose the coordinates to suit us. We shall say that a rotund set D in $\mathbb{R}^d$ is in *correct initial position* if the gauge function $n = n_D$ satisfies

$$n(u, 1) - 1 \sim u^T u/2, \quad u \to 0 \in \mathbb{R}^h. \tag{9.5}$$

A rotund set in correct initial position, rotated around the vertical axis remains in correct initial position.

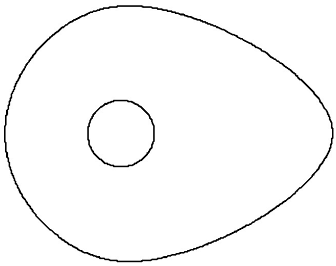
Given a point $p \in \partial D$, there is a linear map which brings D in correct initial position, such that p corresponds to the point $(0, 1)$ on the vertical axis.

To see this, assume D is a subset of $\mathbb{R}^{h+1}$. We shall perform a number of simple linear transformations on D to bring it in correct initial position. Let H be the halfspace *supporting* D at the point p:

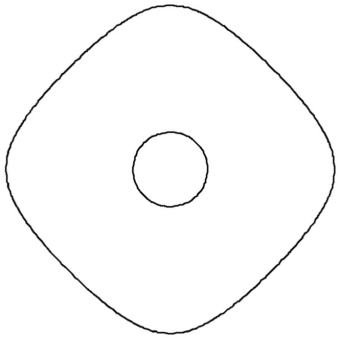
$$H \cap \partial D = \{p\}.$$

First, rotate D in the two-dimensional subspace containing p and the vertical axis, so that $R(H)$ is a horizontal halfspace of the form $\{v \geq t\}$. So $(a, t) = R(p)$ is the point on the boundary of $R(D)$ which maximizes the vertical coordinate. A vertical multiplication $V\colon (u, v) \mapsto (u, v/t)$ maps $R(p)$ into $(a, 1)$. A horizontal *shear* $S : (u, v) = (u - av, v)$ maps this point into the point $(0, 1)$ on the vertical axis. As a result of these transformations H has been mapped onto $\{v \geq 1\}$. The gauge function m of the rotund set $HVR(D)$ has the form

$$m(u, 1) - 1 \sim Q(u)/2, \quad u \to 0 \in \mathbb{R}^h,$$



An analytic egg.



Level sets of the mgf of a symmetric distribution.

where $Q/2$ is the quadratic term in the second-order Taylor expansion of $u \mapsto m(u, 1)$ around the point $u = 0$. From linear algebra it is known that in appropriate orthogonal coordinates $Q(u) = \lambda_1 u_1^2 + \cdots + \lambda_h u_h^2$, with $\lambda_i > 0$ since Q is positive definite. A diagonal transformation in these coordinates will bring D into correct initial position.

So we may write $D = A(D_0)$ where D_0 is in correct initial position, and A is a linear map which sends $(0, 1)$ into p. The map $A = A_p$ is called an *initial transformation*. In the recipe above, all operations are well defined and depend continuously on p, as long as the halfspace supporting D in p is not downward horizontal of the form $\{v \leq s\}$.

Let $\mathcal{J}_p$ denote the set of initial transformations for given $p \in \partial D$. So $A \in \mathcal{J}_p$ if $A^{-1}(D)$ is in correct initial position and $A(0, 1) = p$. The gauge function n_A of $A^{-1}(D)$ satisfies $n_A(u, 1) - 1 \sim u^T u/2$ for $u \to 0$. It has the simple form $n_A = n \circ A$, since

$$n_A(w) < 1 \iff w \in A^{-1}(D) \iff Aw \in D \iff n(Aw) < 1.$$

If D is in correct initial position, and $A^{-1}(D)$ too, then A maps $(0, 1)$ into itself, and also the supporting halfspace $\{v \geq 1\}$. So A is a linear transformation of the horizontal coordinate u. Hence $n_A(u) - 1 \sim (Au)^T (Au)/2$. It follows that A is an orthogonal transformation in $\mathbb{R}^h$. We conclude

$$\mathcal{J}_p = \{AR \mid R \in \mathrm{O}(h)\},$$

where $\mathrm{O}(h)$ is the group of orthogonal linear transformations in $\mathbb{R}^h$, and A is an arbitrary element of $\mathcal{J}_p$. Now define the set of *initial transformations for D* by

$$\mathcal{J} = \bigcup_p \mathcal{J}_p.$$

There is an obvious projection $p\colon \mathcal{J} \to \partial D$ which maps $A \in \mathcal{J}_p$ into $p \in \partial D$. As we have seen above, locally it is possible to define a function $p \mapsto A_p$ from Uß∂D into $\mathcal{J}$ so that $A_p \in \mathcal{J}_p$ for each $p \in U$. Such a map is called a *local section.* It establishes a homeomorphism between $U \times \mathrm{O}(h)$ and $p^{-1}(U)$. This shows that

Theorem 9.9. *The space of initial transformations is a compact fiber bundle.*

See Walschap [2004] for basic information on fiber bundles. For certain dimensions the fiber bundle is non-trivial: there is no global section, the space $\mathcal{J}$ is not homeomorphic to the product space $\partial D \times \mathrm{O}(h)$. For $d = 3$ the space $\mathcal{J}$ of initial maps is like a *Moebiusb@Möbius band.* Locally the band looks like the product of a circle S with the interval $I = [-1, 1]$, but it is not possible to construct a homeomorphism between the Möbius band and the product space $S \times I$. Such topological impossibility results are hard to prove and may be very frustrating. It is known that on the sphere in $\mathbb{R}^3$ there is no vector field without singularities; see Poincaré [1881]. (It is not possible to comb a tennis ball, without creating a crown in some point.) The result also holds in dimension $d > 3$ with d odd; see Brouwer [1911]. This topological result holds for the surface of rotund sets, and has the consequence that it is not possible to choose a coordinate system on the tangent space in every point $p \in \partial D$ which varies continuously with p. Actually such a choice is only possible for $d = 2, 4, 8$ when the tangent bundle to the sphere is parallelizable; see Walschap [2004]. This in turn implies the non-triviality of the fiber bundle $\mathcal{J}$. As a consequence there is a rather unpleasant result about the normalization of *high risk scenarios.*

Theorem 9.10. *In Theorem 9.1, for $d \geq 3$, odd, it is not possible to choose the affine transformations α_H to depend continuously on the halfspace H, even for halfspaces H which do not contain the origin.*

Proof. See Balkema & Embrechts [2004], Theorem 5.6, for details. □

It would be interesting to see to what extent this result restricts the class of possible limit laws. It is not possible to draw an equilateral spherical triangle with side length 100 km around each point on the surface of the earth in a continuous way, even if the surface were a perfect sphere. This means that for $d = 3$ a high risk limit vector (U, V), where U has density $\propto e^{-n_D(u)}$ for an open triangle D, is not possible. For the unit ball B in $\mathbb{R}^3$ one may construct the tangent bundle $T_{\partial B}$ of the two-sphere ∂B by glueing together the sets $\mathbb{C} \times D_-$ and $\mathbb{C} \times D_+$ along the boundaries $\mathbb{C} \times \partial D_\pm$, where $D_\pm$ are copies of the closed disk $\{|z| \leq 1\}$ in $\mathbb{C}$. The glueing condition implies that the two parts $f_\pm$ of a continuous function f on $T_{\partial B}$ should satisfy

$$f_-(z, e^{i\theta}) = f_+(e^{2i\theta} z, e^{i\theta}).$$

On the other hand, for the unit ball B in $\mathbb{R}^4$ the 6-dimensional tangent bundle $T_{\partial B}$ is trivial. It is homeomorphic to the product $\mathbb{R}^3 \times \partial B$, and hence a density $\propto e^{-n_D(u)}$

for a tetrahedron D cannot be excluded a priori. For the unit ball in dimensions 8 and 16 the tangent bundles to the sphere ∂B are also trivial.

The authors would like to thank Dietmar Salamon for an illuminating discussion on this subject. We can only point out here that in certain dimensions it is a waste of time to seek normalizations α_H which depend continuously on H for $H \to \partial$. We leave it to others to clear up the mystery surrounding the question: What distributions on H_+ are taboo as limits?

9.4 Convergence of the quotients. We now turn to the proof of Theorem 9.1. We shall only prove pointwise convergence at this point. Let $f = e^{-\psi \circ n}/C$, where ψ satisfies the conditions of the theorem, and n is the gauge function of a bounded open convex set D which contains the origin. For the moment we make no assumptions about the boundary of D.

Recall that ψ' is positive on $[0, t_\infty)$, ψ tends to infinity in $t_\infty \le \infty$, and $(1/\psi')'(t)$ vanishes in t_∞. Set

$$a = a(t) = 1/\psi'(t), \quad b = b(t) = \sqrt{ta(t)}, \tag{9.6}$$

and define the affine transformations α_t for $t \in (0, t_\infty)$ by

$$\alpha_t(u, v) = (bu, t + av) = (x, y), \quad (u, v) \in \mathbb{R}^{h+1}.$$

We shall give a condition on the gauge function n which implies

$$h_t(w) = \frac{f(\alpha_t(w))}{f(\alpha_t(0))} \to e^{-u^T u/2 - v}, \quad t \to t_\infty$$

uniformly on bounded w-sets in $\mathbb{R}^d$.

Proposition 9.11. *Suppose $n(x, 1) - 1 \sim x^T x/2$ for $x \to 0$. Then the function $\varphi = \psi \circ n$ satisfies*

$$\varphi(bu, t + av) - \varphi(0, t) \to u^T u/2 + v, \quad t \to t_\infty$$

uniformly on bounded (u, v)-sets in $\mathbb{R}^{h+1}$.

Proof. Apply the mean value theorem to ψ:

$$\psi(n(bu, t + av)) - \psi(n(0, t)) = \frac{\psi'(t^*)}{\psi'(t)} \frac{n(bu, t + av) - t}{a},$$

with t^* a real number between $t = n(0, t)$ and $n(bu, t + av)$. The factors $\psi'(t)$ and $a = 1/\psi'(t)$ cancel. We shall see later that $\psi'(t^*)/\psi'(t) \to 1$. First consider the second fraction. Since $(t + av - t)/a = v$, it suffices to show that

$$\frac{n(bu, t + av) - (t + av)}{a} = \frac{t + av}{a}\left(n\left(\frac{bu}{t + av}, 1\right) - 1\right) \to u^T u/2.$$

Here we shall use that

$$\chi(s,u) = (n(su,1)-1)/s^2 \to u^T u/2, \quad s \to 0,$$

where we set $\chi(0,u) = u^T u/2$. Take $s = b/(t+av)$, and recall that $a(t) = o(t)$ since $a'(t) \to 0$. (If t_∞ is finite, then $a(t)$ vanishes in t_∞ since $\psi(t_\infty) = \infty$.) The same limit relations hold for $b(t) = \sqrt{ta(t)}$. So we see that $s = b/(t+av) \to 0$ as $t \to t_\infty$, and

$$\frac{t+av}{a} s^2 \frac{n(su,1)-1}{s^2} = \frac{b^2}{(t+av)a}\chi(s,u) \to u^T u/2, \quad t \to t_\infty.$$

The limit holds uniformly on bounded u-sets in $\mathbb{R}^h$.

Finally, observe that the difference $t^* - t$ is bounded by $\delta = n(bu, t+av) - t$, and that δ/a converges to $u^T u/2 + v$. Theorem 6.1 tells us that $t^* - t = O(a)$, since $\psi'(t_n + s_n a(t_n))/\psi'(t_n) \to 1$ for $t_n \to t_\infty$, and s_n is bounded. □

The same argument gives the limit function $e^{-(m^p(u)+v)}$ under the assumption that $n(x,1) - 1 \sim m^p(x)$ for $x \to 0$, where m^p is the pth power, $p \geq 1$, of the gauge function m of a bounded open convex set in $\mathbb{R}^h$ which contains the origin. Such results yield limit laws for exceedances over horizontal thresholds, see Section 15.2.

Proposition 9.12. *Let p_k be points on ∂D which converge to a point p_0. Suppose n_D is C^2 on a neighbourhood of p_0, and $n_D^*(p_0)$ in (9.4) is positive definite. Let $t_k \to t_\infty$, and let H_k be the halfspace supporting $t_k D$ in $t_k p_k$. There exist affine transformations α_k mapping H_+ onto H_k such that*

$$h_k(w) = \frac{f(\alpha_k(w))}{f(p_k)} \to e^{-(u^T u/2+v)}, \quad k \to \infty, w = (u,v)$$

uniformly on bounded w-sets.

Proof. Choose coordinates such that $p_0 = (0,1)$, and D is in correct initial position. Choose $A_k \in \mathcal{J}_{p_k}$ such that $A_k \to \mathrm{id}$. Set $\varphi_k = \varphi \circ A_k$, where $\varphi = \psi \circ n_D$, and set $\beta_k(u,v) = (b_k u, t_n + a_k v)$ with $a_k = 1/\psi'(t_k)$ and $b_k = \sqrt{a_k t_k}$ as above. Write $\varphi_k \circ \beta_k = \varphi \circ \alpha_k$. Then $\alpha_k = A_k \circ \beta_k$. We claim that $u_k \to u$ and $v_k \to v$ imply

$$\varphi_k(b_k u_k, t_k + a_k v_k) - \varphi(0, t_k) \to u^T u/2 + v.$$

The computation is as above. The lemma below ensures uniformity in k. □

Note the normalization. Let H support tD in the point p_H. Then

$$\begin{aligned} \alpha_H = A_p \circ \alpha_t, \quad & \alpha_t(u,v) = (b_t u, t + a_t v), \tau = n(p_H), \; p = p_H/t, \\ & A_p \in \mathcal{J}_p, \; a_t = 1/\psi'(t), \; b_t = \sqrt{t a_t}. \end{aligned} \tag{9.7}$$

Lemma 9.13. *Let* $\chi_k(s,u) = (n_{A_k}(su,1)-1)/s^2$ *where* $A_k \to I$. *Then for* $s_k \to 0$, $u_k \to u_0$,

$$\chi_k(s_k,u_k) \to u_0^T u_0/2, \quad k \to \infty.$$

Proof. We write χ_k as an average of a second derivative of n_{A_k}. For fixed $u \neq 0$, integration along the ray through u gives

$$\begin{aligned}\chi_k(s,u) &= (n_{A_k}(su) - 1)/s^2 \\ &= \frac{1}{s^2}\int_0^s \int_0^q \frac{\partial^2}{\partial x \partial x} n_{A_k}(tu,1)(u,u)dt dq \\ &= \int_0^1 \int_0^r \frac{\partial^2}{\partial x \partial x} n_{A_k}(stu,1)(u,u)dt dr.\end{aligned}$$

Recall that $n_A = n \circ A$. Hence $n''_A(z)(w,w) = n''(Az)(Aw,Aw)$, and convergence $n''_{A_k} \to n''$ is uniform on some ε-ball around $(0,1)$. In particular, for $s_k \to 0$

$$\frac{\partial^2}{\partial x \partial x} n_{A_k}(s_k t u_k,1)(u_k,u_k) \to n_{xx}(0,1)(u_0,u_0) = u_0^T u_0$$

uniformly in $t \in [0,1]$. The second integral with s_k and u_k tends to $u_0^T u_0/2$. □

The next example is a warning that good asymptotic results for horizontal halfspaces do not guarantee good behaviour for high risk scenarios on halfspaces H_n which are asymptotically horizontal, even when

$$f(\alpha_{H_n}(w))/f(\alpha_{H_n}(0)) \to e^{-(u^T u/2+v)}, \quad w = (u,v) \in \mathbb{R}^{h+1} \tag{9.8}$$

uniformly on bounded sets.

Example 9.14. The vector (X,Y) in $\mathbb{R}^2$ has a standard Gaussian density on $\{y \geq -99\}$. The conditional distribution of X given y for $y \leq -100$ is Student with $f = 99$ degrees of freedom. The variable Y is standard normal, and for $y = \theta - 100$ we shall take the conditional distribution to be a mixture of standard normal, with weight θ, and Student with weight $1-\theta$. Note that for any point (x,y) the density increases as one moves to the point $(0,y)$ on the vertical axis, and increases further as one moves to the origin. Also along the interval $[-100,-99]$. So the density is *unimodal* in the sense that there is only one maximum, but not in our terminology!

The high risk scenarios Z^{H^t} for the horizontal halfplanes $H^t = \{y \geq t\}$ converge to the Gauss-exponential law with the normalization $\alpha_t\colon (u,v) \mapsto (u, t + v/t)$. Let H_n be halfplanes with slope $-\varepsilon_n$ through the point $(0,n)$, with $\varepsilon_n \to 0$. So $H_n = \{y \geq n - \varepsilon_n x\}$ is asymptotically horizontal, and (9.8) holds uniformly on bounded w-sets if we choose the affine transformations α_H as if Z has a standard Gauss distribution. If ε_n is small, the high risk scenarios Z^{H_n} will be asymptotically

Gauss-exponential, since the bivariate Gaussian distribution lies in the domain of the Gauss-exponential limit law. If $\varepsilon_n \to 0$ too slowly, then the halfspace H_n may pick up mass from the Student distribution on the halfspace $\{y \le -99\}$.

How fast should ε_n tend to zero? Let p_n be the mass in $H_n \cap \{y \ge -99\}$, and q_n the mass in $H_n \cap \{y \le -99\}$. The high risk scenarios Z^{H_n} will be asymptotically Gauss-exponential if and only if $q_n = o(p_n)$. A simple calculation shows that $q_n/p_n \to \infty$ for $\varepsilon_n = 1/n$. Also for $\varepsilon_n = 1/n^n$. ◊

Such behaviour can not occur if Z has a *unimodal* density.

9.5 Global behaviour of the sample cloud. The Gauss-exponential point process describes the texture of a large sample cloud from a rotund-exponential density locally at the edge. What does the whole sample cloud look like? Properly scaled, it looks like the set D; see the corollary below. The asymptotics of the number of vertices of the convex hull is treated in Hueter [1999]. Let Z have density $e^{-\psi(n)}$.

One may simulate observations Z by first simulating $T = n_D(Z)$ from the density

$$f_0(t) = t^{d-1} d|D| e^{-\psi(t)}, \quad 0 \le t < y_\infty, \tag{9.9}$$

and then independently choosing $\zeta = Z/T \in \partial D$ uniformly on ∂D by

$$\mathbb{P}\{\zeta \in C\} = |C \cap D|/|D|, \quad C \text{ an open cone.} \tag{9.10}$$

The set D need not be rotund here, but is assumed convex.

The expression (9.9) for the density of T shows that $T \in \mathcal{D}^+(0)$, since $e^{-\psi}$ satisfies the von Mises condition and t (and hence also t^{d-1}) is flat for $e^{-\psi}$. Use $1/\psi'(t) = o(t)$ for $t \to y_\infty$. It follows that

$$\mathbb{P}\{T \ge t\} \sim t^{d-1} d|D| e^{-\psi(t)}/\psi'(t), \quad t \to y_\infty.$$

Choose t_n such that $\mathbb{P}\{T \ge t_n\} \sim 1/n$, and set $a_n = 1/\psi'(t_n)$. Arrange the sample $Z_1, \dots, Z_n$ in decreasing order of the gauge function:

$$T_{n1} > \dots > T_{nn}, \quad T_{ni} = n_D(Z_{ni}),\ \zeta_{ni} = Z_{ni}/T_{ni}.$$

The scaled variables $(T_{ni} - t_n)/a_n$ may be approximated by the points of the limiting Poisson point process on $\mathbb{R}$ with intensity e^{-t}, and the pairs $((T_{ni} - t_n)/a_n, \zeta_{ni})$ by the points of the corresponding *marked Poisson point process*, arranged in decreasing order by the first coordinate.

Proposition 9.15. *Let Z have density $e^{-\psi \circ n_D}$, where ψ satisfies (6.4) and D is a bounded open convex set which contains the origin. We do not assume that D is rotund. Let $Z_1, Z_2, \dots$ be independent observations of Z. Set*

$$T_i = n_D(Z_i), \quad \zeta_i = Z_i/T_i, \quad i = 1, 2, \dots$$

Define t_n and a_n as above. The point process N_n with points $((T_i - t_n)/a_n, \zeta_i)$ converges:

$$N_n \Rightarrow N_0 \ \text{weakly on } [c, \infty) \times \partial D, \quad c \in \mathbb{R}.$$

Here N_0 is the Poisson point process on $\mathbb{R} \times \partial D$ with mean measure $e^{-t}dt \times d\gamma$, where $d\gamma$ is the distribution of ζ; see (9.10).

Proof. The mean measures converge. See Section 6.4. □

Corollary 9.16. *Let M_n be the scaled sample cloud*

$$M_n = \{Z_1/t_n, \dots, Z_n/t_n\}.$$

For any non-empty open cone C and any $\varepsilon \in (0, 1)$,

$$M_n(C \setminus (1+\varepsilon)D) \xrightarrow{\mathbb{P}} 0$$
$$M_n(C \setminus (1-\varepsilon)D) \xrightarrow{\mathbb{P}} \infty.$$

Proof. For any $c \in \mathbb{R}$

$$M_n(C \setminus (1 + ca_n/t_n)D) \Rightarrow N,$$

where N is a Poisson variable with expectation $e^{-c}|C \cap D|/|D|$. Now use $a_n/t_n \to 0$. □

10 The Gauss-exponential domain, unimodal distributions

10.1 Unimodality. Unimodality will play a crucial role in proving $\mathbf{L}^1$-convergence. Recall that a function $f \geq 0$ is unimodal if the level sets $\{f > c\}$ are convex for all $c > 0$. There exist many other definitions, see Dharmadhikari & Joag-Dev [1988]. With the definition above convergence of unimodal functions has some of the features of weak convergence of increasing functions. If the functions $f_0, f_1, \ldots \geq 0$ are unimodal, f_0 is continuous, and if $f_n \to f_0$ on a dense set, then convergence is uniform on compact sets provided the maxima converge.

Proposition 10.1 (Weak Convergence). *Let $f_0, f_1, \ldots \geq 0$ be unimodal functions on an open convex set Dß$\mathbb{R}^d$. Suppose*

1) *$f_n \to f_0$ pointwise on a dense set AßD;*
2) *the level sets $\{f_0 > 0\}$ are open;*
3) *if $\sup f_0 < c < \infty$, then $\sup f_n < c$ eventually.*

Then $f_n(z_n) \to f_0(z_0)$ if $z_n \to z_0$, and if f_0 is continuous at z_0.

Proof. If $f_0(z_0) > c$, then there are points $a_0, \dots, a_d$ in $A \cap \{f_0 > c\}$ such that z_0 lies in the interior U of their convex hull. By unimodality, eventually $f_n > c$ holds on U. If $f_0(z_0) = \sup f_0$ we are done. Otherwise let $b \in (f_0(z_0), \sup f_0)$, and let V be an open simplex with vertices in $A \cap \{f_0 > b\}$. As above, eventually $f_n > b$ on V. Let V_n be the interior of the convex hull of V and z_n. Each $a \in A \cap V_0$ eventually lies in V_n. If $f_{k_n}(z_{k_n}) > b$ infinitely often, then $f_{k_n}(a) > b$ eventually, and hence $f_0(a) \geq b$ for all $a \in A \cap V_0$, contradicting the continuity in z_0. □

We introduce a class of well-behaved unimodal functions which contains the class $\mathcal{RE}$.

Definition. U is the set of all continuous non-negative functions f on $\mathbb{R}^d$ which satisfy:

1) $f(z) < f(0)$ for $z \neq 0$;

2) for $0 < c < f(0)$ the level sets $\{f > c\}$ are bounded and convex;

3) for each $c \in (0, f(0))$ and each boundary point p of $\{f > c\}$, there is a unique *supporting* halfspace $H = H_p$ to $\{f > c\}$ in p, and

$$H_p \cap \{f \geq c\} = \{p\}. \tag{10.1}$$

Condition 1) is not essential. It makes the mode unique and locates it at the origin. Conditions 1) and 3) imply that there are no plateaus: $\{f = c\} = \partial\{f > c\}$ for $0 < c < f(0)$. Condition 3) ensures that the level sets are strictly convex, and their boundary is C^1. The functions $f \in$ U vanish at infinity, but need not be integrable.

For $f \in$ U, there is a one-one correspondence between points $p \in O = \{f > 0\}$, $p \neq 0$, and halfspaces H which intersect O and do not contain the origin, see (10.1). The maps

$$p \mapsto H_p, \quad H \mapsto p_H \tag{10.2}$$

are continuous.

The correspondence between points and halfspaces allows us to write α_p for α_{H_p}, and $p \to \partial_O$ for $H \to \partial$, see (8.7). Define U$_0$ßU as the set of functions $f = e^{-\varphi}$ for which there exist $\alpha_p \in \mathcal{A}$ mapping H_+ onto H_p such that

$$\chi_p = \varphi \circ \alpha_p - \varphi(p) \to \chi, \quad p \to \partial_O, \quad \chi(u, v) = u^T u/2 + v \tag{10.3}$$

uniformly on bounded sets.

Proposition 10.2. *The set* U$_0$ *is closed for positive powers.*

Proof. Suppose $\varphi(\alpha_p(u, v)) - \varphi(p) \to u^T u/2 + v$ and assume further that $c > 0$. Set $\beta_p(u, v) = \alpha_p(u/\sqrt{c}, v/c)$. Then β_p maps H_+ onto the halfspace H_p and $c(\varphi(\beta_p(u, v)) - \varphi(p)) \to c(u^T u/2c + v/c)$. □

Lemma 10.3. *One may choose the normalizations α_p in (10.3) so that $\alpha_p(0,0) = p$, $p \in O \setminus \{0\}$.*

Proof. If $\chi_n = \varphi_n \circ \alpha_n \to \chi$ uniformly on bounded sets for a continuous limit function or $\alpha_n^{-1}(Z_n) \Rightarrow W$, then these limit relations also hold for any sequence $\beta_n \sim \alpha_n$ in the sense that $\alpha_n^{-1}\beta_n \to \text{id}$. First note that $\alpha_p^{-1}(p)$ is boundary point of H_+, and tends to $(0,0)$, since $\chi(\alpha_p^{-1}(p)) \to 0$ by (10.3). So there is a horizontal translation τ_p mapping $(0,0)$ into $\alpha_p^{-1}(p)$. Hence $\beta_p(0,0) = p$ for $\beta_p = \alpha_p \circ \tau_p \sim \alpha_p$, and $\beta_p(H_+) = H_p$. □

10.2* Caps. Let $f \in \mathrm{U}_0$. Write $f = e^{-\varphi}$ on $O = \{f > 0\}$. The function φ has the same level sets as f and is easier to work with. It describes a valley. The minimum is achieved at the origin, and $\varphi(z) \to \infty$ for $z \to \partial_O$. The sets $\{\varphi < c\}$ are open bounded convex sets. The condition $f \in \mathrm{U}_0$ means that (10.3) holds uniformly on bounded w-sets. In particular, successive level sets $\varphi(\alpha_p(w)) = \varphi(p) + m$ look like successive parabolas $v = m - u^T u/2$. In the sequence of parallel parabolas

$$v = m - u^T u/2, \quad m = 0, 1, 2, \dots$$

we may distinguish a tower of parabolic caps

$$\tau^m Q = \{m \le v < m + 1 - u^T u/2\}, \quad \tau(u, v) = (u, v + 1),$$
$$Q = \{0 \le v < 1 - u^T u/2\},$$

balancing one on top of the other at the points $(0, m)$. For the function φ the relation (10.3) gives a similar tower. See the figures below.

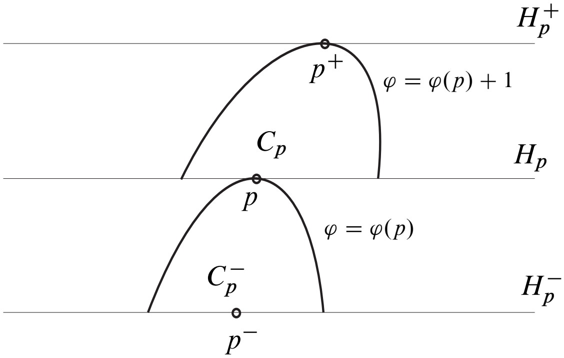
The parabolic caps $\tau^m Q$ all have the same size. This need not hold for the tower of caps for φ. If these caps grow too fast, the integral of $e^{-\chi_p}$ may fail to converge over H_+. So we want to prove that any cap for φ is not much larger than the one below. This will enable us to bound the tails of the integral.

In our arguments we shall only use the fact that $f \in \mathrm{U}_0$. In Section 6 we saw that any df in the domain of the exponential law for exceedances was tail asymptotic to a df with a density $e^{-\psi}$ satisfying the von Mises condition. We surmise that, similarly, distributions in the domain of attraction of the Gauss-exponential high risk limit law are close to a df with density in U_0.

So let us now describe these caps of φ. See the figure below. For convenience, think of the halfspace H_p supporting the convex set $\{\varphi < \varphi(p)\}$ as horizontal. Define the cap

$$C_p = H_p \cap \{\varphi < \varphi(p) + 1\}.$$

Let H_p^+ßH_p support $\{\varphi < \varphi(p) + 1\}$ in p^+. Then $H_p^+ = H_{p^+}$. Similarly, let $H_p^- \supset H_p$ support $\{\varphi < \varphi(p) - 1\}$ in p^-. Then $H_p^- = H_{p^-}$. We call p the



The cap C_p^- dangling from the base point p of the cap C_p.

base point of C_p, and p^+ the *top point*. By using horizontal halfspaces supporting $\{\varphi < \varphi(p) + m\}$, one may construct a tower of such caps balancing on each other, with the top point of one cap being the base point of the next.

The limit relation (10.3) implies

$$\alpha_p^{-1}(C_p) \to Q = \{0 \le v < 1 - u^T u/2\}.$$

It is convenient to choose α_p such that

$$\alpha_p(0,0) = p, \quad \alpha_p(0,1) = p^+, \quad \alpha_p(H_+) = H_p. \tag{10.4}$$

This is possible by the arguments of the proof of Lemma 10.3.

Let τ be the translation $(u, v) \mapsto (u, v+1)$. Then

$$\alpha_p^{-1}(C_p^-) \to \tau^{-1}(Q) = \{-1 \le v < -u^T u/2\}$$
$$\alpha_{p^-}^{-1}(C_p^-) \to Q,$$

which gives

$$\alpha_p^{-1}\alpha_{p^-}(Q) \to \tau^{-1}(Q).$$

This does not imply $\alpha_p^{-1}\alpha_{p^-} \to \tau^{-1}$, since $\tau^{-1}(Q) = \tau^{-1}(R(Q))$ for any horizontal rotation R, but we do have

Proposition 10.4. *The maps $\alpha_p^{-1} \circ \alpha_{p^-} \circ \tau$ are linear and*

$$\alpha_p^{-1} \circ \alpha_{p^-} \circ \tau \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} A & b \\ 0 & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix},$$

where $a = a_p \to 1$, $b = b_p \to 0$, and $R^{-1}A \to I$ for a suitable choice of $R = R_p \in \mathrm{O}(h)$.

Proof. Check that $(0,0) \mapsto (0,1) \mapsto p \mapsto (0,0)$ by (10.4). The zero in the matrix is due to the conservation of the class of horizontal halfspaces. □

Corollary 10.5. *The maps $\beta_p = \alpha_{p^-}^{-1}\alpha_p$ have the form $w \mapsto Tw + q$; the corresponding transformation $\tilde{\beta}_p$ on the vertical coordinate is $v \mapsto v/a + 1$. For any $\varepsilon > 0$, there exists $t_0 > 1$ such that for $\varphi(p) > t_0$,*

$$\|T_p\| < e^{\varepsilon} \quad \|q_p\| < 2, \quad e^{-\varepsilon} < a_p < e^{\varepsilon}. \tag{10.5}$$

Proof. $\beta_p(u,v) = (A^{-1}u + b'v, v/a + 1)$ with $b' = -A^{-1}b/a$. □

The bounds (10.5) will enable us to prove Theorem 9.1. For Theorem 9.2 we need more.

So far, we have not used coordinates. It is convenient to introduce coordinates (u,v) such that p is the origin, and H_p the upper halfspace. Choose coordinates so that α_p is the identity. Now take a fixed halfspace $J = \{v \geq \xi^T u + \xi_0\}$ with slope $\xi \in \mathbb{R}^h$ and $\xi_0 \in \mathbb{R}$, and let $C_J = C_{p(J)}$ be the cap cut off by this halfspace, where we write $p(J) = p_J$; see (10.2). Since $\varphi \circ \alpha_p - \varphi(p)$ is close to $\chi = u^T u/2 + v$, the cap C_J will be close to the parabolic cap

$$\gamma_J^{-1}(Q) = J \cap \{v < 1 + \xi_0 - u^T u/2 - \xi^T \xi/2\}.$$

We again have two limit relations for the cap:

$$\alpha_p^{-1}(C_J) \to \gamma_J^{-1}(Q), \quad \alpha_{p(J)}^{-1}(C_J) \to Q, \quad p \to \partial_O.$$

As above, we obtain $\alpha_p^{-1}\alpha_{p(J)}(Q) \to \gamma_J^{-1}(Q)$, and again, because of the symmetry of Q, we cannot conclude that $\gamma_p = \alpha_p^{-1}\alpha_{p(J)}$ converges to γ_J^{-1}. However, the family (γ_p^{-1}) is relatively compact, and all limit points for $p \to \partial_O$ are of the form $\gamma_J^{-1} \circ R$ with $R \in \mathrm{O}(h)$. This suffices to obtain bounds on the integral of the tail of $\|w\|^m f(\alpha_p(w))/f(p)$ on the halfspace J. By the Transformation Theorem,

$$\int_{J \setminus E_r} \|w\|^m \frac{f(\alpha_p(w))}{f(p)} dw c(p) = \int_{H_+ \setminus rB} \|\gamma_p^{-1}(z)\|^m \frac{f(\alpha_{p(J)}(z))}{f(p(J))} dz,$$

where $c(p) = f(p(J))/f(p)/|\det \gamma_p|$, and E_r is the ellipsoid $\gamma_p(rB)$. Relative compactness of γ_p, $p \to \partial_O$, implies that $c(p)$ is bounded by a constant C. Hence we have the following inequality:

Proposition 10.6. *For any halfspace $J = \{v \geq \xi^T u + \xi_0\}$ there exist t_0, $C_0 > 1$, and $\delta > 0$, such that*

$$\int_{J \setminus RB} \|w\|^m e^{-\chi_p}(w) dw \leq C_0 \int_{H_+ \setminus rB} \|w\|^m e^{-\chi_{p(J)}}(w) dw,$$

$$\varphi(p) > t_0, \quad r = \delta R > 1.$$

10.3* $\mathbf{L^1}$-convergence of densities. Assume $f = e^{-\varphi} \in \mathrm{U}_0$. So $\varphi\circ\alpha_p - \varphi(p) \to \chi$ for $p \to \partial_O$ where $\chi(u,v) = u^T u/2 + v$. We have to bound the tails of certain integrals to obtain the desired limit laws for high risk scenarios and sample clouds. A rough bound will do.

Lemma 10.7. *Let $A_1 ß A_2 ß \cdots$ be bounded Borel sets which cover H_+. Let $g \geq 0$ be bounded by a_n on A_n, and let $h \geq 0$ be bounded by c_n on the complement of A_n. Then*

$$\int_{A_m^c} g(z)h(z)dz \leq \sum_{n>m} a_n |A_n| c_{n-1}.$$

Proof. This is obvious if we replace A_n by $A_n \setminus A_{n-1}$ on the right. □

We shall apply this lemma with $c_n = e^{-n}$, $g(z) = \|z\|^m$, and the hat boxes

$$A_n = \{\|u\| \leq (Cne^{n\delta})^2, 0 \leq v \leq (Cne^{n\delta})^2\}. \tag{10.6}$$

Then $a_n = (C_1 n e^{n\delta})^{2m}$. If δ is small, $\delta(m+d) \leq 1/4$, there exists a constant C_2 such that $a_n |A_n| c_n < C_2/e^{n/3}$ for all $n \geq 1$.

We need an extension of the bounds in (10.5) to products.

Lemma 10.8. *Suppose $\beta_n(w) = T_n w + b_n$ for $n = 1, 2, \ldots$ Then*

$$(\beta_1 \ldots \beta_n)(w) = b_1 + T_1 b_2 + \cdots + T_1 \ldots T_{n-1} b_n + T_1 \ldots T_n w = p_n + S_n w.$$

If $\|T_n\| < e^{\varepsilon}$ and $\|b_n\| < 2$ for $n \geq 1$, then

$$\|S_n\| < e^{n\varepsilon}, \quad \|p_n - p_{n-1}\| < 2e^{n\varepsilon}, \quad \|p_n\| < 2ne^{n\varepsilon}.$$

Let $\tilde{\beta}_n(v) = c_n v + 1$ be the corresponding transformation of the vertical coordinate, and $\tilde{p}_n$ the vertical coordinate of p_n. If $e^{-\varepsilon} < c_n < e^{\varepsilon}$, then $\tilde{a}_n := \tilde{p}_n - \tilde{p}_{n-1} > 0$, and

$$\tilde{a}_{n+1}/\tilde{a}_n = c_n, \quad \tilde{p}_n/\tilde{a}_n < ne^{n\varepsilon}.$$

Proof. By induction. □

We can now formulate and prove the main result: Random vectors with a density in U_0 lie in the domain the Gauss-exponential high risk limit law.

Theorem 10.9. *Let $f \in \mathrm{U}$. Set $O = \{f > 0\}$. Suppose there exist affine transformations α_p mapping H_+ onto H_p such that for $w = (u,v) \in \mathbb{R}^{h+1}$*

$$h_p(w) := \frac{f(\alpha_p(w))}{f(p)} \to e^{-(u^T u/2+v)}, \quad p \to \partial_O.$$

Then f is integrable, and $\int \|z\|^m f(z)dz$ is finite for each $m \geq 1$. Moreover

$$\|w\|^m h_p(w) \to \|w\|^m e^{-(u^T u/2+v)}, \quad p \to \partial_O$$

holds in $\mathbf{L}^1(J)$ for any halfspace $J = \{v \geq \xi^T u + \xi_0\}$.

Proof. Pointwise convergence yields uniform convergence on bounded sets by Proposition 10.1. By Proposition 10.6, we have to show that for any sequence $p_n \to \partial_O$, any $\varepsilon > 0$ and $m \geq 1$, there exists $r > 1$ so large that

$$\int_{H_+ \setminus rB} \|w\|^m h_{p_n}(w)dw < \varepsilon, \quad n \geq n_0, \tag{10.7}$$

where B is the open unit ball. The relation will then hold for p eventually, i.e. for $p \in O$ outside a set $\{f \geq \delta_0\}$.

Cover $\{f < \delta_0/2\}$ by a finite number of halfspaces disjoint from $\{f \geq \delta_0\}$ to show that f is integrable and $\int \|z\|^m f(z)dz$ finite.

To establish (10.7), we use the unimodality of the functions h_p. We shall show that for any $\delta > 0$ there exist $\varepsilon_0 > 0$ and $C > 1$ such that

$$\{h_p > e^{-n}\} \cap H_+ ß A_n, \quad n \geq 1,\ 0 < f(p) < \varepsilon_0,$$

where the sets $A_n = A_n(\delta, C)$ are the hat boxes in (10.6). By the discussion after Lemma 10.7 this yields (10.7).

The construction of the hat boxes is our main task.

Fix a point p far out in O. Write $f = e^{-\varphi}$. Let $H = H_p$ be the halfspace supporting $\{\varphi < \varphi(p)\}$ in p. Choose coordinates (u, v) so that p is the origin, H the upper halfspace above, and $\alpha_p = \text{id}$. So the sets $\{\varphi < \varphi(p) + k\}$ locally look like an increasing sequence of paraboloids, $v < k - u^T u/2$. Let H_k be the horizontal halfspace supported by the convex set $\{\varphi < \varphi(p) + k\}$, for $k \geq -1$, and let $p_k = (u_k, v_k)$ be the point of support. In particular $p_0 = (0, 0)$ and $p_k \approx (0, k)$ for small values of k. The points p_k may drift away from the vertical axis as k grows, and the distance between the halfspaces $a_k = v_k - v_{k-1}$ may diverge to infinity or zero.

Now consider the caps

$$C_k = \{\varphi < \varphi(p) + k\} \cap H_{k-1}.$$

For small k these look like vertical translates of the parabolic cap

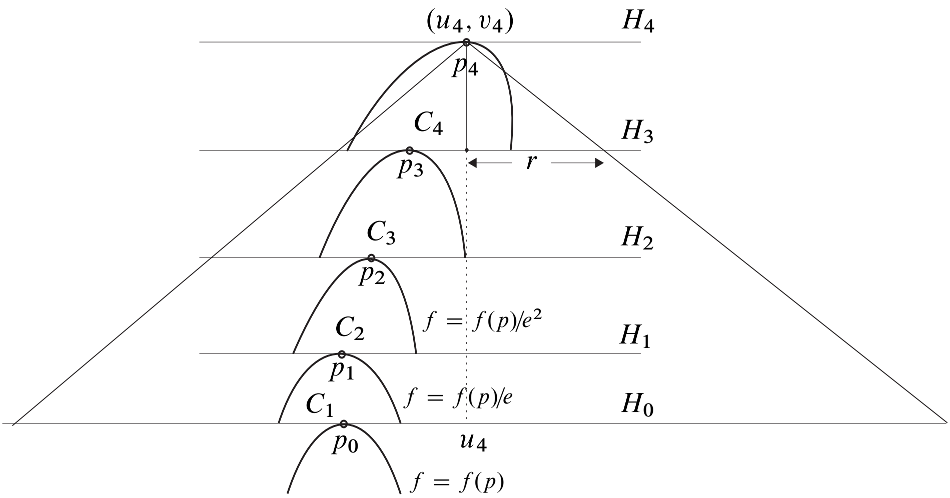
$$Q = \{-1 \leq v < -u^T u/2\}.$$

The cap C_{k+1} balances on the cap C_k in the point p_k.

We want reasonable estimates for the height $a_k = v_k - v_{k-1}$ of the cap C_k, for v_k, for u_k and $u_k - u_{k-1}$, and for the base of the cap. We shall show that the base of C_n

$$C_n \cap \partial H_{n-1}$$

is contained in a disk of radius r_n around the point u_n, where $r_n = 2ne^{n\delta}$, and δ is independent of n, and is small if we start with a point p far out. The cone with top p_n



The disk with center u_4 and radius $r_4 v_4/a_4$ is contained in a disk of radius $R_4 = r_4 u_4/a_4 + \|p_4\|$.

which intersects ∂H_{n-1} in this disk will intersect the coordinate plane $\partial H_0 = \{v = 0\}$ in a disk with radius $r_n v_n/a_n$ and center u_n. This disk is contained in a centered disk of radius $R_n = r_n v_n/a_n + \|p_n\|$, and since $h_p = f/f_p$ it follows that

$$\{h_p > e^{-n}\} \cap H_+ \text{ß} \{\|u\| \leq R_n, 0 \leq v \leq v_n\}.$$

Lemma 10.8 established the bounds

$$\|q_n\| \leq 2ne^{n\delta}, \quad v_n/a_n \leq ne^{n\delta}, \quad v_n \leq ne^{n\delta}, \quad p_n = q_1 + \cdots + q_n.$$

These bounds yield hat boxes of the right size. □

10.4 Conclusion. Let Z have density $f = e^{-\psi \circ n}$ on O where $\psi \colon [0, t_\infty) \to \mathbb{R}$ satisfies (6.4), $O = t_\infty D$ if t_∞ is finite and $O = \mathbb{R}^d$ else, and n is the gauge function of a bounded open convex set D containing the origin. We have the following results:

Proposition 10.10. *Suppose $n(x, 1) - 1 \sim m^p(x)$ for some $p \geq 1$ and some gauge function m of an open bounded convex set in $\mathbb{R}^h$ containing the origin. Define*

$$\alpha_t(u, v) = (bu, t + av), \quad a = a_t = 1/\psi'(t), \ b = b_t = (ta_t)^{1/p}.$$

Then

$$W_t = \alpha_t^{-1}(Z^{H^t}) \Rightarrow W, \quad t \to t_\infty, \ H_t = \{y \geq t\}, \tag{10.8}$$

where W has density $g(u, v) \propto e^{-(m^p(u)+v)}$. All moments converge: for $s > 0$,

$$\mathbb{E}\|W_t\|^s \to \mathbb{E}\|W\|^s, \quad t \to t_\infty.$$

Proof. Convergence of the quotients was established in Proposition 9.11 and the discussion following the proof. In the proof of $\mathbf{L}^1$ convergence, the caps C_k are asymptotically parabolic for $p = 2$. Under the assumption of the proposition $\alpha_n^{-1}(C_n) \to Q_p = \{-1 \leq v < -m^p(u)\}$. The proof carries through in this case. A more direct proof is given in Section 15.2. □

If (10.8) holds, we say that Z lies in the horizontal domain of attraction of W. This domain will be studied in more detail in Section 15.2.

Example 10.11. This example treats the l^p norm on $\mathbb{R}^d$ for $p \neq 2$. Let $n(z) = \|z\|_p = (z_1^p + \cdots + z_d^p)^{1/p}$ with $p \geq 1$. For $p > 1$ the set $D = \{n < 1\}$ is strictly convex. Write $s(z) = z_1^p + \cdots + z_d^p$. Then $s''(z)$ is diagonal with entries z_i^{p-2} on $(0, \infty)^d$. As s and n have the same level sets, in particular $\{s < 1\} = \{n < 1\} = D$, we see that $\partial D \cap (0, \infty)^d$ is C^2 with positive curvature. If a coordinate of $z \in \partial D$ vanishes, the curvature in that direction is infinite for $p < 2$ and zero for $p > 2$. The proposition above describes the asymptotic behaviour of high risk scenarios for horizontal halfspaces (corresponding to the bad behaviour in the North Pole); the proposition below and its corollary describe the asymptotic behaviour of the high risk scenarios for halfspaces with direction in the open positive orthant (corresponding to the good behaviour of n_D on this open cone). It is also possible to describe the asymptotic behaviour in the remaining directions, but we shall not do so. There is an abrupt change in the limit law as one moves to the boundary of the open orthant. Limit powers different from two only occur in isolated directions. There are two explanations for this phenomenon:

1) The boundary of a convex set may look like a quadratic function at all points, but asymptotics with a power $p \neq 2$ can only occur on a null set;

2) the limit density $e^{-(\|u\|_p^p+v)}$ for $p \neq 2$ has a two-dimensional group of symmetries; for $p = 2$ the dimension of the symmetry group is much larger. ◊

Theorem 10.12. *Suppose n_D is C^2 on a neighbourhood of $p_0 \in \partial D$ and $n_D^*(p_0)$, see (9.4), is positive definite. Let the points $p_n \in \partial D$ converge to p_0, and let $t_n \to t_\infty$ from below. Define H_n to be the halfspace supporting $t_n D$ at the point $t_n p_n$. There are affine transformations α_n mapping H_+ onto H_n such that*

$$W_n = \alpha_n^{-1}(Z^{H_n}) \Rightarrow W$$

where W has a Gauss-exponential distribution. Moreover all moments converge:

$$\mathbb{E}\|W_n\|^s \to \mathbb{E}\|W\|^s, \quad s > 0.$$

Suppose $n\mathbb{P}\{Z \in H_n\} \to c_0 \in (0, \infty)$. Then $n\alpha_n^{-1}(\pi) \to \rho$ weakly on halfspaces $J = \{v \geq \xi^T u + \xi_0\}$, where ρ has density $c_1 e^{-(u^T u/2+v)}$ with $c_1 = c_0/(2\pi)^{h/2}$.

Proof. The convergence of the quotients $h_p = f \circ \alpha_p / f(p)$ was described in Proposition 9.12. The $\mathbf{L}^1$ convergence is proved in Theorem 10.9. □

If the gauge function is C^2 on an open set U, it is C^2 on the larger set $V = \bigcup_{t>0} tU$ by homogeneity.

Corollary 10.13. *Suppose n is C^2 on the homogeneous open set Vß$\mathbb{R}^d$ and n^*, see (9.4), is positive definite in each point of the compact set Kß$V \cap \partial D$. Then*

$$\alpha_n^{-1}(Z^{H_n}) \Rightarrow W$$

for any sequence of halfspaces H_n supporting $t_n D$ in $t_n p_n$ with $p_n \in K$. Moreover, all moments converge.

Proof. It suffices to prove this for subsequences p_{k_n} which converge. See Theorem 10.12. □

Proof of Theorem 9.1. Take $V = \mathbb{R}^d \setminus 0$ and $K = \partial D$. □

Proof of Theorem 9.2. The complement of the cone C_m may be covered by $2h$ halfspaces $J_i^{\pm} = \{v \geq -2m \pm 2mhv_i\}$. Convergence $\int_J \|w\|^m d\rho_n \to \int_J \|w\|^m d\rho$ for the mean measures ρ_n and ρ is a consequence of the $\mathbf{L}^1$-convergence of $\|w\|^m f(\alpha_n(w))/f(p_n) \to \|w\|^m e^{-(u^T u/2+v)}/c_1$ on J with $d\rho_n = nc_n\alpha_n^{-1}(d\pi)$. □

11 Flat functions and flat measures

One might compare the role played by flat functions and flat measures in the multivariate theory to that of slowly varying functions in the description of the domain of attraction of the Pareto laws for univariate exceedances. They allow us to describe variations in the tail behaviour which keep the distribution within the domain of attraction, and even retain the original normalizations.

11.1 Flat functions. In this section we extend the class $\mathcal{RE}$ of rotund-exponential densities of Theorem 9.1,

$$f_0 = e^{-\psi \circ n}/C_0. \tag{11.1}$$

The density f_0 above satisfies the limit relation

$$(f_0 \circ \beta_H)(w)/(f_0 \circ \beta_H)(0) \to e^{-v-u^T u/2} \tag{11.2}$$

uniformly on compact w-sets in $\mathbb{R}^d$, and in $\mathbf{L}^1(H_+)$, with β_H as in (9.7). Let f be a continuous positive density on $O = \{f_0 > 0\}$ which satisfies the same limit relation with the same normalizations β_H. Then one may write $f = Lf_0$, where L is a continuous positive function on O which satisfies

$$(L \circ \beta_H)(w)/(L \circ \beta_H)(0) \to 1, \quad H \to \partial_O \tag{11.3}$$

uniformly on compact w-sets.

Relation (11.3) states that far out in O, the function L behaves locally like a positive constant. Such a function will be called *flat*. The condition (11.3) is weaker than asymptotic equality. If L satisfies (11.3) uniformly on compact sets, then $f = Lf_0$ satisfies (11.2) uniformly on compact sets. If Lf_0 is integrable, then one may define the probability density $f_L = Lf_0/C_L$, and ask whether the associated probability distribution lies in the domain of the Gauss-exponential law. In this section, we shall show that this is the case. We shall give a simple sufficient condition for (11.3) in terms of partial derivatives. As a concrete application, we shall show that the multivariate hyperbolic distributions belong to the domain of attraction of the Gauss-exponential high risk limit law. The more geometric point of view allows us to formulate a simple sufficient condition for smooth *unimodal* densities to lie in $\mathcal{D}(0)$, in terms of the first and second derivatives.

11.2 Multivariate slow variation. We start by looking at densities of the form $L(z)f_0(z)/C$, where f_0 is a Gaussian density and L a continuous function which behaves locally like a positive constant when $\|z\| \to \infty$. The question is, how much can one alter a Gaussian density while retaining the high risk limit behaviour (with the original normalizations β_H).

Example 11.1. Let f_0 be an arbitrary Gaussian density on $\mathbb{R}^d$. For $z \in \mathbb{R}^d$ write $z = r\theta$, where $r = \|z\|$ and $\theta \in \partial B$ is a unit vector. The function rf_0 is integrable and the corresponding density lies in the domain of the Gauss-exponential distribution. This is also true for $r^c f_0$ for any $c > 0$, and for $Q^c f_0$ where Q is a positive quadratic function on $\mathbb{R}^d$. The function $e^{\sqrt{r}} f_0$ normalized to a probability density also lies in the domain of attraction, and so does $\chi(\theta) f_0$ for any continuous positive function χ on the unit sphere. Proofs are given below. ◊

In the first part of this section, we consider a class $\mathcal{L}$ of continuous positive functions L on $\mathbb{R}^d$ which satisfy a growth condition which may be regarded as the multivariate additive version of slow variation:

$$L(z + w)/L(z) \to 1, \quad \|z\| \to \infty, \; w \in \mathbb{R}^d. \tag{11.4}$$

If $\lambda\colon \mathbb{R}^d \to \mathbb{R}$ is a C^∞ function whose partial derivatives vanish in infinity, then $L = e^\lambda$ satisfies (11.4). The converse is also true. Continuity of L implies that if (11.4) holds pointwise, it holds uniformly on bounded w-sets; see Bingham, Goldie & Teugels [1989]. Let $\lambda = \log L$, and define $\lambda_0 = \pi_0 * \lambda$ as the convolution of λ with a C^∞ probability density π_0 with compact support. Then

$$\lambda_0(z) - \lambda(z) = \int \pi_0(w)(\lambda(z-w) - \lambda(z))dw \to 0,$$

and the partial derivative of λ_0 of any order at a point z_0 is the convolution of the difference $\lambda(z) - \lambda(z_0)$ with the corresponding partial derivative of π_0. Hence it vanishes at infinity. Thus we have constructed a C^∞ function λ_0 whose partial derivatives vanish at infinity, such that e^{λ_0} is asymptotic to L.

In the univariate case, the functions

$$r, \quad r^7, \quad e^{\sin\sqrt{r}}, \quad e^{\sqrt{r}}, \quad e^{-\sqrt{r}}, \quad r^5 e^{\sqrt{r}\sin r^{1/3}}$$

all satisfy the functional relation

$$f(r+s)/f(r) \to 1, \quad r \to \infty, \tag{11.5}$$

for each $s \in \mathbb{R}$, since $(\log f)'$ vanishes in infinity. Functions in $\mathcal{L}$ satisfy the relation $L(\theta r + \theta s)/L(\theta r) \to 1$, $r \to \infty$, in every direction $\theta \in \partial B$.

In the multivariate setting, one would like to know how the limit relations in different directions are coordinated. The answer is surprising. Given any countable family F of continuous positive functions f on $(0, \infty)$ which satisfy (11.5), for instance the family

$$\alpha r^\gamma e^{\beta r^\theta \sin(r^\eta)}, \quad \alpha, \beta, \gamma, \theta, \eta \in \mathbb{Q}, \ \alpha, \theta > 0, \ \eta \geq 0, \ \theta + \eta < 1,$$

there exists a function $L \in \mathcal{L}$ with the property: for each $f \in F$ there is a dense set S_f of directions θ in the unit sphere $S = \partial B$ such that

$$L(\theta r)/f(r) \to 1, \quad r \to \infty, \ \theta \in S_f, \ f \in F.$$

A construction is given in Balkema [2006]. Here we only want to warn the reader that functions L which satisfy (11.4) are not as tame as they may seem.

Theorem 11.2. *Let the vector $Z \in \mathbb{R}^d$ have density $f_0(z) = e^{-n(z)^2/2}/C$, where n is the gauge function of a rotund set D, and let β_H satisfy (9.7). Suppose $L \in \mathcal{L}$. Then the product $f = Lf_0$ is integrable and satisfies (11.2) uniformly on bounded w-sets, and in $\mathbf{L}^1(H_+)$. In particular the random vector with density Lf_0/C_L lies in the domain of attraction of the Gauss-exponential law with the same normalization as f_0, and the normalized densities converge in $\mathbf{L}^1$.*

Proof. We give only a sketch, since we shall prove a more general result in Theorem 11.4 below. The function $\varphi = n^2/2$ has second derivative $nn'' + n'(n')^T$ which is positive definite and homogeneous of degree zero. So $(\varphi - \lambda)''$ is positive definite outside a compact set if we choose λ, with $e^{\lambda} \sim L$, to be C^{∞} with partial derivatives of all orders vanishing in infinity. This means that the densities g_H of the normalized high risk scenarios are strongly unimodal (logconcave). Hence Lf is integrable, and pointwise convergence in (11.2) implies convergence in $\mathbf{L}^1$.

By (9.7) we may write $\beta_H = A \circ \beta_t$ for an initial linear transformation $A \in \mathcal{J}$. Since $\mathcal{J}$ is compact, and $\psi(t) = t^2/2$ gives $b_t = 1$ and $a_t = 1/t < 1$ eventually, for any ball rB the diameters of the ellipsoids $\beta_H(rB)$ are uniformly bounded. Hence (11.3) holds uniformly on compact w-sets. This yields (11.2). □

We shall now prove a more general result for rotund-exponential densities. Instead of strong unimodality our proof makes use of the fact that $f_0^{1/2}$ is integrable and the corresponding density belongs to the domain of attraction of a high risk limit law. First we have to say more precisely what we mean by a flat function, see (11.3). We shall give a formal definition which also applies to heavy tailed limit laws, or limit laws with bounded support, see Section 12.

Definition. Let Z lie in the domain of attraction of W, i.e. $\beta_H^{-1}(Z^H) \Rightarrow W$ for $0 < P\{Z \in H\} \to 0$. Assume that the high risk limit vector W has density g on H_+, that $\{g > 0\}$ is a convex set, open in H_+, and that g is continuous on $\{g > 0\}$. A function L is *flat for* Z if L is defined, positive and continuous on the interior O of the convex support of Z, or on $O \setminus K$ for some compact subset K of O, if $P\{Z \in O\} = 1$, and if (11.3) holds uniformly on compact subsets of $\{g > 0\}$.

Remark 11.3. If L and L_0 are flat for Z, then so are LL_0, and L^t for $t \in \mathbb{R}$.

11.3 Integrability. We can now formulate the main result of this section.

Theorem 11.4. *Let Z have density $f_0 \in \mathrm{U}_0$. If L is flat for Z, then the product Lf_0 is integrable and the random vector with density Lf_0/C_L lies in the domain of attraction of the Gauss-exponential distribution.*

Proposition 11.5 (Basic Inequality). *Let Z have density $f = e^{-\varphi} \in \mathrm{U}$. Set $O = \{f > 0\}$. For each $p \in O \setminus \{0\}$, let $\beta_p \in \mathcal{A}$ map H_+ onto H_p and 0 into p; see Lemma 10.3. Assume that*

$$h_p(w) = (f \circ \beta_p)(w)/f(p) \to h(w), \quad p \to \partial_O \tag{11.6}$$

uniformly on compact subsets of H_+, and in $\mathbf{L}^1$, for some continuous integrable positive function h on H_+. Let W have density $g = h/C$ on H_+.

Assume that $L = e^{\lambda}$ is positive and continuous on O, and

$$L_p(w) = (L \circ \beta_p)(w)/L(p) \to 1, \quad p \to \partial_O,\ w \in H_+ \tag{11.7}$$

uniformly on $\{h \geq c\}$ *for some* $c < 1$. *Then* Z *lies in the domain of attraction of the high risk limit vector* W, L *is flat for* Z, *and for all* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that*

$$|\lambda(z) - \lambda(p)| < \varepsilon + \varepsilon(\varphi(z) - \varphi(p)), \quad z \in H_p \cap O,\ 0 < f(p) < \delta. \tag{11.8}$$

Proof. Let $\varepsilon > 0$. Write $K = \{h \geq c\}ßH_+$. First observe that the function h has convex level sets $\{h > t\}$ on H_+, since this holds for the functions h_p by Lemma 8.7. Uniform convergence $h_p \to h$ on K implies that there exists $\delta_0 > 0$ such that $h_p(w) < \sqrt{c}$ for $h(w) = c$ and $f(p) < \delta_0$. Hence

$$f(p) < \delta_0,\ z \in H_p \cap O,\ f(z)/f(p) \geq \sqrt{c} \ \Rightarrow\ \beta_p^{-1}(z) \in K.$$

By assumption, there exists $\delta > 0$ such that

$$f(p) < \delta,\ z \in H_p \cap O,\ \beta_p^{-1}(z) \in K \ \Rightarrow | \lambda(z) - \lambda(p)| < \varepsilon.$$

For $p \in O \setminus \{0\}$, $z \in H_p \cap O$, choose points $p_0, \dots, p_n$ on the line segment $[p, z]$ such that $p_0 = p$, $f(p_k) = \sqrt{c} f(p_{k-1})$ for $k = 1, \dots, n$, $f(z) \geq \sqrt{c} f(p_n)$, and set $p_{n+1} = z$. Now suppose $f(p) < \delta_1 = \delta_0 \wedge \delta$. Then $h(\beta_{p_k}^{-1}(p_{k+1})) \geq c$ for $k = 0, \dots, n$. Hence $|\lambda(p_k) - \lambda(p_{k-1})| < \varepsilon = \varepsilon(\varphi(p_k) - \varphi(p_{k-1}))/\log(1/\sqrt{c}$, and

$$|\lambda(z) - \lambda(p)| < \varepsilon + \varepsilon_0(\varphi(z) - \varphi(p)),$$
$$\varepsilon_0 = \varepsilon/\log(1/\sqrt{c}), \quad z \in H_p \cap O,\ 0 < f(p) < \delta.$$

In particular, L is flat for Z since $\varepsilon > 0$ is arbitrary. □

Proof of Theorem 11.4. Flatness implies (11.1) for Lf uniformly on compact subsets of H_+. The basic inequality gives $L(z)/L(p) \leq 2(f(p)/f(z))^{\varepsilon}$ for $z \in H_p$, $f(p) < \delta$. Hence Lf is integrable if $\sqrt{f}$ is integrable. If ψ satisfies (6.4) then so does $\psi/2$. Apply Theorem 8.8 on power families to give $\mathbf{L}^1$ and pointwise convergence of the quotients $h_p^{1/2}$. By dominated convergence, $((Lf) \circ \beta_p)(w)/(Lf)(p) \to h$ in L^1. □

11.4* The geometry. In the remainder of this section we develop the theory of flat functions further.

First we give a geometric formulation of condition (11.3). A positive continuous function $L = e^{\lambda}$ on O is flat if it is asymptotically constant on the ellipsoids $E_q = E_H = \beta_H(B)$ for $H = H_q$ with $q \in O \setminus \{0\}$. For rotund-exponential densities $f = e^{-\psi \circ n}$ the ellipsoids E_q have the form $A \circ \beta_t(B)$, where A is an initial map and $\beta_t(u, v) = (b_t u, t + a_t v)$ with $a_t = 1/\psi'(t)$ and $b_t = \sqrt{t a_t}$. Since $a_t = o(t)$ for $t \to t_\infty$ by Theorem 6.1, we see that

$$a_t \ll b_t \ll t, \quad t \uparrow t_\infty.$$

Hence the diameter of the ellipsoids $\beta_t(B)$ vanishes for $t \to t_\infty$ if t_∞ is finite, and otherwise it is $o(t)$. Since the set $\mathcal{J}$ of initial maps is compact by Theorem 9.9, these bounds also hold for the ellipsoids E_H. The ellipsoids E_H are like buttons attached to the surface of the rotund set tD at the point $z = z_H$, where H supports tD.

Note that $q \mapsto E_q$ is continuous, in spite of Theorem 9.10. In principle, such a continuous family of ellipsoids defines a Riemannian metric with continuous geodesics. The distance between neighbouring points is roughly $2n\varepsilon$ for $\varepsilon > 0$ small, where n is the number of ellipsoids E_z^ε needed to form a chain between the two points. Here we define

$$E_p^r = \beta_p(rB) = p + r(E_p - p).$$

One may think of the family of ellipsoids (E_q) as the *geometry associated with the density* f_0. Since the geometry is determined by the normalization, the density Lf_0, with L flat, has the same geometry as f_0. If we replace the normalization β_H by a normalization $\tilde{\beta}_H$ which gives the same limit law, then by the Convergence of Types Theorem the associated ellipsoids are asymptotic: For any $\varepsilon > 0$ there is a compact subset K of O such that $E_q^{1-\varepsilon}\text{ß}\widetilde{E}_q\text{ß}E_q^{1+\varepsilon}$ for $q \in O \setminus K$.

Example 11.6 (Functions which are flat for all densities $f_0 = e^{-\psi\circ n}/C_0$).

1) If t_∞ is finite, then $O = \{f_0 > 0\}$ is a bounded set. Let L be a continuous positive function, defined on an open neighbourhood V of ∂O. The restriction of L to $O \cap V$ is flat.

2) Let χ_0 be a positive continuous function on the unit sphere ∂B in $\mathbb{R}^d$. Then $z \mapsto \chi(z) = \chi_0(z/\|z\|)$ on $\mathbb{R}^d \setminus \{0\}$ is *positive-homogeneous* of degree zero, and flat for all f_0 in (11.1). So is $z \mapsto r = \|z\|$. ◊

Recall that the standardized generalized multivariate *hyperbolic distribution* has density

$$f(z) \propto (1 + z^T z)^{c_1} K_{\lambda - d/2}(c_2\sqrt{1 + z^T z})e^{\gamma z}, \quad z \in \mathbb{R}^d. \tag{11.9}$$

Here K_ν denotes a modified Bessel function of the third kind with index ν, and the constants $c_1 \in \mathbb{R}$, $\lambda \in \mathbb{R}$, $c_2 > 0$, and $\gamma \in \mathbb{R}^d$ with $\|\gamma\| < c_2$ are parameters, with $c_1 = (\lambda - d/2)/2$.

Proposition 11.7. *The Gauss-exponential domain contains the hyperbolic densities* (11.9).

Proof. The asymptotic behaviour of the Bessel function does not depend on the index. For any $\nu \in \mathbb{R}$

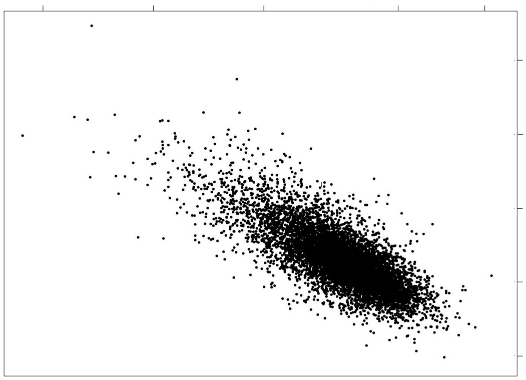
$$K_\nu(t) \sim \sqrt{\pi/2t}e^{-t}, \quad t \to \infty.$$

Since $e^{-c\sqrt{1+z^T z}} \sim e^{-cr}$ for $\|z\| = r \to \infty$, we see that

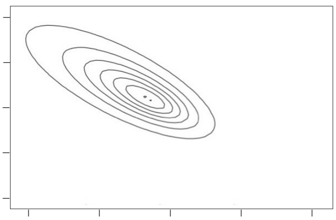
$$f(z) \sim c_0 r^{2c_1 - 1/2} e^{-c_2 r} e^{\gamma z} = c_0 r^{2c_1 - 1/2} e^{-n(z)} = L(z)e^{-n(z)}, \quad r = \|z\| \to \infty,$$

where n is the gauge function of the rotund set $D = \{z \in \mathbb{R}^d \mid c_2^2 z^T z < 1 + \gamma z\}$, an ellipsoid, which is excentric for $\gamma \neq 0$, and L is flat for e^{-n}, by Remark 11.3 and the example above. □

The simulated 10 000-point sample cloud from an asymmetric hyperbolic distribution and the level curves below were kindly made available by Alex McNeil using the S-plus programs in McNeil, Frey & Embrechts [2005]. The level curves of the density are close to scaled copies of an off-center ellipse.



Sample cloud.



Level curves.

We shall now take a more geometric point of view. Let d be a distance associated with the ellipsoids E_p as described above. Then flat functions are slowly varying in the sense that

$$z_n \to \partial,\ d(z_n', z_n) \text{ bounded } \Rightarrow\ L(z_n') \sim L(z_n).$$

The geometry is asymptotic in the sense that for bounded domains O, for any $r > 1$ there exists a compact set K_rßO such that

$$E_p^r \text{ß} O, \quad p \in O \setminus K_r.$$

In the geometry (E_p), the set O becomes more spacious as one approaches the boundary. In this sense it behaves like hyperbolic geometry.

The normalizations α_p may be replaced by the normalizations $\alpha_p \circ R_p$, where R_p are arbitrary rotations around the vertical axis in u, v-space. This does not affect convergence. Neither does it affect the geometry. That may explain why the geometry may be chosen to vary continuously, whereas in certain dimensions it is not possible to choose the normalizations α_p to depend continuously on the point $p \in O \setminus \{0\}$; see Theorem 9.10.

If $L = e^{\lambda}$, where λ is C^1, then one may formulate sufficient conditions for flatness in terms of the partial derivatives. Let λ_r be the radial derivative and $\lambda_0(p)$ the maximum of the partial derivatives along unit vectors in the tangent plane to $\{f = f(p)\}$ in p:

$$\lambda_0(p) = \max\{d\lambda(p + te)/dt \mid e \in \partial H_p,\ \|e\| = 1\}.$$

One may show that L is flat for the density $e^{-\psi \circ n}/C$ if

$$\lambda_r(p) = o(\psi'(n(p))), \quad \lambda_0(p) = o(\sqrt{\psi'(n(p))/n(p)}).$$

Example 11.8. Suppose $Z = (X, Y) \in \mathbb{R}^2$ is standard Gaussian. We have seen that $r^{\sin\theta}$ is flat, and so is $e^{\sqrt{r}\sin\theta}$ since the partial derivatives λ_x and λ_y of the exponent vanish, and hence these functions lie in $\mathcal{L}$. The function $e^{r^{7/4}}$ is also flat for Z (but only if the Gaussian distribution is spherically symmetric). Flat functions $L = e^{\lambda}$ which increase so fast in any given direction are rather stiff, they have to increase at this rate in every direction. Along circles the slope of λ tends to zero, and hence $\lambda(r, 0) - \lambda(0, r) = o(r)$. $\diamondsuit$

Assume some extra smoothness. For a C^2 function φ on a vector space the linear approximation in the point q does not depend on the coordinates, neither does the second-order Taylor approximation, or the difference $\varphi''(q)$, which is a quadratic form on the vector space V. In our case we consider for given q the quadratic form $Q = \varphi^*(q) = \varphi''(q) + \varphi'(q) \otimes \varphi'(q)$. For $\varphi = \psi \circ n$ as in Theorem 9.1, the form Q is positive definite outside the origin, and we define the ellipsoid $E_q^* = q + \{Q < 1\}ßV$. The definition of this ellipsoid is geometric and does not depend on the coordinates. We can now introduce affine coordinates on the vector space V which are adapted to the ellipsoid E_q^*. These coordinates are a map $\gamma_q : \mathbb{R}^h \times \mathbb{R} \to V$. They are chosen so that the origin is the point q, the upper halfspace is the halfspace which supports the rotund set $\{\varphi < \varphi(q)\}$ in the point q, and the unit ball in the new coordinates is the ellipsoid E_q^*. Hence $E_q^* = \gamma_q(B)$. The coordinate map γ_q transforms the geometry (E_z^*) on V into a geometry (F_w^q) on $\mathbb{R}^h \times \mathbb{R}$:

$$F_w^q = \gamma_q^{-1}(E_z^*), \quad z = \gamma_q(w) \neq 0.$$

We claim that the geometry (F_w^q) converges for $q \to \partial$. The limit is the parabolic geometry (F_w). Here are two definitions of *parabolic geometry* on $\mathbb{R}^h \times \mathbb{R}$:

1) The *Riemannian geometry* induced by the positive definite quadratic form $\chi^* = \chi'' + \chi' \otimes \chi'$, where $\chi(u, v) = u^T u/2 + v$;

2) The family (F_w), where F_w is the ellipsoid $\sigma(B)$, $w = \sigma(0)$, with $\sigma \in \mathcal{G}$. Here $\mathcal{G}$ is the symmetry group of the measure ρ_0^* with density $e^{-\chi}$; see Section 4.

The equivalence of the two definitions follows from the invariance of χ^* under $\mathcal{G}$.

The parabolic geometry is not very intuitive. It preserves the vertical direction but not the (Euclidean) shape of an ellipsoid. The ellipsoids

$$E_w = w + \{\chi^*(w) < 1\}$$

become flat and sloped as one moves away from the vertical axis. In the plane the ellipse E_p around the point $p = (a, -a^2/2)$ on the parabola $v = -u^2/2$ may be visualized as the ellipse inscribed in the parallelogram formed by the two vertical lines $u = a \pm 1$ on either side of p and the two lines $v = a^2/2 - au \pm 1$ above and below the tangent line to the parabola $v = -u^2/2$ in the point p. These inscribed ellipses are the image of the unit disk inscribed in the square $\{|u| = 1, |v| = 1\}$ under the element $\sigma \in \mathcal{G}$ which maps the origin into p. The transformations might be called a parabolic translation since it preserves all parabolas $v = c - u^2/2$. The parabolic geometry on $\mathbb{R}^{h+1}$ is a true geometry in the sense of Klein's Erlangen program. In fact it is equivalent to the subgeometry of Euclidean geometry which preserves the vertical direction. (The equivalence is established by the map $(u, v) \mapsto (u, v + u^2/2)$ which preserves vertical lines and transforms parabolas into horizontal planes; see Section 8.3. The map transforms the Gauss-exponential measure $d\rho = e^{-(v+u^T u/2)} du dv$ into the product measure $e^{-v} du dv$. However, the map is non-linear, and does not preserve the class of ellipsoids. It is not clear how it should be used to analyse the asymptotic behaviour of densities in U_0.) The transformation group $\mathcal{G}$ of the parabolic geometry on $\mathbb{R}^d$ has dimension $1, 2, 4, 7, 11, \ldots$ for $d = 1, 2, \ldots,$ consisting of a d-dimensional group of parabolic translations, including the vertical translations, and a $(h^2 - h)/2$-dimensional group of rotations around the vertical axis. Each of these symmetries yields a bijection of the family of ellipsoids $(E_w, w \in \mathbb{R}^d)$.

Lemma 11.9. *Let Σ be a positive definite quadratic form on the vector space V, and $\xi \neq 0$ a linear functional. There exists a unique $c_0 \in \mathbb{R}$ such that $\Sigma - c_0 \xi \otimes \xi$ has rank less than the dimension of V.*

Proof. Choose coordinates $\zeta_1, \ldots, \zeta_d$ such that the ellipsoid $\{\Sigma < 1\}$ is the unit ball. Write $\xi = c_1\zeta_1 + \cdots + c_d\zeta_d$ and set $c_0 = 1/(c_1^2 + \cdots + c_d^2)$. Then $\sqrt{c_0}\xi$ is a unit vector. Replace $\zeta_1, \ldots, \zeta_d$ by orthonormal coordinates $\zeta_1', \ldots, \zeta_d'$ such that $\zeta_1' = \sqrt{c_0}\xi$. Then $\Sigma - tc_0(\xi \otimes \xi)$ has matrix $\mathrm{diag}(1 - t, 1, \ldots, 1)$. □

Theorem 11.10. *Let $f = e^{-\varphi} \in \mathrm{U}$ be C^2 on $O = \{f > 0\}$ and suppose $\varphi^* = \varphi'' - \varphi' \otimes \varphi'$ is positive definite and $\varphi' \neq 0$ on $O \setminus K$ for some compact set KßO. Let $c(w)$ be the constant such that $\varphi^*(w) - c(w)(\varphi' \otimes \varphi')(w)$ has rank $< d$. For $p \in O \setminus K$, choose affine normalizations β_p mapping H_+ onto H_p such that*

$$\varphi_p(w) = \varphi(\beta_p(w)) - \varphi(p)$$

satisfies

$$\varphi_p'(0) = (0, 1), \quad \varphi_p''(0) = \mathrm{diag}(1, \ldots, 1, c(w)).$$

Suppose $E_p^r = \beta_p(rB) ß O$ *for* $p \in O$, $f(p) < \delta_r$. *Then* $f \in \mathrm{U}_0$ *if*

$$\varphi''_{p_n}(w_n) \to \mathrm{diag}(1, \dots, 1, 0), \quad p_n \to \partial_O,\ w_n \to w_0,\ w_0 \in \mathbb{R}^d. \tag{11.10}$$

Proof. It suffices to show that $\varphi_p(u, v) \to v + u^T u/2$ pointwise for $p \to \partial_O$. For this it suffices that

$$\varphi_p(0, 0) \to 0, \quad \varphi'_p(0) \to (0, 1), \quad p \to \partial_O,$$

and that (11.10) holds. □

Corollary 11.11. *If f satisfies the conditions of the theorem, then f is integrable and the corresponding probability distribution lies in the domain $\mathcal{D}(0)$ of the Gauss-exponential law.*

Remark 11.12. The assumptions of the theorem may be whittled down. If the sets $C_n = \{\varphi \le n\}$ are compact subsets of O, then one of these sets contains K. If we assume that O is connected, then so is this set C_n, since $\varphi' \ne 0$ implies that $f = e^{-\varphi} \wedge e^{-n}$ has no saddle points, and no local maxima outside the plateau C_n. The condition φ^* is positive definite implies that level surfaces $\{\varphi = c\}$ for $c > n$ are locally convex, and hence convex, see Valentine [1964] Theorem 4.4. The infinitesimal behaviour expressed in φ^* determines the global behaviour, unimodality of f. As in Theorem 6.10 it suffices that the condition (11.10) holds for $w_0 \in qB$ for some $q > 0$, and that $E_p^q ß O$ eventually.

Let us check that the rotund-exponential densities satisfy the conditions of the theorem above, with $K = \{0\}$.

Proposition 11.13. *Let $\varphi_{A,t} = \psi \circ n \circ A \circ \beta_t$. Then for $t \to t_\infty$, for any $R > 1$*

$$\varphi'_{A,t}(w) \to (u, 1)^T, \quad \varphi''_{A,t}(w) \to I_h = \mathrm{diag}(1, \dots, 1, 0)$$

uniformly for $A \in \mathcal{J}$, $\|w\| \le R$, $w = (u, v)$, with $\mathcal{J}$ as in Theorem 9.9.

Proof. The proof is an application of the chain rule. First observe that the normalizations A and β_t have been chosen so that $\varphi'_{A,t}(0, 0) = (0, 1)$. Set $m = n \circ A$, $z_t = \beta_t(w) = (b_t u, t + a_t v)$, see (9.6), $t^* = m(z_t)$, $B_t = \mathrm{diag}(b_t, \dots, b_t, a_t)$ the linear part of β_t. Then $\varphi'_{A,t}(w) = \psi'(t^*) n'_A(z_t) B_t$. Observe that $t^* - t = O(a_t)$ implies $\psi'(t^*) \sim \psi'(t) = 1/a_t$. By homogeneity $m_x(x, y)x + m_y(x, y)y = m(x, y)$. Also $m_x(x, 1) = x^T + e(x)$, where the components of $e(x)$ are $o(\|x\|)$, since $A \in \mathcal{J}$ implies $n_{xx}(0, 1) = I$. So $m_y(x, 1) - 1 = O(x^T x)$. By homogeneity,

$$m'(b_t u, t + a_t v) = m'(b_t u/(t + a_t v), 1) = (b_t u/t, 1) + (o(b_t/t), O(b_t^2/t^2)).$$

This gives $\varphi'_{A,t}(w) \to (u, 1)$. The second derivative may be written as

$$\varphi''_{A,t}(w) = a'(t^*)\varphi'_{A,t}(w) \otimes \varphi'_{A,t}(w) + \psi'(t^*) B_t n''_A(z_t) B_t.$$

The first term vanishes in t_∞, since $a'(t)$ vanishes and $t^*/t \to 1$. By continuity, $m''(z) \to I_h$ for $z \to (0,1)$, and hence by homogeneity we may write $m''(b_t u, t + a_t b) = (I_h + E)/t$ where the entries of E all are $o(1)$. Then this also holds for the entries of $\psi'(t)B_t E B_t/t$. Now observe that $\psi'(t)B_t I_h B_t/t = I_h$. Finally it should be checked that the relations also hold for $A_n \to A \in \mathcal{J}$, $t_n \to t_\infty$ and $w_n \to w \in \mathbb{R}^d$. □

For $f \in \mathrm{U}_0$ the geometry converges to the parabolic geometry.

Proposition 11.14. *Let $f \in \mathrm{U}_0$, $p_n \to \partial_O$ where $O = \{f > 0\}$, $r > 1$, $q_n \in E^r_{p_n}$, $\alpha^{-1}_{p_n}(q_n) \to w_0$. Then $w_n = \alpha^{-1}_{p_n}(E_{q_n}) \to F_{w_0} = \sigma_{w_0}(B)$, where σ_{w_0} is the symmetry mapping the origin into w_0.*

Proof. Write $f = e^{-\varphi}$ on O. Then

$$\varphi_n(w) = \varphi(a_{p_n}(w)) - \varphi(p) \to \chi(w) = u^T u/2 + v, \quad w = (u,v) \in \mathbb{R}^{h+1}.$$

Convergence $w_n \to w_0$ implies convergence $c_n = \varphi_n(w_n) \to c_0 = \chi(w)$ by Proposition 10.1. Let C_n be the cap

$$C_n = \{\varphi_n < c_n + 1\} \cap H_n,$$

where H_n is the halfspace supporting $\{\varphi_n < c_n\}$ in w_n. Since $\varphi_n \to \chi$ is uniform on bounded sets by Proposition 10.1, the level sets $\{\varphi_n < c_n + 1\}$ converge to the parabola $\{v < c_0 + 1 - u^T u/2\}$, and the caps C_n converge to the parabolic cap $C_0 = \{\chi < c_0 + 1\} \cap H_0$, where H_0 supports the paraboloid $\{\chi < c_0\}$ at w_0. The semi-ellipsoid $F_n \cap H_n$ with $F_n = \alpha^{-1}_{p_n}(E_{q_n})$ fits into the cap C_n just as the semi-ellipsoid $F_{w_0} \cap H_0$ fits into the cap C_0. Indeed, $\alpha^{-1}_{q_n}(C_n) \to Q_+$ and $\alpha^{-1}_{q_n}(E_{q_n}) \to B$, and the symmetry which maps B into F_{w_0} maps Q_+ into C_0. Instead of the ellipsoids E_p one might use the corresponding caps to define the geometry. Convergence $C_n \to C_0$ is equivalent to convergence $F_n \to F_{w_0}$. □

11.5 Excess functions. Univariate exceedances are handled effectively by tail functions T, where $T(y) = 1 - F(y) = P\{Y > y\}$. For a random vector Z the tail function is the *excess function* $e_Z(H) = P\{Z \in H\}$. By a well-known result for multivariate characteristic functions the excess function of the vector Z determines the distribution; see Section 3.1. In general, excess functions are hard to characterize. Here we shall give an asymptotic formula for the excess function of densities $f = Le^{-\psi \circ n}/C$ with L flat.

Theorem 11.15. *Let D be a rotund set in $\mathbb{R}^d$ with gauge function n. Let Z have density*

$$f(z) \sim L(z) f_0(z)/C, \quad z \to \partial_O, \ O = \{f_0 > 0\},$$

where $f_0 = e^{-\psi \circ n}/C_0$ *is rotund-exponential, and* L *is flat for* f_0*. Let* $H = H_z$ *be the closed halfspace supporting* tD *in the point* $z = z_H = tp$*. Then*

$$P\{Z \in H\} \sim f(z_H)L_0(z_H), \quad \mathbb{P}\{Z \in H\} \to 0+,$$

where L_0 *is flat for* f_0*,* $L_0(tp) = C(p)t^{h/2}/\psi'(t)^{1+h/2}$ *with*

$$C(p) = |\det A_p| = 1/\sqrt{\det Q(p)}. \tag{11.11}$$

Here A_p *is any initial map for* p *and* $Q(p) = (n'' + n' \otimes n')(p)$*. Moreover,*

$$f(z_H) \sim \max_{z \in H} f(z) \sim \max_{z \in \partial H} f(z).$$

Proof. By (8.11), we have $P\{Z \in H_z\}/f(z) \sim (2\pi)^{h/2}|\det \beta_z|$, where $\beta_z = A_p\beta_t$; see (9.7). Now $\det \beta_t = a_t b_t^h = t^{h/2}/\psi'(t)^{1+h/2}$, and $|\det A_p| = 1/\sqrt{\det Q(p)}$ since $n_A = n \circ A$. The asymptotic relations in the last line hold since $\max f(H) = \max_{w \in H_+} f(\beta_H(w))$ and $L \circ \beta_H$ is asymptotically constant on bounded sets. □

11.6* Flat measures. So far, life was simple because the vectors Z had a well-behaved density f, yielding the limit relation

$$(f \circ \beta_p)(w)/f(p) \to e^{-v-u^Tu/2}, \quad p \to \partial_O, \ O = \{f > 0\}$$

uniformly on compact sets of $\mathbb{R}^d$ (and in $\mathbf{L}^1(H_+)$). Such simple densities are in accordance with our basic assumption that the underlying sample cloud is bland, consisting of a dark convex central region surrounded by a homogeneous halo. A distribution with a density in U_0, or of the form $f_0 = e^{-\psi \circ n}/C_0$ as in Theorem 9.1, or of the form $f = Lf_0/C$ with L flat for f_0, may fit such a data set well.

Our basic limit relation (8.2) is phrased in terms of weak convergence. Hence the theory developed in this book also should allow discontinuous densities or even discrete probability distributions.

Example 11.16. Suppose that the random vector Z has integer components and $P\{Z = k\} \sim f(k)$ for $\|k\| \to \infty$, where $f = Lf_0$ with $f_0 = e^{-\psi \circ n}/C_0$, and L flat for f_0. Assume $t_\infty = \infty$ and $\psi'(t)$ vanishes for $t \to \infty$. This implies that the ellipsoids E_p will contain arbitrarily large balls as $\|p\| \to \infty$. So asymptotically the counting measure on the integer lattice will behave like *Lebesgue measure* on these large ellipsoids. Does Z lie in the Gauss-exponential domain? The theorem below will handle this question. ◊

Let Z have a probability distribution of the form

$$d\pi = f_0 d\mu,$$

where μ is a roughening of Lebesgue measure for the family of ellipsoids $E_p = \beta_p(B)$ associated with the density $f_0 = e^{-\psi \circ n}/C_0$. What this means will be explained below. Let us assume that the normalizations β_H associated with f_0 in (9.7) may be used to normalize the distribution π^H of the high risk scenario Z^H

$$\beta_H^{-1}(d\pi^H)(w) = (f_0 \circ \beta_H)(w)\beta_H^{-1}(d\mu)/C_H \to e^{-v-u^T u/2}dw/(2\pi)^{h/2}$$

weakly for $H \to \partial$, see (8.9). By Theorem 9.1,

$$f_0 \circ \beta_H(w)/f_0 \circ \beta_H(0) \to e^{-v-u^T u/2}$$

uniformly on compact sets. This means that

$$\beta_H^{-1}(d\mu)/C'_H \to d\lambda, \quad C'_H = (2\pi)^{h/2}C_H/(f_0\beta_H)(0),$$

where λ is Lebesgue measure on H_+ and $\to$ denotes vague convergence. Similarly, the last limit relation together with the second implies the first in the sense of vague convergence.

Let us call a Radon measure μ_0 on $\mathbb{R}^d$ a *roughening* of Lebesgue measure *for the Euclidean norm* if there exists a countable partition $\mathcal{F}$ of $\mathbb{R}^d$ into bounded Borel sets F such that $\mu_0 F/|F| \to 1$ and such that the diameter of F_p vanishes for $\|p\| \to \infty$, where F_p is the element of the partition $\mathcal{F}$ containing the point p. The measure μ_0 translated over z will converge vaguely to Lebesgue measure λ for $\|z\| \to \infty$. Using this simple concept as a guide, we now define the asymptotic relation we are really interested in:

Definition. A Radon measure μ on $O = \{f_0 > 0\}$ is called a *roughening of Lebesgue measure for* the density $f_0 = e^{-\psi \circ n}/C_0$ in Theorem 9.1 if there exists a countable partition $\mathcal{F}$ of O into Borel sets F such that

1) $\lambda F > 0$ and $\mu F/\lambda F \to 1$;

2) for all $\varepsilon > 0$ there exists a compact set $K ß O$ such that $F_p \in E_p^\varepsilon$ for $p \in O \setminus K$.

The measure μ is called *flat for* f_0 if 1) above is replaced by the condition

1′) $\lambda F > 0$ and $\mu F_p/\lambda F_p \sim L(p)$ for $p \to \partial_O$ for a function L, flat for f_0.

In Section 6 it was shown that $f_0 L d\mu \in \mathcal{D}^+(0)$ if f_0 satisfies the von Mises condition, if L is flat for f_0, and μ is a roughening of Lebesgue measure. In the multivariate setting a similar result holds. We first prove vague convergence. Recall that $f = e^{-\varphi} \in \mathrm{U}_0$ if f is a unimodal function which satisfies some regularity conditions, $f \in \mathrm{U}$, and if there exist $\alpha_p \in \mathcal{A}$ mapping H_+ onto H_p and $(0, 0)$ into p such that

$$\varphi_p(w) = \varphi(\alpha_p(w)) - \varphi(p) \to \chi(w) = v + u^T u/2, \quad w = (u, v) \in \mathbb{R}^{h+1}.$$

Proposition 11.17. *Let $d\nu = f d\mu$ with $f \in \mathrm{U}_0$ and μ on $O = \{f > 0\}$ flat for f. Then*

$$\alpha_p^{-1}(\nu)/C(p) \to \rho \text{ vaguely on } \mathbb{R}^d, \quad p \to \partial_O,$$

where ρ has density $e^{-(v+u^T u/2)}$ and $C(p) = f(p)L(p)/|\det A_p|$ writing $\alpha_p(w) = A_p w + b_p$.

Proof. The left hand side has the form

$$\frac{f(\alpha_p(w))}{f(p)} \frac{d\alpha_p^{-1}(\mu)|\det A_p|}{L(p)}.$$

The first factor converges pointwise to $e^{-\chi}$, and even uniformly on bounded sets by Proposition 10.1. It suffices to show that the second factor converges vaguely to Lebesgue measure. For this it suffices to show that for any $r > 1$ the diameter of the sets $\alpha_p^{-1}(F)$ with $F \in \mathcal{F}$, the partition associated with the roughening, and $\alpha_p^{-1}(F) \cap rB \neq \emptyset$, goes to zero uniformly in F for $p \to \partial_O$. Here we use the equivalence of the parabolic and the Euclidean geometry on compact sets. Let $\mathcal{G}$ be the group of symmetries of the parabolic geometry, the symmetry group of ρ. There is a constant $R > 1$ such that each ellipsoid $\sigma(B)$ of the parabolic geometry which intersects the ball rB is contained in the ball $\sigma(0) + RB$.

Now let $\varepsilon > 0$ and let $\delta \in (0, 1)$. Let $p_n \to \partial_O$ and let $F(q_n) \in \mathcal{F}$ intersect $E_{p_n}^r = \alpha_{p_n}(rB)$. Then $F(q_n)$ß$E_{q_n}^{\delta}$ eventually. By convergence to the parabolic geometry, $\alpha_{p_n}^{-1}(F(q_n))$ß$\sigma_{w_n}(2\delta B)$ eventually, where σ_{w_n} is a symmetry mapping the origin into $w_n = \alpha_{p_n}^{-1}(q_n)$. Take $\delta = \varepsilon/2R$ to conclude that the diameter of $\alpha_{p_n}(F(q_n))$ does not exceed 2ε. □

The theorem below concerns functions of the form $f_0 L$, where f_0 is a rotund-exponential density. We surmise that a similar result also holds for $f_0 \in \mathrm{U}_0$.

Theorem 11.18. *Suppose $f_0 = e^{-\psi \circ n}/C_0$ satisfies the conditions of Theorem 9.1. Let L be flat for f_0, and let μ be a measure on $O = \{f_0 > 0\}$ which is flat for f_0. Set $f = Lf_0$. Then $f d\mu$ is a finite measure and the corresponding probability measure $f d\mu/C$ lies in the domain of attraction of the Gauss-exponential limit law with the normalizations of f_0.*

Proof. The proof is rather technical. Basically, for weak convergence one still has to show that

$$\int_{U_r} f(z)dz < \varepsilon \int_H f(z)dz, \quad r > r_\varepsilon, \; f_0(p_H) < \delta, \; f = f_0 L,$$

where U_r is the union of all sets in the partition $\mathcal{F}$ which contain a point of the boundary of H, but which lie outside the ellipsoid $E_H^r = \alpha_H(rB)$. See Balkema & Embrechts [2004], Theorem 7.1, for details. □

12 Heavy tails and bounded vectors

This section contains a short description of high risk limit laws with heavy tails, and of bounded limit vectors. The global limit theory for heavy tails will be treated in a more general setting in Section 16.

12.1 Heavy tails. We start with a simple result.

Theorem 12.1. *Let Z in $\mathbb{R}^d$ have a spherically symmetric density $f(z) = f_0(\|z\|)$. If f_0 varies regularly with exponent $-(\lambda + d)$ for some $\lambda > 0$, then Z lies in the domain of attraction of a high risk limit vector $W = (U, V)$ on H_+, with density*

$$g(w) = \frac{1/C}{((1+v)^2 + u^T u)^{(\lambda+d)/2}}, \quad w = (u, v) \in \mathbb{R}^h \times [0, \infty), \tag{12.1}$$

where

$$C = C(\lambda, d) = (\pi^{h/2}/\lambda)(\Gamma((\lambda+1)/2)/\Gamma((\lambda+d)/2)). \tag{12.2}$$

Proof. Recall that regular variation means that $f_0(rs)/f_0(s) \to 1/r^{\lambda+d}$ as $s \to \infty$, for fixed $r > 0$. This implies

$$h_s(w) = \frac{f(sw)}{f(sp_0)} \to h(w) = \frac{1}{\|w\|^{d+\lambda}}, \quad s \to \infty, \tag{12.3}$$

where p_0 is any unit vector.

By a well-known inequality, see Bingham, Goldie & Teugels [1989],

$$f_0(rs)/f_0(s) \le 2/r^{\lambda/2+d}, \quad s \ge s_0, \ r \ge 1.$$

This ensures convergence of the integrals

$$\int_{\|z\|\ge 1} \frac{f(sz)dz}{f_0(s)} = b(d) \int_1^\infty \frac{r^{d-1} f_0(rs)}{f_0(s)} dr \to b(d) \int_1^\infty \frac{dr}{r^{\lambda+1}} = \frac{b(d)}{\lambda},$$

where $b(d) = 2\pi^{d/2}/\Gamma(d/2)$ is the area of the surface ∂B of the unit ball in $\mathbb{R}^d$. Let W_s denote the vector Z/s conditioned to lie outside the unit ball. Then W_s has density

$$g_s(w) = c_s \frac{f(sw)}{f_0(s)} \to \frac{b(d)}{\lambda \|w\|^{\lambda+d}}, \quad s \to \infty, \ \|w\| \ge 1.$$

Let H be the closed halfspace $\{\theta \ge s\}$. By symmetry, we may assume that θ is the vertical unit vector. Then Z^H/s is distributed like W_s conditioned to lie in the halfspace $\{v \ge 1\}$, and $Z^H/s - \theta \Rightarrow W$, where W lives on $\{v \ge 0\}$ with the density g

above. The constant C is the value of the integral

$$\int_0^\infty \int_{\mathbb{R}^{d-1}} \frac{dvdu}{((1+v)^2+u^Tu)^{(\lambda+d)/2}} = \int_0^\infty \frac{dv}{(1+v)^{\lambda+1}} b(h) \int_0^\infty \frac{r^{d-2}dr}{(1+r^2)^{(\lambda+d)/2}}.$$

The second integral yields a beta function as in (13.1). □

Definition. The distribution with density g on H_+ defined in (12.1) and (12.2) is called a *Euclidean Pareto distribution with exponent* λ.

The vertical component V and the horizontal component U are not independent. The vertical component has a Pareto distribution, $P\{V > v\} = 1/(1+v)^\lambda$. The vector U has a spherically symmetric density on $\mathbb{R}^h$ which is proportional to $J_{\lambda+h}(\|u\|)/\|u\|^{\lambda+h}$, where

$$J_c(s) = \int_{1/s}^\infty \frac{dy}{(1+y^2)^{(c+1)/2}} = \int_{1/\sqrt{1+s^2}}^1 (1-r^2)^{c/2-1}dr, \quad c > 0,\ s \geq 0. \quad (12.4)$$

Spherical Pareto distributions extend to infinite Radon measures. It is convenient to use a normalization for this Radon measure which is better adapted to the basic limit relation (12.3). Let ρ denote the measure on $O = \mathbb{R}^d \setminus \{0\}$ with density $1/\|w\|^{d+\lambda}$. Then all closed halfspaces $H ß O$ have finite mass, and the associated probability measure $d\rho_H = 1_H d\rho/\rho(H)$ is the image of the *Euclidean Pareto* limit distribution under an affine map σ_H from H_+ onto H, which is a symmetry of the excess measure ρ.

It is interesting to compare the behaviour of a sample from the standard Gaussian distribution on $\mathbb{R}^d$ and from a spherical *Student distribution*. For the Gaussian distribution a sample of size n will form a black cloud of radius $r_n \sim \sqrt{2\log n}$ with a halo on a scale of $1/r_n$; for the Student distribution the sample has no central black region. If one scales the sample of n observations from the density f in Theorem 12.1 by s_n, where $s_n \to \infty$ is defined by $\mathbb{P}\{Y \geq s_n\} = 1/n$, one obtains the Poisson point process on $\mathbb{R}^d$ with intensity $c/\|w\|^{t+d}$ as limit when $n \to \infty$. For heavy tails the distinction between the local and the global behaviour of the sample cloud is absent.

Let us now look at *flat* functions for heavy tails. Recall that a positive continuous function L on $\mathbb{R}^d$ is flat for Z if $L \circ \beta_H(w)/L \circ \beta_H(0) \to 1$ uniformly on compact w-sets in H_+, where $\beta_H^{-1}(Z^H) \Rightarrow W$. In our case this means that L is asymptotically constant on rings $R_s = \{2s \leq \|z\| \leq 4s\}$ for $s \to \infty$, since L is asymptotically constant on the intersection of the ring R_s with any halfspace tangent to the ball $B(0,s)$. So L is asymptotic to a spherically symmetric function $L_0(\|z\|)$, where $L_0\colon [0,\infty) \to (0,\infty)$ varies slowly in infinity. Flat functions do not extend the class of densities f introduced in Theorem 12.1 beyond asymptotic equality.

For roughening Lebesgue measure, the diameter of the sets F_p in the partition $\mathcal{F}$ should be $o(\|p\|)$ for $\|p\| \to \infty$. In particular, the counting measure on the lattice $\mathbb{Z}^d$ is a roughening of Lebesgue measure for all densities f in Theorem 12.1. We shall not enter into details here.

For random vectors Z there is a rich limit theory for Z^s, the vector Z conditioned to lie outside the ball sB of radius s, see Mikosch [2005]. The obvious limit relation is

$$Z^s/s \Rightarrow X, \quad s \to \infty. \tag{12.5}$$

In *polar coordinates* one may write $X = \Theta R$, with $R = \|X\| \geq 1$ and Θ a random element of the unit sphere. In Brozius & de Haan [1987] it is shown that R has a Pareto distribution, $P\{R > r\} = 1/r^\lambda$ for $r \geq 1$, for some parameter $\lambda > 0$. The vector Θ may have any distribution on the unit sphere, and is independent of R. If we condition X to lie in a halfspace H_θ supporting the unit ball at θ, and if Θ charges the halfspace $\{\theta \geq 0\}$ but not its boundary $\{\theta = 0\}$, we obtain a limit distribution ρ_θ for high risk scenarios Z^{H_n} with halfspaces $H_n = \{\theta_n \geq s_n\}$, where $s_n \to \infty$ and $\theta_n \to \theta$. High risk scenarios in different directions converge to different limit laws. These depend continuously on the direction θ provided Θ charges all halfspaces $\{\theta \geq 0\}$ and no hyperplanes $\{\theta = 0\}$.

In the case of heavy tails we have two models for describing extremal behaviour: the description in terms of high risk scenarios by conditioning on halfspaces, and the description (12.5), where the vector is conditioned to lie outside a ball whose radius s tends to infinity. The choice of the appropriate model depends on the sample set and the application; see Section 8.5. If the polar coordinates have a natural interpretation and the irregularities in the halo are clear and persistent for $s \to \infty$, and not due to elliptic level sets, then the more versatile second model is appropriate. If the irregularities fade out at infinity the simpler first model is relevant. In the first model the limit distribution is determined by one real parameter $\lambda > 0$, the tail exponent; in the second model there is an additional probability distribution on the unit sphere which has to be determined. We shall return to this issue in Section 16.

Example 12.2 (A continuous density f on the plane, which vanishes on the first and third quadrants, and which lies in the domain of a *Euclidean Pareto* law.). By a rotation through $\pi/4$, we may assume that f vanishes for $|y| \geq |x|$ and is positive for $|y| < |x|$. The function $h(x, y) = 0 \vee (|x| - |y|) \wedge 1$ has this property. Introduce the family of increasing ellipses

$$E_q : \left(\frac{x}{qe^q}\right)^2 + \left(\frac{qy}{e^q}\right)^2 < 1, \quad q \in [1, \infty).$$

The function

$$f_0(x, y) = h(x, y)/e^{3q}, \quad (x, y) \in \partial E_q, \ q \geq 1$$

is well defined on the complement of E_1, and is integrable, since E_n lies inside the disk D_n of radius $e^{4n/3}$ eventually, and $\sum |D_n|/e^{3n}$ converges. The function f_0 may be adapted on an ellipse E_{q_0} so as to become a continuous probability density f which vanishes on $A_0 = \{|x| \le |y|\}$ and is positive on $A_1 = \{|x| > |y|\}$.

The diagonal matrix $\beta_q = e^q \operatorname{diag}(q, 1/q)$ maps the unit disk B onto the ellipse E_q. Its inverse maps A_0 onto the thin wedge $\{|y| \ge q^2|x|\}$ and A_1 onto the complement of this wedge. Let Z have density f. For $q \ge q_0$ the density g_q of $W_q = \beta_q^{-1}(Z)$ is constant on the unit circle except for a neighbourhood of diameter $O(1/q^2)$ around the points $(0,1)$ and $(0,-1)$ where it is less. The image of E_{q+s} under β_q^{-1} is an ellipse with semi axes $e^s(1+s/q)$ and $e^s/(1+s/q)$, which is close to a disk with radius e^s for q large, and contained in a disk of radius $e^{4s/3}$ for all $s \ge 0, q \ge 1$. So

$$e^{3q} g_q(u,v) \to 1/r^3, \quad r^2 = u^2 + v^2$$

pointwise for $u \ne 0$, and also in $\mathbf{L}^1$ on $r \ge \varepsilon$ for any $\varepsilon > 0$. This shows that Z lies in the domain of attraction of the Euclidean Pareto law with exponent $\lambda = 1$. $\Diamond$

In the spirit of this example one may construct unimodal densities in $\mathcal{D}(\tau)$ which are not spherically symmetric, as in Example 7 in the Preview.

12.2 Bounded limit vectors. Densities with bounded support play only a minor role in risk theory. However the associated theory is simple and gives insight in the ideas underlying high risk limit theory, in particular in the role of rotund sets.

Let D be a rotund set in $\mathbb{R}^d$. If Z is uniformly distributed on D, then for any halfspace H intersecting D the high risk vector Z^H is uniformly distributed on the cap $D \cap H$. Now suppose the volume $|D \cap H|$ tends to zero. Since the curvature of the boundary is continuous, it is approximately constant on the cap when $|D \cap H|$ is small, and the high risk vector Z^H, properly normalized, will converge in distribution to a random vector W which is uniformly distributed on the parabolic cap $Q_+ = Q \cap H_+$, where Q is the paraboloid

$$Q = \{(u,v) \mid v \le 1 - u^T u\}\text{ß}\mathbb{R}^h \times \mathbb{R}. \tag{12.1}$$

Proposition 12.3. *Let D be a rotund set. There exist affine transformations β_H mapping H_+ onto H such that $\beta_H^{-1}(D) \to Q$ for $0 < |D \cap H| \to 0$, where Q is the paraboloid in (12.1).*

Proof. First assume D supports the halfspace $\{y \ge 1\}$ in $(0,1)^T$, and the upper boundary satisfies $1 - \partial^+ D(x) \sim x^T x$ for $x \to 0$. Define

$$\alpha_\sigma(u,v) = (\sqrt{\sigma}u, 1 - \sigma + \sigma v) = (x,y), \quad \sigma \in (0,1).$$

The upper boundary of the convex set $D_\sigma = \alpha_\sigma^{-1}(D)$ satisfies

$$1 - \partial^+ D_\sigma(u) = (1 - \partial^+ D(\sqrt{\sigma}u))/\sigma \to u^T u, \quad \sigma \to 0$$

uniformly on bounded u-sets. This proves the limit relation for halfspaces H of the form $\{y \geq t\}$. In order to establish the limit for halfspaces H diverging in an arbitrary direction we use Lemma 9.13, as in the proof of Proposition 9.12. □

As a corollary, we see that the *uniform distribution* on the parabolic cap Q_+ is a high risk limit law, whose domain of attraction contains the *uniform distribution* on any *rotund* set. The standard Poisson point process on the open *paraboloid* Q is the vague limit in law of the normalized sample clouds $\beta_{H_n}^{-1}(N_n)$, where H_n are closed halfspaces such that $|H_n \cap D| \sim |Q_+|/n$. The Poisson point process on Q thus describes the local texture at the edge of the sample cloud for large samples.

Let us briefly discuss some issues related to these limit results.

1) A density f on D which lies in the domain of attraction of the uniform distribution on Q_+ need not be constant. It is clear that f lies in the domain of the uniform distribution if f extends to a continuous function on the closure of D which is positive on the boundary. In general the function f need not have a continuous extension to the closure of D, even if it is C^1 on D. If $L = e^\lambda$ is positive on D and the partial derivatives of λ are bounded on D then the function L is integrable, and the density L/C will lie in the domain of the uniform distribution.

One can give precise conditions on the radial and tangential derivatives. Write $z = (1-s)\zeta$ with $\zeta \in \partial D$ and $s \in (0,1)$, and let H support $(1-s)D$ in z. Define $\lambda_R(t\zeta) = |d\lambda(t\zeta)/dt|$ and let $\lambda_T(z)$ be the maximum of $d\lambda(z+te)/dt$, $t = 0$, over all unit vectors e with $z + te \in \partial H$ as in Section 11.4. If

$$\lambda_R(z) = o(1/s), \quad \lambda_T(z) = o(1/\sqrt{s}), \quad s = s(z) = (1-n(z))_+, z \to \partial_D, \quad (12.2)$$

then L is flat for the uniform distribution on D in the sense that

$$L \circ \beta_H(w)/L \circ \beta_H(0) \to 1, \quad 0 < P\{Z \in H\} \to 0$$

holds uniformly on compact w-sets in Q_+. The condition on the radial derivative implies that on rays the function L is slowly varying as one approaches the boundary of D. One can show that L is integrable and that the density L/C lies in the domain of the uniform distribution on Q_+.

2) Let D be rotund and let γ denote the probability distribution on ∂D, determined by conditioning Lebesgue measure on $\mathbb{R}^d$ with respect to the gauge function n_D. For the unit ball γ is the uniform distribution on the unit circle. One can prove that the distribution γ on ∂D lies in the domain of a high risk limit law. (And so does $Ld\gamma/C$ for any continuous positive function L on ∂D). The limit law is singular with respect to Lebesgue measure on $\mathbb{R}^d$. It is the uniform distribution on the upper boundary of Q_+: the limit vector $W = (U, V)$ has the form $V = 1 - U^T U$, where U is uniformly distributed on the disk $\|u\| < 1$ in $\mathbb{R}^h$. The corresponding excess measure ρ is *Lebesgue measure* on $\mathbb{R}^h$ lifted to ∂Q by the map $u \mapsto (u, 1 - u^T u)$.

3) The uniform distribution on Q_+ and on the upper boundary of Q_+ are only two possible limit laws. Let s be the tent function on the rotund set D introduced in (12.2). If the random vector Z on D has density $f(z) \propto s(z)^{c-1}$ for some $c > 0$, then Z lies in the domain of attraction of the vector W on Q_+ with density $q(w)^{c-1}/C_c$, where $q(u, v) = 1 - u^T u - v$ measures the vertical distance from $w = (u, v)$ to the upper boundary of Q_+. This remains true if Z has density $\propto e^{\lambda} s^{c-1}$ where λ satisfies (12.2).

4) If λ satisfies (12.2), then e^{λ}/s lies in the domain of the uniform distribution on the upper boundary of Q_+ if the function is integrable.

Definition. The high risk limit distributions on Q_+ and on the upper boundary of Q_+ are called the *parabolic power laws*.

For rotund sets the behaviour in all boundary points is the same asymptotically. Now let us see what happens if D has a *vertex* in a point p_0.

Assume $D = (0, 1)^d$, the unit cube in $\mathbb{R}^d$, and $p_0 = 0$. Consider halfspaces $H = H_\xi = \{\xi \leq 1\}$ with $\xi = (x_1, \ldots, x_d) \in (1, \infty)^d$. The closure of $H \cap D$ is the convex hull Σ_ξ of the points $0, e_1/x_1, \ldots, e_d/x_d$, where $e_1, \ldots, e_d$ is the standard basis of $\mathbb{R}^d$. If Z is uniformly distributed over D, then Z^H is uniformly distributed over the simplex Σ_ξ. The linear transformation $\beta_\xi = \mathrm{diag}(1/x_1, \ldots, 1/x_d)$ maps the unit simplex Σ with vertices $0, e_1, \ldots, e_d$ onto Σ_ξ. Hence β_ξ^{-1} maps the high risk vector Z^{H_ξ} onto the vector W which is uniformly distributed on the unit simplex Σ. The excess measure ρ is *Lebesgue measure* on the open positive quadrant $(0, \infty)^d$. The connected symmetry group $\mathcal{G}$ consists of all positive diagonal linear maps. For any halfspace H with finite positive ρ mass, there exists an element $\sigma_H \in \mathcal{G}$ mapping Σ onto $H \cap [0, \infty)^d = \Sigma_\xi$ such that $d\rho^H = 1_H d\rho/\rho(H) = \sigma_H(d\rho_0)$ is the image of ρ_0, the uniform distribution on Σ. There are other Radon measures on $(0, \infty)^d$ which are semi-invariant under the group $\mathcal{G}$. These have density $x_1^{c_1} \ldots x_d^{c_d}$, with $c_i > -1$ for $i = 1, \ldots, d$. These measures, restricted to Σ, and normalized, are possible limit laws for Z^H when Z is a vector on $(0, 1)^d$ or $(0, \infty)^d$, and the halfspaces H have the form $\{\xi \leq 1\}$ with $\xi \in (0, \infty)^d$.

In this situation there is a finite-dimensional family of limit laws which holds for halfspaces whose normal points into the open negative orthant. One might speak of a *local limit law*. The associated measure ρ on $(0, \infty)$ has a large group of symmetries. The high risk scenarios Z^H all describe the behaviour of Z in the neighbourhood of one particular boundary point. In that respect the limit behaviour is not very interesting. On the other hand, the problem of describing the asymptotic behaviour of the convex hull of a sample of size n from the uniform distribution in a polygon in $\mathbb{R}^2$ has attracted considerable attention since Rényi & Sulanke [1963], see for instance Cabo & Groeneboom [1994]. The model also is of some interest for finance since prices are by nature positive.

There is a variant where the *vertex* has a different structure: D is the cap of a cone $C = \{\|u\| < v\}$. In this case ρ is Lebesgue measure and the symmetry group $\mathcal{G}$ is the *Lorentz group* acting on the cone C of future events in relativity theory. Let $q = v - \|u\|$ denote the vertical distance to the boundary ∂C. The densities q^{c-1} with $c > 0$ are also semi-invariant under $\mathcal{G}$. See also Section 15.3.

13 The multivariate GPDs

It is time now to introduce the complete class of *multivariate generalized Pareto distributions*, GPDs. This section is meant for reference rather than for detailed reading. Below we shall give for every dimension $d > 1$

1) the standard multivariate GPD's π_τ as a continuous one-parameter family;

2) the power families of the *Euclidean Pareto* and the parabolic power distributions in a simple form;

3) the associated spherical probability distributions μ_τ on $\mathbb{R}^h$; and

4) the excess measures ρ_τ with their symmetry groups.

It will be shown how projection onto lower dimensional subspaces acts on the limit laws π_τ and on distributions in their domains of attraction $\mathcal{D}(\tau)$. We shall discuss the role of spherical symmetry and independence.

We shall use the following notation in this section:

$$h = d - 1 \geq 1, \quad W = (U, V) \in H_+ = \mathbb{R}^h \times [0, \infty), \quad \nu = 1/|\tau| \text{ for } \tau \neq 0.$$

There are other candidates for the term multivariate generalized Pareto distribution, see Tajvidi [1995] and Balkema & Qi [1998].

13.1 A continuous family of limit laws. For the Gauss-exponential limit law the representation in 1) and 2) is the same, and the corresponding spherical probability distribution in 3) is the standard multivariate Gaussian distribution. The associated Radon measure ρ_0 has density $e^{-(u^T u/2+v)}$. A Gaussian density in dimension $d = 2, 3, 4, \ldots$ is determined by $5, 9, 14, \ldots$ parameters; a Gauss-exponential density on $H_+ ß\mathbb{R}^d$ by $4, 8, 13, \ldots$ parameters.

In the univariate case the exponential distribution forms the central distribution in the family of GPDs, linking the heavy tailed Pareto distributions and the power laws with finite upper endpoint. The multivariate situation is similar. In all cases the vertical component of the high risk limit vector $W = (U, V)$ has a univariate GPD. We shall use the shape parameter τ of this univariate law, see (5) in the Preview, to classify the multivariate distributions. This is the *Pareto parameter* of the distribution, and of the associated excess measure. The multivariate GPD has cylinder symmetry with respect to the vertical axis. This means that aside from the shape parameter the only other parameters are the two scale parameters for the horizontal and vertical components. For the standardized distributions π_τ we choose the vertical scale parameter so that the component V has the standard univariate GPD, G_τ in (5) in the Preview. The horizontal scale parameter is determined asymptotically for $\tau \to 0$ by the continuity condition at $\tau = 0$.

In the univariate case, the shape parameter τ varies over the whole real line; in the multivariate case the parameter τ varies over the set $[-2/h, \infty)$.

1) The Gauss-exponential distribution π_0 is the central term in a continuous one-parameter family of high risk limit distributions π_τ, $\tau \geq -2/h$, on the upper halfspace H_+ in $\mathbb{R}^d$, the standard multivariate GPDs. For $\tau > -2/h$ the distribution π_τ has a density $g_\tau(u, v)$ of the form

$$g_\tau(u,v) = \begin{cases} \left((1+\tau v)^2 + \tau u^T u\right)^{-1/2\tau - d/2} / C_\tau & \tau > 0, \\ e^{-(v + u^T u/2)}/(2\pi)^{h/2} & \tau = 0, \\ \left(1 + \tau v + \tau u^T u/2\right)_+^{-1/\tau - 1 - h/2} / C_\tau & -2/h < \tau < 0, \end{cases}$$

where the constants C_τ have the value

$$C_\tau = \begin{cases} (\nu\pi)^{h/2}\Gamma((\nu+1)/2)/\Gamma((\nu+1+h)/2) & \tau > 0, \\ (2\nu\pi)^{h/2}\Gamma(\nu - h/2)/\Gamma(\nu) & -2/h < \tau < 0. \end{cases}$$

For $\tau = -2/h$ the probability measure π_τ is singular. It lives on the parabolic cap

$$\{2v = h - u^T u\} \cap \{v \geq 0\}\text{ß}\mathbb{R}^d,$$

and projects onto the uniform distribution on the disk $\{u^T u < h\}$ in the horizontal coordinate plane. Thus in the case $\tau = -2/h$, one may write the limit vector as $W = (U, V)$, where U is uniformly distributed on the centered disk of radius $\sqrt{h}$ in $\mathbb{R}^h$, and the vertical coordinate $V = (h - U^T U)/2$ is a function of U.

For each $\tau \geq -2/h$ the vertical coordinate V has a standard univariate GPD with parameter τ, and the horizontal component U has a spherically symmetric density. Expectations exist for $\tau < 1$; *variances* for $\tau < 1/2$:

$$E(U,V) = \frac{1}{1-\tau}(0,1), \quad \mathrm{var}(U,V) = \frac{1}{1-\tau}\,\mathrm{diag}\left(1,\dots,1,\frac{1}{(1-\tau)(1-2\tau)}\right).$$

The expressions for g_τ for positive and negative parameter values τ differ. For $d > 1$ the family of multivariate GPDs is continuous in τ, but no longer analytic.

2) The positive and negative parameter values determine two power families of limit densities. In the calculations below we leave out the integration constants. These may be computed from the identities

$$\int_0^\infty \frac{r^{2a-1}dr}{(1+r^2)^{b+a}} = \int_0^1 r^{2a-1}(1-r^2)^{b-1}dr = \frac{1}{2}B(a,b) = \frac{\Gamma(a)\Gamma(b)}{2\Gamma(a+b)}. \quad (13.1)$$

For $\tau > 0$ the vector $(X, Y) = (U/\sqrt{\nu}, V/\nu)$ has a Euclidean Pareto density proportional to

$$1/((1+y)^2 + x^T x)^{(\nu+d)/2}.$$

This density was treated in Section 8.

Let the vector S in $\mathbb{R}^h$ have density proportional to $1/(1+s^T s)^{(\nu+d)/2}$. Then $(1+y)S$ has density proportional to $1/((1+y)^2+s^T s)^{(\nu+d)/2}$. This is the conditional density of X given $Y = y$. So we see that (X, Y) for $\tau > 0$ is distributed like $((1+Y)S, Y)$, where Y is independent of S.

For $\tau \in (-2/h, 0)$ the vector $(X, Y) = (U/\sqrt{2\nu}, V/\nu)$ has a parabolic power density proportional to

$$(1 - y - x^T x)_+^{q-1}, \quad \nu = q + h/2.$$

Here $P\{Y \geq y\} = (1-y)_+^\nu$ and X has density proportional to $(1 - x^T x)_+^q$.

Let $S \in \mathbb{R}^h$ have a spherical beta density, proportional to $(1 - s^T s)_+^{q-1}$, $q > 0$. Then $\sqrt{(1-y)}S$ has density proportional to $(1 - y - s^T s)_+^{q-1}$ for $0 \leq y < 1$. So (X, Y) for $-2/h < \tau < 0$ is distributed like $(\sqrt{(1-Y)}S, Y)$ with S independent of Y. This also holds for the boundary case $\tau = -h/2$, where S is uniformly distributed over the boundary of the unit disk in $\mathbb{R}^h$, and X is uniformly distributed over the disk. Note that S and X both have a *spherical beta density*, but with different exponents.

13.2 Spherical distributions. 3) Introduce the family of spherical probability distributions μ_τ on $\mathbb{R}^h$ with densities proportional to

$$1/(1 + \tau s^T s)^{(h+1+\nu)/2}, \ \tau > 0; \quad (1 + \tau s^T s/2)_+^{\nu-1-h/2}, \ \tau \in (-2/h, 0).$$

By continuity μ_0 is the standard Gauss distribution on $\mathbb{R}^h$, and $\mu_{-2/h}$ is the uniform distribution on the sphere of radius $\sqrt{h}$ in $\mathbb{R}^h$. Note that μ_τ for $\tau > 0$ is *not* the spherical Student distribution with ν degrees of freedom. (It lives on $\mathbb{R}^h$, not $\mathbb{R}^d$.)

Theorem 13.1 (Structure). *Let S have distribution μ_τ on $\mathbb{R}^h$, and let V have a standard univariate GPD on $[0, \infty)$ with df G_τ, as defined in Section 13.1. Assume that S and V are independent. Set $W = ((1 + \tau V)S, V)$ for $\tau \geq 0$ and write $W = (\sqrt{(1 + \tau V)}S, V)$ for $-2/h \leq \tau \leq 0$. Then W has the standard multivariate GPD π_τ on H_+.*

Proof. For $\tau > -2/h$ this follows from the density of the conditional distribution of U given $V = v$ where (U, V) has density g_τ above, see the arguments under 2). For $\tau = -2/h$ it follows from the fact that U is distributed on a sphere of radius $\sqrt{2\nu - v}$ in $\mathbb{R}^h$ if $V = v$, just as $\sqrt{(1 + \tau v)}S$. □

Theorem 13.2 (Projection). *Let $W = (U_1, \dots, U_h, V)$ have distribution π_τ on H_+, and let $S = (S_1, \dots, S_h)$ have distribution μ_τ on $\mathbb{R}^h$. For $m = 0, \dots, h-1$ the vector $(U_1, \dots, U_m, V)$ has distribution π_τ on $\mathbb{R}^m \times [0, \infty)$, and $(S_1, \dots, S_m)$ has distribution μ_τ on $\mathbb{R}^m$.*

Proof. The densities proportional to $1/(1 + s^T s)^{(\nu+d)/2}$ or to $(1 - s^T s)_+^{\nu-1-h/2}$ are stable under orthogonal projections as is seen by integrating out the variable s_1. This then also holds for the distributions μ_τ for $\tau \neq 0$ and $\tau > -2/h$, and (by continuity) also for the two exceptional values of τ. Now apply Theorem 13.1 to obtain the result for the GPDs. □

13.3 The excess measures and their symmetries. 4) In Section 8 we extended the Gauss-exponential limit law on H_+ to a Radon measure on the whole space $\mathbb{R}^d$. The other high risk limit laws have similar extensions to infinite Radon measures. If one adapts the normalization so as to achieve maximal simplicity, one obtains the measures:

ρ_τ on $\mathbb{R}^d \setminus \{0\}$, with density $1/r^{\lambda+d}$ for the Euclidean Pareto limit laws, $\tau = 1/\lambda > 0$;

ρ_0 on $\mathbb{R}^d$, with density $e^{-v} e^{-u^T u/2}$ for the Gauss-exponential limit law;

ρ_τ on the paraboloid $Q = \{v + u^T u < 0\}$, with density $q^{\lambda-1}$ for the parabolic power laws, where $q(u, v) = -(v + u^T u)$ for $\tau = -1/(h/2 + \lambda) < 0$; and

$\rho_{-2/h}$ on ∂Q is *Lebesgue measure* on $\mathbb{R}^h$ lifted to the parabolic surface $v + u^T u = 0$ for the parabolic power law with $\tau = -2/h$.

Recall that the *symmetry group* $\mathcal{G}$ of a measure μ is the set of all $\sigma \in \mathcal{A}$ for which there exists a constant $c_\sigma > 0$ such that $\sigma(\mu) = c_\sigma \mu$. In Section 8 it was shown that the symmetry group of the measure ρ on $\mathbb{R}^d$ with density $e^{-v} e^{-\|u\|^2/2}$ is generated by the vertical translations $(u, v) \mapsto (u, v + t)$, $t \in \mathbb{R}$, the orthogonal transformations which leave the points on the vertical axis in their place, and the parabolic translations (8.6). The symmetry group for the parabolic power measures with densities $q^{\lambda-1}$ on the paraboloid Q is the same except that we replace the vertical translations by the linear maps $(u, v) \mapsto (cu, c^2 v)$, $c > 0$, which map Q onto itself. This is also the symmetry group of the singular measure on ∂Q. The symmetry group of the spherical Pareto measures with densities $1/r^{d+1/\tau}$ on $\mathbb{R}^d \setminus \{0\}$ is different. It is generated by the group of orthogonal transformations on $\mathbb{R}^d$ together with the scalar maps $w \mapsto cw$, $c > 0$.

Proposition 13.3. *Let $\mathcal{S}$ be the symmetry group of the infinite Radon measure ρ_τ associated with a Pareto-parabolical high risk limit law π_τ on H_+. For $\tau = 0$ let $J_0 = H_+$, for $\tau > 0$ let $J_0 = \{v \geq 1\}$, and for $\tau < 0$ let $J_0 = \{v \geq -1\}$. For each closed halfspace H with $0 < \rho_\tau(H) < \infty$ there is a symmetry σ mapping H onto J_0. The probability measures $d\rho_\tau^H = 1_H d\rho_\tau / \rho_\tau(H)$ have the shape of π_τ.*

Proof. As for Lemma 8.1. □

13.4 Projection. For coordinatewise maxima there is an obvious projection theorem:

$$(Z_1, \dots, Z_d) \in \mathcal{D}^\vee(W_1, \dots, W_d) \Rightarrow (Z_1, \dots, Z_m) \in \mathcal{D}^\vee(W_1, \dots, W_m)$$

for $m = 1, \dots, d-1$. A similar result holds for the domains $\mathcal{D}(\tau)$ of the multivariate GPDs: If $(Z_1, \dots, Z_d) \in \mathcal{D}(\tau)$, then $(Z_1, \dots, Z_m) \in \mathcal{D}(\tau)$ for $m = 1, \dots, d-1$.

Theorem 13.4 (Projection Theorem). *Let $\Gamma\colon \mathbb{R}^d \to \mathbb{R}^m$ be an affine surjection, $\Gamma(z) = Az + b$, with $b \in \mathbb{R}^m$, and the m rows of the matrix A independent. If $Z \in \mathcal{D}(\tau, \mathbb{R}^d)$, then $\Gamma(Z) \in \mathcal{D}(\tau, \mathbb{R}^m)$.*

Proof. Let $W_H = \alpha_H^{-1}(Z^H) \Rightarrow W$ where α_H maps H_+ onto H. By cylinder symmetry of the limit distribution, we may replace α_H by $\alpha_H \circ R_H$ for any rotation R_H around the vertical axis, $R_H \in \mathrm{O}(\mathbb{R}^h)$. We need only consider high risk scenarios Z^H for halfspaces $H = \Gamma^{-1}(J)$ for halfspaces $J \text{ß} \mathbb{R}^m$. Let $K = \{A = 0\}$ be the kernel of A. Then H is a union of translates of K, and so is ∂H. We may choose R_H such that $\tilde{\alpha}_H = \alpha_H \circ R_H$ maps $\{0\} \times \mathbb{R}^k \times \{0\} \text{ß} \mathbb{R}^{m-1} \times \mathbb{R}^k \times \mathbb{R}$ into a translate of the $k = d - m$-dimensional kernel K. This allows us to regard Γ as a coordinate projection in terms of the coordinates introduced by $\tilde{\alpha}_H$. One may write $\Gamma \circ \tilde{\alpha}_H = \beta_J \circ p$ for an affine transformation β_J on $\mathbb{R}^m$ where p is the projection $(u_1, \dots, u_h, v) \mapsto (u_1, \dots, u_{m-1}, v)$:

$$\begin{array}{ccc}
\mathbb{R}^{m-1} \times \mathbb{R}^k \times \mathbb{R} & \xrightarrow{\tilde{\alpha}_H} & \mathbb{R}^d \\
\downarrow p & & \downarrow \Gamma \\
\mathbb{R}^{m-1} \times \mathbb{R} & \xrightarrow{\beta_J} & \mathbb{R}^m .
\end{array}$$

It now follows that $\beta_J^{-1}(\Gamma(Z)) = p(\tilde{\alpha}_H^{-1}(Z)) \Rightarrow p(W) = (U_1, \dots, U_{m-1}, V)$ for $H \to \partial$. □

13.5 Independence and spherical symmetry. Independence and spherical symmetry are two powerful instruments for creating multivariate distributions. Centered Gaussian distributions with iid components are the only spherical distributions with independent components. The theory of coordinatewise maxima is able to accommodate both spherical symmetry and independence. However the *exponent measure* corresponding to the independent part lives on the boundary. For $\tau \le 0$ this boundary has coordinates equal to $-\infty$. Compare Example 7.4.

In geometric extreme value theory spherical symmetry plays a prominent role. A rather trivial form of the multivariate theory states that for any density f_0 on $[0, \infty)$ in

$\mathcal{D}^+(\tau)$ the corresponding spherical density f (where $\|Z\|$ has density f_0) belongs to $\mathcal{D}(\tau)$. This chapter tries to answer the question: How far can one relax the condition of spherical symmetry, while retaining the desired limit behaviour for the high risk scenarios?

Theorem 13.5. *If $Z \in \mathcal{D}(\tau)$ has independent components, then Z is Gaussian* (*and* $\tau = 0$).

Proof. It suffices to prove that the component Z_1 is Gaussian. Write $Z = (X, Y)$, and let $Z^t = (X, Y^t)$ be the high risk scenario for the horizontal halfspace $H^t = \{y \geq t\}$. Then $Y \in \mathcal{D}(\tau)$, i.e. $(Y^t - t)/a_t \Rightarrow V$ where V has a GPD G_τ on $[0, \infty)$. Let $\alpha_t(u, v) = (u, t + a_t v)$. Then $\alpha_t^{-1}(Z^t) \Rightarrow (X, V)$ with V and X independent. So (X, V) has a multivariate GPD. By independence $\tau = 0$. Hence X is Gaussian. □

IV Thresholds

This chapter treats exceedances over horizontal and elliptic thresholds.

The first four sections of this chapter treat exceedances. There are two sections on exceedances over linear thresholds, and two on exceedances over elliptic thresholds. In both cases the first section treats the general theory, and the second investigates specific cases. For exceedances over horizontal thresholds the link to the theory of multivariate GPDs in Chapter III is clear: rather than letting halfspaces drift away arbitrarily, we now assume divergence in a fixed direction. For convenience we assume the halfspaces to be horizontal: $H_t = \{y \geq t\}$ where y denotes the vertical coordinate. Exceedances over elliptic thresholds are associated with heavy tails. The Euclidean Pareto measures in Chapter III have spherically symmetric densities $c/\|w\|^{d+1/\tau}$. The limit measures in the present chapter need not have such a simple form.

Both in the case of horizontal halfspaces and in the case of complements of open ellipsoids we condition on a decreasing family of closed sets, with vanishing probability, and ask for a limit vector to describe the asymptotic behaviour of the conditional distributions. Both theories may be regarded as an alternative to the limit theory for coordinatewise maxima of random vectors in $[0, \infty)^d$. One may for instance choose the vertical axis along the diagonal, and condition on $Y = \eta(Z) = Z_1 + \cdots + Z_d \geq y$ for $y \to y_\infty$; or one may choose an increasing sequence of centered coordinate ellipsoids E_n, and condition on $Z \notin E_n$.

In both theories there is a one-dimensional group of affine transformations γ^t, $t \in \mathbb{R}$, acting on the excess measure

$$\gamma^t(\rho) = e^t \rho, \quad t \in \mathbb{R}.$$

For exceedances over horizontal thresholds the transformations γ^t map horizontal halfspaces into horizontal halfspaces, $\gamma^t \in \mathcal{A}^h$, and the corresponding group of positive affine transformations $\tilde{\gamma}^t$ on the vertical axis is precisely the symmetry group of $\tilde{\rho}$, the vertical component of the excess measure ρ. For exceedances over elliptic thresholds the γ^t, $t > 0$, are linear expansions. In the limit theory for coordinatewise maxima the symmetries γ^t were CATs, affine transformations whose linear part is diagonal.

The one-parameter symmetry groups γ^t, $t \in \mathbb{R}$, allow a complete classification of the limit distributions. For exceedances over horizontal thresholds the excess measure ρ is determined by a probability measure ρ^* on $\mathbb{R}^h$, and the symmetry group γ^t in $\mathcal{A}^h$; for exceedances over elliptic thresholds ρ is determined by a probability measure ρ^* on the unit sphere, in appropriate coordinates, and a linear expansion group γ^t.

There are differences too.

1) For exceedances over elliptic thresholds we are able to give a constructive characterization of the domain of attraction of excess measures with a continuous positive density.

2) For exceedances over horizontal thresholds convergence of the high risk scenarios entails convergence of the sample clouds.

3) Horizontal thresholds are easier to handle since there is only one decreasing family of horizontal halfspaces.

4) For elliptic thresholds convergence of convex hulls holds automatically; for horizontal thresholds no simple conditions are known.

The last section of this chapter offers background material. It is devoted to regular variation of affine transformations, and excess measures on $\mathbb{R}^d$. It contains subsections on the Jordan form of a linear transformation, on Lie groups, and on the Meerschaert Spectral Decomposition Theorem.

14 Exceedances over horizontal thresholds

14.1 Introduction. Often one has a good idea of the direction in which *risk* is located. One may even have a variate which measures loss, such as for instance minus the value of one's portfolio. In this section we assume that risk occurs at large values of the vertical coordinate (denoted by y, v, or η). So we are interested in high risk scenarios for *horizontal* halfspaces

$$H^y = \{\eta \geq y\}, \quad y \uparrow y_\infty,$$

where y_∞ denotes the upper endpoint of the distribution of the vertical component $Y = \eta(Z)$ of the vector Z describing the state of the system. The vertical component measures loss, or discomfort, or some other quantity of interest; the horizontal component $X \in \mathbb{R}^h$ of $Z = (X, Y)$ contains additional information. By ignoring the horizontal coordinate one is back in the univariate theory of exceedances discussed in Section 6. We are interested in the asymptotic behaviour of the *high risk scenarios*

$$Z^y = Z^{H^y}, \quad y \uparrow y_\infty.$$

Do there exist affine normalizations α_y such that

$$\alpha_y^{-1}(Z^y) \Rightarrow W, \quad \mathbb{P}\{Y \geq y\} \to 0+, \tag{14.1}$$

where we assume W to be non-degenerate, and to live on a horizontal halfspace $J_0 = \{v \geq j_0\}$, for instance the upper halfspace H_+, and we assume that α_y maps J_0 onto H^y.

Definition. $\mathcal{A}^h$ is the group of affine transformations mapping horizontal halfspaces into horizontal halfspaces. The matrix representation is given in (14.3) below.

Each $\alpha \in \mathcal{A}^h$ determines a transformation of the vertical coordinate, $\tilde{\alpha}$. The map $\alpha \mapsto \tilde{\alpha}$ is a homomorphism of the group $\mathcal{A}^h$ onto the group $\mathcal{A}^+$ of positive affine transformations $y \mapsto ay + b, a > 0$, on $\mathbb{R}$. It follows that (14.1) implies convergence of the vertical coordinate:

Proposition 14.1. *Suppose* (14.1) *holds for* $Z = (X, Y)$ *and* $W = (U, V)$ *in* $\mathbb{R}^{h+1}$ *with* $\alpha_t(H_+) = H^t$. *Then* $\tilde{\alpha}_t^{-1}(Y^{[t,\infty)}) \Rightarrow V$, *and* V *has a GPD on* $[0, \infty)$.

Definition. The vector $Z = (X, Y) \in \mathbb{R}^{h+1}$ lies in the *domain of horizontal attraction* of $W = (U, V)$ if (14.1) holds, with α_t mapping the standard horizontal halfspace $J_0 = \{v \geq j_0\}$ onto H^t, and if the distribution of W is non-degenerate. Notation $Z \in \mathcal{D}^h(W)$.

Proposition 14.1 allows us to apply the univariate theory to the vertical coordinate Y. The vertical component V of the limit vector has a GPD. The shape parameter τ of the df of V will be used to classify the multivariate limit distributions. The theory of exceedances over horizontal thresholds may be regarded as a refinement of the univariate theory for exceedances. The univariate theory describes the distribution of Y, given that Y exceeds a threshold y, for $y \to y_\infty$; the multivariate theory describes the distribution of the state of the system, $Z = (X, Y)$, under the same conditions. We shall see below that, as in the univariate case, there exists an infinite Radon measure ρ such that

$$\rho_y = \alpha_y^{-1}(\pi)/\mathbb{P}\{Y \geq y\} \to \rho, \quad y \to y_\infty$$

holds weakly on all horizontal halfspaces J on which ρ is finite, where π denotes the distribution of Z. The projection of ρ on the vertical coordinate, $\tilde{\rho} = \eta(\rho)$, is precisely the univariate Radon measure which extends the distribution of V. The measure $\tilde{\rho}$ is a univariate excess measure. It lives on an unbounded interval

$$(j_*, j^*) = \{\tilde{\gamma}^t(j_0) \mid t \in \mathbb{R}\}, \quad J_0 = \mathbb{R}^h \times [j_0, \infty). \tag{14.2}$$

The $\tilde{\gamma}^t$ in $\mathcal{A}^+$ satisfy $\tilde{\gamma}^t(\tilde{\rho}) = e^t\tilde{\rho}$. In appropriate coordinates one of the following holds:

$$\tilde{\rho}[y, \infty) = e^{-y}, \ y \in \mathbb{R}; \quad = 1/y^\lambda, \ y > 0; \quad = |y|^\lambda, \ y < 0,$$

as in (6.8), where $\lambda = 1/|\tau|$ for $\tau \neq 0$. This is as it should be. The normalized sample clouds from the distribution π of Z converge to the Poisson point process N with mean measure ρ; and the normalized sample clouds of the vertical coordinate converge to the vertical coordinate $\eta(N)$, a Poisson point process with mean measure $\tilde{\rho}$.

One could regard N as a *marked point process* (with the distribution of the horizontal mark dependent on the vertical coordinate). We prefer to think of the point process N in geometric terms. One of the questions then is what happens if there is a slight departure from the horizontal. Recall that the limit measure ρ is *sturdy* if $\rho(J_n) \to \rho(J_0) = 1$ for halfspaces $J_n \to J_0$. For sturdy measures $1_{J_n} d\rho \to 1_{J_0} d\rho$ weakly whenever $J_n \to J_0$. The measure ρ is sturdy precisely if the vertical direction η lies in the intrusion cone Δ of ρ, see Section 5.4. If ρ is not sturdy there exists a sequence $J_n \to J_0$ such that $\rho(J_n) = \infty$ for all $n \geq 1$, and ρ is called *flimsy*.

Definition. Suppose $\rho_n = n\pi_n \to \rho$ vaguely, where we set $\pi_n = \alpha_{y_n}^{-1}(\pi)$ with $\mathbb{P}\{Y \geq y_n\} \sim 1/n$ and $\alpha_t(J_0) = \{y \geq t\}$. The probability measure π (or Z) is *steady* if $1_{J_n} d\rho_n \to 1_{J_0} d\rho$ weakly whenever $J_n \to J_0$.

For steady π the normalized sample cloud converges weakly to the limiting Poisson point process on halfspaces sufficiently close to horizontal. The vertical coordinate η lies in the *convergence cone* Γ of $n\pi_n$.

The main objective of this section is to determine the limit laws for exceedances over horizontal thresholds. A complete classification is possible since limit distributions extend to excess measures. The Extension Theorem 14.12 allows us to replace the limit relation (14.1) by

$$e^t \beta(t)^{-1}(\pi) \to \rho \text{ weakly on } J = \mathbb{R}^h \times [v, \infty), \quad t \to \infty,\ \rho(J) < \infty,$$

where $\beta\colon [0, \infty) \to \mathcal{A}^h$ is continuous and varies like γ^t as defined in (12) in the Preview. This is the main result of the present section. The proof is technical, and the reader is advised to skip it and proceed directly to Section 14.6. Section 14.9 lists the limit laws in $\mathbb{R}^3$.

Section 14.10 gives conditions under which convergence is preserved when some of the coordinates of the horizontal components of the vector $Z = (X, Y)$ are deleted. Projection allows us to reduce conditions for sturdiness of the limit measure, and spectral stability, to lower dimensional limit measures.

A summary of the main results may be found in Section 14.13. The two final subsections treat the domains of attraction for limit distributions for exceedances over horizontal thresholds. Probability distributions in the domains of attraction may be described as perturbations of typical distributions. Typical distributions have the property that the conditional distributions of the horizontal component, given the vertical component, all have the same shape. A look at Section 18.1 on multivariate regular variation may be helpful at this point.

14.2 Convergence of the vertical component. We start by looking at the vertical component, the backbone of the vector $Z = (X, Y)$ in $\mathcal{D}^h(\rho)$.

For exceedances over horizontal thresholds the normalizations are transformations $\alpha \in \mathcal{A}^h$. These map horizontal halfspaces into horizontal halfspaces. The associated matrices of size $1+h+1$ and $1+1$,

$$\alpha = \begin{pmatrix} 1 & 0 & 0 \\ p & A & q \\ b & 0 & a \end{pmatrix}, \quad \tilde{\alpha} = \begin{pmatrix} 1 & 0 \\ b & a \end{pmatrix}, \tag{14.3}$$

define the affine transformation $\alpha\colon (x, y) \mapsto (p + Ax + qy, b + ay)$ on $\mathbb{R}^{h+1}$, and the positive affine transformation $\tilde{\alpha}\colon y \mapsto b + ay$ acting on the vertical coordinate. Note that α maps the vertical axis into the line $p + \mathbb{R}q$. The blocked $1+h+1$-matrices are convenient for explicit computations. Thus it is clear that $\alpha \mapsto \tilde{\alpha}$ is a homomorphism from $\mathcal{A}^h$ into $\mathcal{A}^+$: If $\gamma = \alpha\beta$ then $\tilde{\gamma} = \tilde{\alpha}\tilde{\beta}$.

Definition. The parameter τ of the limit GPD of the vertical coordinate is the *Pareto parameter* of the multivariate limit distribution.

Before proceeding, let us note that Proposition 14.1 yields information about the limit theory of multivariate *GPD*s developed in Chapter III. It gives a qualitative description of the tails of distributions in the domains of attraction of multivariate GPDs. Assume $Z \in \mathcal{D}(\tau)$. The halfspace H then is allowed to diverge in any direction. But every direction ξ may be regarded as the vertical direction. So ξZ lies in the domain of attraction $\mathcal{D}^+(\tau)$ of the GPD G_τ for each non-zero linear functional ξ on $\mathbb{R}^d$. If Z has heavy tails then $t^s \mathbb{P}\{\xi Z \geq t\}$ vanishes for $s < 1/\tau$ when $t \to \infty$, and diverges to ∞ for $s > 1/\tau$. If Z has light tails, all moments of $\|Z\|$ are finite. If τ is negative, Z has bounded support.

In order to be able to speak of limit laws we had to impose a regularity *condition on the boundary* of the distribution π of the vector Z: hyperplanes tangent to the convex support should not carry positive mass, see (8.3). Here we can say more.

Proposition 14.2. *Suppose π lies in the domain $\mathcal{D}(\tau)$ of a multivariate GPD. Then*

$$\pi(\partial H)/\pi(H) \to 0, \quad \pi(H) \to 0+.$$

Proof. Otherwise the limit law would have positive mass on the horizontal hyperplane ∂H_+ and the vertical component V of the limit vector would have mass in the origin. This is not possible since V has a GPD by Proposition 14.1. □

14.3* A functional relation for the limit law. We now start with the proof of the Extension Theorem, which will be formulated in Section 14.6.

It is convenient to use risk as parameter, and write Z_q for the high risk scenario Z^t with $\mathbb{P}\{Z \in H^t\} = q$. Similarly we write W_q for the limit vector W conditioned to lie in the horizontal halfspace J for which $\mathbb{P}\{W \in J\} = q$. Observe that $(W_q)_p = W_{pq}$ for $p, q \in (0, 1]$.

The vectors W_q are well defined since the df G of the vertical component is continuous on $\mathbb{R}$ and strictly increasing on the interval $\{0 < G < 1\}$. It helps to assume in first instance that the df F of Y also has these properties. From the univariate theory in Section 6 it is known that the distribution tail $1-F$ is asymptotic to a continuous function $1-F_0$ on $\mathbb{R}$ which is strictly decreasing on $\{0 < F < 1\}$. We shall write

$$q = q(t), \quad t = t(q), \quad q(t) = 1 - F_0(t). \tag{14.4}$$

We begin with a simple result on weak convergence of high risk scenarios. Recall that $G^{\leftarrow}\colon (0,1) \to \mathbb{R}$ denotes the left-continuous inverse of the univariate df G.

Lemma 14.3. *Let* $W_n \Rightarrow W = (U, V) \in \mathbb{R}^{h+1}$. *Let* V *have df* G. *Let* $H, H_1, H_2, \ldots$ *be horizontal halfspaces. Suppose* $q_n := \mathbb{P}\{W_n \in H_n\} \to \mathbb{P}\{W \in H\} = q \in (0,1)$, *where* $H, H_1, H_2, \ldots$ *are horizontal halfspaces. If* $G^{\leftarrow}$ *is continuous in* q *then* $W^{H_n} \Rightarrow W^H$.

Proof. First note that the dfs G_n of the vertical component of W_n converge weakly to G, and hence $G_n^{\leftarrow} \to G^{\leftarrow}$ weakly. The continuity in q ensures that $t_n \to t$, where we write $H_n = \mathbb{R}^h \times [t_n, \infty)$ and $H = \mathbb{R}^h \times [t, \infty)$. It also ensures that $\mathbb{P}\{V = t\} = 0$. So $1_{H_n} \to 1_H$ holds W-a.s. For any bounded continuous function φ we then have $\mathbb{E}(\varphi 1_{H_n})(W_n)/q_n \to \mathbb{E}(\varphi 1_H)(W)/q$. □

Now suppose W is a limit vector for exceedances over horizontal thresholds. The lemma implies that W_q is distributed like $\gamma(W)$ for some $\gamma \in \mathcal{A}^h$. Here is a more general result.

Proposition 14.4. *Suppose* $\alpha_{t_n}^{-1}(Z^{t_n}) \Rightarrow W$ *and* $\mathbb{P}\{Y \geq t_n\} = q_n \to 0+$. *Let* G *be the df of the vertical component of* W. *Suppose* $G^{\leftarrow}$ *is continuous in* $q \in (0,1)$. *If* $\mathbb{P}\{Y \geq s_n\} \sim q q_n$ *then* $\alpha_{t_n}^{-1}(Z^{s_n}) \Rightarrow W_q$. *If also* $\alpha_{s_n}^{-1}(Z^{s_n}) \Rightarrow W$ *then*

1) *The sequence* $(\alpha_{t_n}^{-1}\alpha_{s_n})$ *is relatively compact.*
2) *For any limit point* γ *of this sequence,* $\gamma(W)$ *is distributed like* W_q.
3) *For any such* γ *there exist symmetries* σ_n *of* W *in* $\mathcal{A}^h$ *such that*

$$\alpha_{t_n}^{-1}\alpha_{s_n}\sigma_n \to \gamma.$$

Proof. Convergence to W_q follows from the lemma with $W_n = \alpha_{t_n}^{-1}(Z^{t_n})$ and $W_n^{H_n} = \alpha_{t_n}^{-1}(Z^{s_n})$ with $H_n = \alpha_{t_n}^{-1}(H^{s_n})$. The limit vectors W and W_q are of the same type by the Convergence of Types Theorem; see the Preview. The CTT yields the three conclusions above. □

14.4* Tail self-similar distributions. We shall now first investigate random vectors W with the property that for certain $q \in (0,1)$ there exist $\gamma \in \mathcal{A}^h$ such that the high risk scenario W_q is distributed like $\gamma(W)$.

Definition. A random vector W on H_+ with distribution ρ_0 is *tail self-similar* for $\gamma \in \mathcal{A}^h$ if the high risk scenario W^H for $H = \gamma(H_+)$ is distributed like $\gamma(W)$, and if $q(\gamma) := \mathbb{P}\{W \in \gamma(H_+)\} \in (0, 1)$.

Example 14.5. Let $W = (U, V) \in \mathbb{R} \times [0, \infty)$. Suppose V is discrete with a *Pascal* distribution on the non-negative integers: $P\{V \geq t\} = q^{[t]}$, where $[t]$ denotes the integer part of t. Given $V = n$ let U be normal with unit variance and expectation n^2. Then W is tail self-similar for $\gamma\colon (u, v) \mapsto (u + 2v + 1, v + 1)$. Observe that $\alpha_t^{-1}(W^t) \Rightarrow W$ if we choose $\alpha_t = \gamma^{-[t]}$ for $t \geq 0$. Yet $W \notin \mathcal{D}_h(W)$. ◊

We claim that a tail self-similar distribution ρ_0 extends to an infinite Radon measure ρ which satisfies

$$\gamma(\rho) = \rho/q, \quad q = q(\gamma). \tag{14.5}$$

The construction of ρ is simple. The difference $S = H_+ \setminus \gamma(H_+)$ is a horizontal slice, and $d\mu = 1_S d\rho_0$ is a measure of mass $1 - q$. The restriction of ρ_0 to the slice $S_1 = \gamma(S) = \gamma(H_+) \setminus \gamma^2(H_+)$ is $q\gamma(\mu)$, and similarly for the slices $S_2, S_3, \ldots,$ with $S_k = \gamma^k(S)$. Now define

$$\rho = \sum_{k=-\infty}^{\infty} q^k \gamma^k(\mu) \quad \text{on} \quad \bigcup_{k=-\infty}^{\infty} \gamma^k(S) = \mathbb{R}^h \times (j_*, j^*). \tag{14.6}$$

Here (j_*, j^*) is the union of the sets $\tilde{\gamma}^{-n}([0, \tilde{\gamma}(0)))$, where $\tilde{\gamma}$ is the positive affine transformation on $\mathbb{R}$ associated with γ in (14.3). Then ρ is an infinite measure, which is finite on $\mathbb{R}^h \times [v, \infty)$ for any $v > j_*$, and ρ restricted to H_+ is ρ_0. The projection $\tilde{\rho}$ of ρ on the vertical axis satisfies $\tilde{\gamma}(\tilde{\rho}) = \tilde{\rho}/q$. By construction (14.5) holds.

It is not obvious that a different tail self-similarity, $\gamma_1(W) = W^{H_1}$, for the vector W will yield the same extension ρ. We shall return to this issue in Section 18.10. In order to understand tail self-similar distributions we need to understand the symmetry group of ρ.

The symmetry group of ρ in $\mathcal{A}^h$ is the set $\mathcal{G}$ of all $\gamma \in \mathcal{A}^h$ for which there exists a constant $q = q(\gamma)$ such that $\gamma(\rho) = \rho/q$. The set $\mathcal{G}$ is a group, and the map $\gamma \mapsto q(\gamma)$ is a homomorphism from $\mathcal{G}$ into the multiplicative group of the positive reals, $(0, \infty)$. We shall show that there are only two options: Either the image $q(\mathcal{G})$ is a discrete subgroup of the form $\{p^k \mid k \in \mathbb{Z}\}$ for some $p \in (0, 1)$, or $q(\mathcal{G}) = (0, \infty)$. It is the latter case in which we are interested. We shall see that in that case the vertical component V of the vector W has a GPD on $[0, \infty)$ and ρ is an excess measure. At the end of this section we give more details on the structure of the group $\mathcal{G}$, its relation to the group $\mathcal{S}_h$ of symmetries of ρ_0 in $\mathcal{A}^h$, and the choice of γ in (14.5), which need not be unique.

Let Γ be the set of all $\gamma \in \mathcal{A}^h$ for which $W^{\gamma(H_+)}$ is distributed like $\gamma(W)$. Then

$$\gamma(d\rho_0) = 1_{\gamma(H_+)} d\rho_0 / q(\gamma), \quad q(\gamma) = \rho_0(\gamma(H_+)). \tag{14.7}$$

We list some simple properties of the set Γ.

Lemma 14.6. *The following hold:*

1) *If* $\gamma \in \Gamma$ *and* HßH_+ *has positive mass, then* $\gamma(1_H d\rho_0) = 1_{\gamma(H)} d\rho_0/q(\gamma)$.
2) Γ *contains the group* $\mathcal{S}_h$ *of symmetries of* ρ_0 *in* $\mathcal{A}^h$.
3) *If* $\gamma_1, \gamma_2 \in \Gamma$, *then* $\gamma = \gamma_1\gamma_2 \in \Gamma$ *and* $q(\gamma) = q_1 q_2$.
4) *If* $\gamma_1, \gamma_2 \in \Gamma$ *and* $q_2 \le q_1$ *then* $\gamma := \gamma_1^{-1}\gamma_2 \in \Gamma$ *and* $q(\gamma) = q := q_2/q_1$.

Proof. The statements 1), 2) and 3) are obvious. For 4) we have to prove that $\gamma(d\rho_0) = 1_H d\rho_0/q$ with $H = \gamma(H_+)$. Apply γ_1 to both sides to obtain an equivalent equality, and then use 3):

$$\gamma_2(d\rho_0) = \gamma_1(1_H d\rho_0)/q = 1_{\gamma_1(H)} d\rho_0/q/q_1.$$

Since $\gamma_1(H) = H_2$ and $qq_1 = q_2$ this equality holds, and $\gamma(d\rho_0) = 1_H d\rho_0/q$. □

These results yield a dichotomy: Either the image $q(\Gamma)$ contains a maximal element $p \in (0,1)$, and $q(\Gamma) = \{p, p^2, \dots\}$, or $q(\Gamma)$ is dense in $(0,1]$. The dichotomy will allow us to show that the extension ρ does not depend on γ.

Theorem 14.7. *Let* ρ_0 *be a probability measure on* H_+. *Let* Qß$(0,1)$ *be non-empty. For each* $q \in Q$ *let* $\gamma_q \in \mathcal{A}^h$ *satisfy*

$$\gamma_q(d\rho_0) = 1_{H_q} d\rho_0/q, \quad H_q = \gamma_q(H_+).$$

Let ρ_q *be the extension of* ρ_0 *corresponding to* γ_q. *These extensions are equal. Let* ρ *denote this extension.*

1) *If there exists a real* $p \in (0,1)$ *such that each* $q \in Q$ *is a positive integer power of* p, *then there exists* $\gamma \in \mathcal{A}^h$ *such that*

$$\gamma^k(\rho) = \rho/p^k, \quad k \in \mathbb{Z}.$$

2) *Otherwise there is a one-parameter group* γ^t, $t \in \mathbb{R}$, *in* $\mathcal{A}^h$, *such that*

$$\gamma^t(\rho) = e^t\rho, \quad t \in \mathbb{R}. \tag{14.8}$$

The vertical projection $\tilde\rho_0$ *of* ρ_0 *then has tail function*

$$\tilde\rho_0[v, \infty) = 1 - G_\tau(av), \quad v \ge 0$$

for some $a > 0$, *where* G_τ *is the standard GPD with parameter* τ *in* (5) *in the Preview. Let* (j_*, j^*) *be the minimal interval in* $\mathbb{R}$ *on which the extension* $\tilde\rho$ *of* $\tilde\rho_0$ *lives. Then* ρ *lives on* $\mathbb{R}^h \times (j_*, j^*)$, *and* $\tilde\rho$ *is the vertical projection of* ρ.

Proof. The construction (14.6) shows that the extension of ρ_0 for γ and for γ^m is the same for $m = 2, 3, \ldots$. In the discrete case there is a $\beta \in \Gamma$ such that $q(\beta) = p$. Then $\gamma_q = \beta^{m_q}\sigma_q$ for some $m_q \geq 1$ and some symmetry σ_q of ρ_0 in $\mathcal{A}^h$. Hence $\rho_\beta = \rho_{\gamma_q}$. This holds for each $q \in Q$.

In the continuous case the image $q(\Gamma)$ of the set Γ in (14.7) is dense in $(0, 1]$. This implies that the vertical projection $\tilde{\rho}_0$ has a GPD on $[0, \infty)$. Hence $q\colon \Gamma \to (0, 1]$ is continuous. Choose $\gamma_n \in \Gamma$ such that $q_n := q(\gamma_n) \to 1-0$. Then $W^{q_n} \Rightarrow W$ gives $\gamma_n(\rho_0) \to \rho_0$. The Convergence of Types Theorem yields a sequence of symmetries $\sigma_n \in \mathcal{A}^h$ of ρ_0 such that $\gamma_n' := \sigma_n^{-1}\gamma_n \to \mathrm{id}$, and $\gamma_n' \in \Gamma$ by the lemma above. From Lemma 18.79 it follows that there exist $\tau_n \in \mathcal{S}_h$, and a generator C such that

$$\beta_{k_n}^{[tm_n]} \to e^{tC} = \gamma^t, \quad t \in \mathbb{R},\ \beta_{k_n} = \tau_n^{-1}\gamma_{k_n}'$$

for some subsequence $k_n \to \infty$, and $m_n \to \infty$. Then $\gamma^s \in \Gamma$ for a dense set of $s > 0$ by writing $\gamma^s = \gamma_q\sigma_q$ for $q = e^{-s} \in q(\Gamma)$. By continuity $\gamma^s(W)$ is distributed like $W^{e^{-s}}$ for all $s \geq 0$. The extension $\rho = \rho_s$ induced by γ^s does not depend on s. Apply the uniqueness in the discrete case to $s = 1/2^n$, and let $n \to \infty$. □

Remark 14.8. The one-parameter group γ^t in (14.8) need not be unique. This issue is discussed in Sections 14.7 and 18.1.

14.5* Domains of attraction.

For $q(\Gamma)$ discrete, it is simple to write down the domain of attraction explicitly. This description remains valid in the continuous case. A drawback of this description for the continuous case is the introduction of a spurious period.

Let $\gamma(\rho) = \rho/p$ with $p \in (0, 1)$. Recall that ρ was constructed slicewise as a sum of transforms $\gamma^k\mu/p^k$, where μ was the restriction of ρ_0 to the slice $S = H_+ \setminus \gamma(H_+) = \mathbb{R}^h \times [0, c)$. So μ has mass $1 - p$. We shall define the distribution π of our random vector $Z = (X, Y)$ as a mixture of probability measures $\pi_n = \beta_n(\mu_n)$ with mixing distribution $q_0, q_1, \ldots$. The conditions are:

1) $q_{n+1}/q_n \to p$;

2) μ_n lives on S for $n \geq 1$, and converges weakly to the probability measure $\mu/(1-p)$;

3) $\beta_n = \gamma_1 \ldots \gamma_n$ with $\gamma_n \to \gamma$, and $\tilde{\gamma}_n(0) = c$ for $n \geq 1$.

Set $\tilde{\beta}_n(0) = y_n$. Then $\tilde{\beta}_n(c) = y_{n+1}$ by 3), and so the slices $S_n = \mathbb{R}^h \times I_n$ with $I_n = \tilde{\beta}_n[0, c) = [y_n, y_{n+1})$ nicely fill up $\mathbb{R}^h \times [y_1, y_\infty)$ with $y_\infty = \lim y_n \leq \infty$. We define the initial measure π_0 to be any probability measure on $\mathbb{R}^h \times (-\infty, y_1)$.

The slicing sequence $y_1, y_2, \ldots$ is basically a univariate construct. In the case of thin tails the increments $a_n = y_n - y_{n-1}$ are asymptotically equal, see (6.5), otherwise $a_{n+1}/a_n \to c$ with $c > 1$ for heavy tails, and $c \in (0, 1)$ for bounded tails.

Theorem 14.9. *Let ρ be an infinite measure on $\mathbb{R}^d$ which satisfies $\gamma(\rho) = \rho/p$ for some $\gamma \in \mathcal{A}^h$ and $p \in (0,1)$. Suppose $\tilde{\gamma}(0) = c > 0$. Set $j_* = \inf \gamma^{-n}(0)$ and $j^* = \sup \gamma^n(0)$. Assume ρ lives on $\mathbb{R}^h \times (j_*, j^*)$, $d\rho_0 = 1_{H_+} d\rho$ is a probability measure, and ρ does not charge the horizontal coordinate plane.*

If π is the probability measure defined above then $\beta_n^{-1}(\pi/r_n) \to \rho$ weakly on $\mathbb{R}^h \times [\tilde{\gamma}^k(0), \infty)$ for any $k \le 0$ where $r_n = \pi\{y \ge y_n\}$.

The converse also holds: If $Z = (X, Y)$ has distribution π, and $\alpha_n^{-1}(Z^{y_n}) \Rightarrow W$ where W has distribution ρ_0, α_n maps H_+ onto H^{y_n}, and $q_n := \mathbb{P}\{Y \ge y_n\}$ satisfies $q_{n+1}/q_n \to p$, then π has the form above for a sequence $\beta_n = \alpha_n \sigma_n$ with $\sigma_n \in \mathcal{S}_h$.

Proof. The first part is an immediate consequence of the construction. First note that $r_n/q_n \to 1 + p + p^2 + \cdots = 1/(1-p)$. Next for $k < 0$

$$\beta_n^{-1}(q_{n+k}\pi_{n+k})/r_n = \gamma_n^{-1} \cdots \gamma_{n+k+1}^{-1} \beta_{n+k}^{-1}(q_{n+k}\pi_{n+k})/r_n \to p^k \gamma^k \mu/(1-p).$$

A similar relation holds for $k \ge 0$. Combining a finite number of these slices S_{n+j}, $j = k, \dots, m$, we obtain weak convergence on $\bigcup\{\gamma^j(S) \mid k \le j \le m\}$. Exponential decrease ensures that $r_{n+m}/r_n \le 2p^m < \varepsilon$ eventually for any $\varepsilon > 0$. Hence weak convergence holds on $\{v \ge \tilde{\gamma}^k(0)\}$ for $k \le 0$.

For the second part write $\alpha_n^{-1}(Z^{y_{n+1}}) = W_n^{s_n}$ with $W_n = \alpha_n^{-1}(Z^{y_n}) \Rightarrow W$. By the CTT $W_n^{s_n} \Rightarrow W^s$. By the lemma below there exist symmetries $\sigma_n \in \mathcal{S}_h$ such that $\beta_n = \alpha_n \sigma_n$ satisfies $\beta_n^{-1}\beta_{n+1} \to \gamma$. Since the σ_n are symmetries they do not affect convergence: $\beta_n^{-1}(Z^{y_n}) = \sigma_n(W_n) \Rightarrow W$. By assumption α_n maps H_+ onto H^{y_n}. Hence $\tilde{\beta}_n^{-1}(y_n) = 0$. Also $\tilde{\beta}_n^{-1}(y_{n+1}) \to c$. Hence we may find $\beta_n' \sim \beta_n$ in $\mathcal{A}^h$ such that $\tilde{\beta}_n'[0, c) = [y_n, y_{n+1})$. □

Lemma 14.10. *Let $\mathcal{S} ß \mathcal{A}$ be a compact group, and $\gamma \in \mathcal{A}$. Suppose $\gamma^{-1}\mathcal{S}\gamma = \mathcal{S}$. Let the sequence $\alpha_0, \alpha_1, \dots$ in $\mathcal{A}$ satisfy*

$$\alpha_{n+1} = \alpha_n \gamma_n \sigma_n, \quad \gamma_n \to \gamma,\ \sigma_n \in \mathcal{S}.$$

There is a sequence $\beta_0, \beta_1, \dots$ such that

$$\beta_n = \alpha_n \tau_n, \quad \beta_{n+1} = \beta_n \gamma_n', \quad \tau_n \in \mathcal{S},\ \gamma_n' \to \gamma.$$

Proof. Write $\alpha_{n+1} = \alpha_n \tau_n^{-1} \gamma_n' \tau_{n+1}$ starting with $\alpha_1 = \alpha_0 \,\mathrm{id}\, \gamma_0 \sigma_0$, and setting

$$\alpha_{n+1} = \alpha_n(\tau_n^{-1}\tau_n)\gamma_n({\tau'}_n^{-1}\tau_n')\sigma_n = \alpha_n \tau_n^{-1}\gamma_n'\tau_{n+1}$$

with $\tau_{n+1} = \tau_n'\sigma_n$ and $\tau_n' = \gamma^{-1}\tau_n\gamma \in \mathcal{S}$ to ensure that $\gamma_n' = \tau_n\gamma_n\gamma^{-1}\tau_n^{-1}\gamma \to \gamma$ (since $\gamma_n\gamma^{-1} \to \mathrm{id}$ implies $\tau_n\gamma_n\gamma^{-1}\tau_n^{-1} \to \mathrm{id}$ for any relatively compact sequence (τ_n)). □

One could develop a theory for tail self-similar measures as a discrete counterpart to the theory of high risk scenarios and excess measures. We shall not do so.

14.6 The Extension Theorem. The Extension Theorem states that for $\pi \in \mathcal{D}^h(\rho_0)$ with normalizations α_y the measures $\alpha_y^{-1}(\pi)/\pi(H^y)$ converge vaguely to the excess measure ρ extending ρ_0. *Weak convergence* holds on every horizontal halfspace $J = \mathbb{R}^h \times [v, \infty)$ on which ρ is finite. If we choose y_n so that $\pi(H^{y_n}) \sim 1/n$, then, by the results in Section 5, the n-point sample clouds N_n from the distribution π, properly normalized, converge to a Poisson point process N with mean measure ρ weakly on J.

We shall first formulate a simple extension result. Weak convergence on the upper halfspace H_+ entails weak convergence on all horizontal halfspaces $\mathbb{R}^h \times [v, \infty)$ for $v > j_*$.

Proposition 14.11. *Suppose $\pi \in \mathcal{D}^h(\rho_0)$ with normalizations α_y. The vertical coordinate of the limit vector $W = (U, V)$ has a GPD. Let (j_*, j^*) in (14.2) be the interior of the support of the excess measure $\tilde{\rho}$ on $\mathbb{R}$ extending the distribution of V. The distribution ρ_0 of W extends to an excess measure ρ on $\mathbb{R}^h \times (j_*, j^*)$, and $\alpha_y^{-1}(\pi/\pi(H^y)) \to \rho$ weakly on $\mathbb{R}^h \times [v, \infty)$ for each $v > j_*$ as $\pi(H^y) \to 0+$.*

We shall prove more. Since each limit distribution ρ_0 has a unique extension to an excess measure ρ on $\mathbb{R}^h \times (j_*, j^*)$, we may write $\mathcal{D}^h(\rho)$ for $\mathcal{D}^h(\rho_0)$.

Theorem 14.12 (Extension Theorem). *Suppose $\pi \in \mathcal{D}^h(\rho)$ for an excess measure ρ which satisfies $\gamma^t(\rho) = e^t \rho$ for a one-parameter group $\gamma^t \in \mathcal{A}^h$, $t \in \mathbb{R}$. There exists a continuous curve $\beta\colon [0, \infty) \to \mathcal{A}^h$ which varies like γ^t, such that*

$$e^t \beta(t)^{-1}(\pi) \to \rho \ \text{ weakly on } \mathbb{R}^h \times [v, \infty), \quad t \to \infty,\ v > j_*,$$

where (j_*, j^*) in (14.2) is the domain of the vertical component $\tilde{\rho}$ of the excess measure ρ.

There is a relation between the normalizations α_y of the high risk scenarios Z^y, and the curve β. Let F be the df of the vertical component Y of the vector $Z = (X, Y)$ with distribution π. There exists a continuous df F_0 which is strictly increasing on $\{0 < F_0 < 1\} = (y_0, y_\infty)$, with the same upper endpoint y_∞ as F, and with finite lower endpoint y_0, and which satisfies $(1 - F(y))/(1 - F_0(y)) \to 1$ for $y \to y_\infty$ from below.

Proposition 14.13. *Let F and F_0 be as above. Define $y\colon [0, \infty) \to [y_0, y_\infty)$ by $1 - F_0(y(t)) = e^{-t}$. Let $\mathcal{S}^h$ denote the set of measure preserving transformations of ρ in $\mathcal{A}^h$. There exist $\sigma(t) \in \mathcal{S}^h$, $t \geq 0$, such that*

$$\alpha_{y(t)} \sim \beta(t)\sigma(t), \quad t \to \infty.$$

We shall now give the proof of these interrelated results.

Proof. Let $\alpha_y^{-1}(Z^y) \Rightarrow W$. The high risk scenarios W_q, $0 < q < 1$, all have the same shape. Hence the distribution of W extends to an excess measure ρ on $\mathbb{R}^h \times (j_*, j^*)$. This excess measure is unique. There exists a one-parameter group $\gamma^t \in \mathcal{A}^h$ such that $\gamma^t(\rho) = e^t \rho$ for $t \in \mathbb{R}$. This group need not be unique.

Set $p = 1/e$, $1 - F_0(y_n) = 1/e^n$, and $\alpha_n = \alpha(y_n)$. By Lemma 14.10 $\alpha_n = \beta_n \sigma_n$, where $\beta_n^{-1}\beta_{n+1} \to \gamma$ and $\rho_n := \beta_n^{-1}(\pi)/p^n \to \rho$ weakly on $\gamma^{-m}(H_+)$ for each $m \geq 1$. Embed β_n, $n \geq n_0$, in a continuous curve $\beta\colon [0,\infty) \to \mathcal{A}^h$ which varies like γ^t as in Section 18.2. Then

$$\beta(t)^{-1}(\pi)/p^t = (\beta(t)^{-1}\beta_{[t]}\gamma^{[t]-t})(\gamma^{t-[t]}\rho_{[t]}/p^{t-[t]}) \to \rho$$

weakly on $\gamma^{-m}(H_+)$ for all $m \geq 1$ since the first factor tends to id, and the second to ρ. Let $t_n \to \infty$. Then $\beta(t_n)^{-1}(Z^{y(t_n)}) \Rightarrow W$, and $\alpha(y(t_n))^{-1}(Z^{(y(t_n))}) \Rightarrow W$. By the CTT the sequence $\beta(t_n)^{-1}\alpha(y(t_n))$ is relatively compact, and all limit points are symmetries of W. These symmetries lie in $\mathcal{S}^h$ since $\alpha(y)$ and $\beta(t)$ do. Choose $\sigma(t) \in \mathcal{S}^h$ at minimal distance to $\beta(t)^{-1}\alpha(y(t))$. This is possible since for non-degenerate probability measures the set of symmetries is compact. Then $\alpha(y(t_n)) \sim \beta(t_n)\sigma(t_n)$. □

14.7 Symmetries. If the high risk scenarios from the distribution π converge to a non-degenerate limit vector W, then the distribution ρ_0 of W extends to an excess measure ρ and

$$e^t\beta(t)^{-1}(\pi) \to \rho \ \text{ weakly on } \mathbb{R}^h \times [v,\infty), \quad t \to \infty,\ v > j_*,$$

for a continuous normalization curve $\beta\colon [0,\infty) \to \mathcal{A}^h$, which varies like γ^t, where $\gamma^t(\rho) = e^t\rho$ for $t \in \mathbb{R}$. The symmetry group $\mathcal{S}$ of ρ_0 is compact by Theorem 18.75. If it is finite the symmetry group γ^t of the excess measure ρ is unique, as we shall see below. Moreover any continuous normalization curve then varies like γ^t, and is asymptotic to $\beta(t)\sigma$ for an element $\sigma \in \mathcal{S} \cap \mathcal{A}^h$.

Let us first ask: Are symmetries of ρ_0 symmetries of ρ? We give some examples.

Example 14.14. Let ρ be *Lebesgue measure* on the cone $C = \{v < -|u|\}$ in the lower half of $\mathbb{R}^2$, and ρ_0 the restriction to the halfplane $J_0 = \{v \geq -1\}$. Then W is uniformly distributed on the triangle $T = \{-1 \leq v < -|u|\} = C \cap J_0$. It has a discrete symmetry group of six elements, the permutation group of the vertices of T. Since ρ may be rotated into Lebesgue measure on the positive quadrant it has a two-dimensional symmetry group isomorphic to the group of positive diagonal matrices. The measure preserving symmetries are diagonal matrices $\mathrm{diag}(c, 1/c)$, $c \neq 0$. ◊

Example 14.15. Let ρ_1 have density $c/\|w\|^4$ on $\mathbb{R}^3 \setminus \{0\}$, and let $d\rho_2 = 1_{C^c}d\rho_1$ where C is the closed cone in $\{v \leq 0\}$ intersecting the horizontal plane $\{v = -1\}$ in the square $[-1,1]^2$. Choose $c > 0$ so that $d\rho_0 = 1_{J_0}d\rho_1$ is a probability measure on $J_0 = \{v \geq 1\}$. The measures ρ_1 and ρ_2 are excess measures with respect to the scalar expansions $\gamma^t : w \mapsto e^t w$. Let $\mathcal{S}_i$ denote the group of measure preserving transformations of the measure ρ_i. Then $\mathcal{S}_0 = \mathrm{O}(2)$, $\mathcal{S}_1 = \mathrm{O}(3)$, and $\mathcal{S}_2$ is the discrete subgroup of O(2) consisting of the symmetries of the square $[-1, 1]^2$. ◊

Example 14.16. The Gauss-exponential excess measure on $\mathbb{R}^3$ has a three-dimensional group of measure preserving symmetries; the symmetry group of the corresponding high risk limit distribution is a one-dimensional subgroup. ◊

Theorem 14.17. *Let ρ be an excess measure for exceedances over horizontal thresholds, and $\rho(H_+) = 1$. Let W be the high risk limit vector with distribution $d\rho_0 = 1_{H_+} d\rho$. Assume $\gamma^t(\rho) = e^t \rho$, $t \in \mathbb{R}$, with $\gamma^t \in \mathcal{A}^h$. Let ρ live on $\mathbb{R}^h \times (j_*, j^*)$, where (j_*, j^*) is the orbit $\tilde{\gamma}^t(0)$, $t \in \mathbb{R}$, where $\tilde{\gamma}^t$ describes the action of γ^t on the vertical component of γ^t. Assume ρ_0 is non-degenerate. Let $\sigma \in \mathcal{A}^h$. Then $\sigma(W)$ is distributed like W if and only if $\sigma(\rho) = \rho$.*

Proof. If $\sigma(\rho) = \rho$ and σ maps horizontal halfspaces into horizontal halfspaces then $\sigma(J_0) = J_0$. Hence $\sigma(\rho_0) = \rho_0$. Conversely let $\mathcal{S}_0$ denote the set of all $\sigma \in \mathcal{A}^h$ which satisfy $\sigma(\rho_0) = \rho_0$. Then $\mathcal{S}_0$ is a group, and $\sigma \mapsto \gamma^{-t}\sigma\gamma^t$ is an automorphism of $\mathcal{S}_0$ since $\sigma(W^y)$ is distributed like W^y for $y \in (0, j^*)$ if we let W^y denote the vector W conditioned to lie in the halfspace $\mathbb{R}^h \times [y, \infty)$, and

$$\gamma^{-t}\sigma\gamma^t(W) \stackrel{\mathrm{d}}{=} \gamma^{-t}\sigma(W^{\tilde{\gamma}^t(0)}) \stackrel{\mathrm{d}}{=} \gamma^{-t}(W^{\tilde{\gamma}^t(0)}) \stackrel{\mathrm{d}}{=} W.$$

The automorphisms form a group. Hence $\gamma^t\sigma\gamma^{-t} \in \mathcal{S}_0$ for $t > 0$. Let ρ_s be the restriction of ρ to $\mathbb{R}^h \times [\tilde{\gamma}^s(0), \infty)$. Then $\gamma^t\sigma\gamma^{-t}(\rho_0) = \rho_0$ gives $\gamma^{-t}(\rho_0) = \sigma\gamma^{-t}(\rho_0)$, and hence

$$e^{-t}(\rho_{-t}) = \gamma^{-t}(\rho_0) = \sigma\gamma^{-t}(\rho_0) = e^{-t}\sigma(\rho_{-t}).$$

Therefore $\sigma(\rho_{-t}) = \rho_{-t}$ for all $t > 0$, which implies $\sigma(\rho) = \rho$ since ρ lives on $\mathbb{R}^h \times (j_*, j^*)$. □

We shall now discuss the effect of extra symmetries on the normalization.

Example 14.18. The vector $W = (U, V) \in \mathbb{R}^{h+1}$ with U standard Gauss and V standard exponential and independent of U lies in the domain of the Gauss-exponential measure ρ:

$$e^t \beta(t)^{-1}(\pi) \to \rho \ \text{ weakly on } \mathbb{R}^h \times [c, \infty), \quad t \to \infty,\ c \in \mathbb{R}. \tag{14.9}$$

One may choose $\beta(t)(u, v) = (u, v+t)$, but also $\beta(t)(u, v) = (\sigma_t(u), v+t)$ for any curve $\sigma : [0, \infty) \to \mathrm{O}(h)$. Because of symmetry, there are many continuous curves β for which (14.9) holds. ◊

Theorem 14.19. *If the symmetry group in $\mathcal{A}^h$ of the high risk vector W is discrete then every continuous normalization curve $\beta\colon [0,\infty) \to \mathcal{A}^h$ in (14.9) varies like γ^t.*

Proof. By the Extension Theorem one may choose β to vary like γ^t. Let α also be a continuous normalization curve. By the Convergence of Types Theorem, for any sequence $t_n \to \infty$ $\alpha(t_n)^{-1}\beta(t_n)$ is relatively compact, and all limit points σ are symmetries of W, which lie in $\mathcal{A}^h$, since $\alpha(t)$ and $\beta(t)$ do. Choose symmetries $\sigma(t)$ such that $\alpha(t) \sim \beta(t)\sigma(t)$ for $t \to \infty$. Since the symmetry group is compact by Theorem 18.75, it is finite, and the continuous curve $\sigma(t)$ is eventually constant. So

$$\alpha(t_n)^{-1}\alpha(t_n+s_n) = \sigma^{-1}\beta(t_n)^{-1}\beta(t_n+s_n)\sigma \to \sigma^{-1}\gamma^s\sigma, \quad t_n \to \infty,\ s_n \to s.$$

Let $\bar\gamma^s = \sigma^{-1}\gamma^s\sigma$ have generator $\bar C$, and let $\gamma^s = e^{sC}$. If $\bar C = C$ then $\bar\gamma^s = \gamma^s$ for all s; if $\bar C \neq C$ then the group of measure preserving symmetries of ρ has dimension at least one. By the previous proposition it is finite. Contradiction. □

14.8 The Representation Theorem. *Excess measures* for exceedances are related to product measures. Suppose $\gamma^t(\rho) = e^t(\rho)$ for $t \in \mathbb{R}$, with $\gamma^t \in \mathcal{A}^h$. Let $0 < \rho(J_0) < \infty$ for $J_0 = \mathbb{R}^h \times [j_0,\infty)$, and let $\tilde\rho$ live on $(j_*, j^*) = \{\tilde\gamma^t(j_0) \mid t \in \mathbb{R}\}$.

Introduce the map

$$\Phi\colon (u,t) \mapsto \gamma^t(u,j_0), \quad \Phi\colon \mathbb{R}^{h+1} \to O = \mathbb{R}^h \times (j_*, j^*).$$

The map Φ is a homeomorphism. It is continuous. The map $t \mapsto \tilde\gamma^t(j_0)$ is a homeomorphism, and for fixed t the map $\Phi(\cdot,t)$ is an affine transformation on $\mathbb{R}^h$, which depends continuously on t. The map Φ transforms the group of vertical translations on $\mathbb{R}^{h+1}$ into the one-parameter group γ^t on O:

$$\begin{array}{ccc} \mathbb{R}^{h+1} & \xrightarrow{\tau^t} & \mathbb{R}^{h+1} \\ \Phi\big\downarrow & * & \big\downarrow\Phi \\ O & \xrightarrow[\gamma^t]{} & O \end{array}$$

Write $\mu = \Phi^{-1}(\rho)$. Then for any Borel set E_0ß$\mathbb{R}^h$

$$\mu(E_0 \times [t,\infty)) = e^{-t}\mu(E_0 \times [0,\infty)), \quad t \in \mathbb{R}.$$

Hence there is a finite measure ρ^* on $\mathbb{R}^h$ such that

$$\rho = \Phi(\rho^* \times \varepsilon), \quad \varepsilon(dt) = e^{-t}dt.$$

Introduce the class $\mathcal{J}$ of *invariant* Borel sets EßO. A set EßO is invariant if $\gamma^t(E) = E$ for all $t \in \mathbb{R}$. The class $\mathcal{J}$ is a σ-algebra. Sets in $\mathcal{J}$ are images of vertical Borel sets in

$\mathbb{R}^{h+1}$ under Φ. The measure ρ^* may be defined, using the one-to-one correspondence between sets $E \text{ß} \mathcal{J}$ and Borel sets E_0 in $\mathbb{R}^h$, by

$$\rho^*(E_0) = \rho(E \cap J_0), \quad E_0 \times \{j_0\} = E \cap \{v = j_0\}.$$

The total mass of ρ^* is $\rho^*(\mathbb{R}^h) = \rho(J_0)$. If $\rho(J_0) = 1$ then ρ^* is a probability measure.

Definition. The finite measure ρ^* is the *spectral measure* associated with ρ. The probability measure $\rho_0^* = \rho^*/\rho(J_0)$ is the *spectral distribution*.

The symmetries of ρ reflect a basic property of the limit vector: For any $v \in [0, j^*)$ the high risk scenario W^{H_v} has the same shape as W – since it also is limit of high risk scenarios Z^{H_y}. The product measure underlying the excess measure ρ yields a simple representation theorem for W.

Theorem 14.20 (Representation Theorem). *Let W be a limit vector for exceedances over horizontal thresholds. The excess measure ρ which extends the distribution of W satisfies $\gamma^t(\rho) = e^t \rho$ for a one-parameter group in $\mathcal{A}^h$. There exists a random vector $U^* \in \mathbb{R}^h$ and an independent standard exponential variable T such that W is distributed like $\gamma^T(U^*, 0)$.*

Corollary 14.21. *Let $W = (U, V)$. Suppose $v \geq 0$ and $\mathbb{P}\{V \geq v\} = e^{-t} > 0$. Conditionally on $V = v$ the vector (U, v) is distributed like $\gamma^t(U^*, 0)$. The conditional distributions depend continuously on v; they all have the same shape; U^* is the conditional distribution of U given $v = 0$.*

14.9 The generators in dimension $d = 3$ and densities. The limit vectors $W = (U, V)$ for exceedances over horizontal thresholds may be classified by the shape parameter τ of the *GPD* of the vertical component V. It is convenient to choose the horizontal halfspace J_0 on which the limit vector W lives to depend on the sign of τ: Take $J_0 = \{v \geq 1\}$ for $\tau > 0$ (heavy tails); $J_0 = \{v \geq -1\}$ for $\tau < 0$ (V bounded); and $J_0 = H_+$ for $\tau = 0$ (exponential tails). The corresponding excess measures ρ live on $\{v > 0\}$, $\{v < 0\}$ and $\mathbb{R}^d$ respectively.

As an appetizer we give the *Jordan form* for the generators of the excess measures in dimension $d = 3$. Even in dimension $d = 3$ there is a great variety of possible generators. One-dimensional groups γ^t, $t \in \mathbb{R}$, of affine transformations on $\mathbb{R}^3$ may be represented by a group of matrices e^{tC} of size 4, where the generator C has top row zero.

Example 14.22. If $\tilde{\gamma}^t$ is the translation $y \mapsto y + t$, the generator C has the form below, with $\varphi > 0$, $\lambda, \mu \in \mathbb{R}$. The second row $(1, 0, 0, 0)$ is the vertical coordinate η, the first row the virtual coordinate ζ_0.

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 1 & \lambda \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda & -\varphi \\ 0 & 0 & \varphi & \lambda \end{pmatrix}.$$

Cylinder symmetric limit distributions on H_+ are described by the third matrix with $\mu = \lambda \in \mathbb{R}$. The limit vector is symmetric for reflection in the vertical axis if $(-U, V)$ is distributed like (U, V). These vectors have symmetry groups generated by the third and fifth matrix above. $\diamondsuit$

Example 14.23. In case γ is not a translation it is convenient to replace H_+ by $J_0 = \{v \geq 1\}$ for expansions, $\tau > 0$, and by $J_0 = \{v \geq -1\}$ for contractions, $\tau < 0$, so that $\tilde{\gamma}^t(v) = e^{\tau t}v$. The vertical coordinate η now is the first row in which τ occurs. We distinguish three affine and five linear symmetry groups. The generator C has the form below, with $\varphi > 0$, $\mu, \nu \in \mathbb{R}$.

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \tau \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \tau & 0 \\ 0 & 0 & 1 & \tau \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \tau & 0 \\ 0 & 0 & 0 & \mu \end{pmatrix},$$

$$\begin{pmatrix} \tau & 0 & 0 \\ 1 & \tau & 0 \\ 0 & 1 & \tau \end{pmatrix}, \quad \begin{pmatrix} \tau & 0 & 0 \\ 1 & \tau & 0 \\ 0 & 0 & \mu \end{pmatrix}, \quad \begin{pmatrix} \tau & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 1 & \mu \end{pmatrix}, \quad \begin{pmatrix} \tau & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{pmatrix}, \quad \begin{pmatrix} \tau & 0 & 0 \\ 0 & \mu & -\varphi \\ 0 & \varphi & \mu \end{pmatrix}.$$

The seventh matrix with $\nu = \mu$ yields the cylinder symmetric distributions. $\diamondsuit$

The eight matrices in Example 14.23 give a rough initial classification. The behaviour of the excess measure is determined by the sign of the parameters μ and ν, and the sign of the differences $\mu - \tau$ and $\nu - \tau$. A special case is $\nu = \mu$. So for $\tau > 0$ (and for $\tau < 0$) there are $41 = 1 + 1 + 5 + 1 + 5 + 5 + 18 + 5$, rather than eight, qualitatively different excess measures, and for $\tau = 0$ there are $21 = 1+3+8+3+3$.

Choose w_1 to be the vertical coordinate, and w_0 the virtual coordinate. Sometimes one may write

$$C = \begin{pmatrix} J_\tau & 0 \\ 0 & C^* \end{pmatrix}, \quad J_0 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad J_\tau = \begin{pmatrix} 0 & 0 \\ 0 & \tau \end{pmatrix}, \ \tau \neq 0. \tag{14.10}$$

The last three matrices in the two examples above have this form. The group γ^t then preserves the vertical axis, and

$$e^{tC} = \begin{pmatrix} Q_\tau(t) & 0 \\ 0 & e^{tC^*} \end{pmatrix}, \quad Q_0(t) = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}, \quad Q_\tau(t) = \begin{pmatrix} 1 & 0 \\ 0 & e^{\tau\tau} \end{pmatrix}, \ \tau \neq 0. \tag{14.11}$$

Conditional on $V = \tilde{\gamma}^t(j_0)$ the horizontal component U now is distributed like $e^{tC^*}(U^*)$ where U^* has distribution ρ^*. In particular $(-U, V)$ is distributed like (U, V) if U^* is distributed like $-U^*$.

The distribution of the limit vector W depends on the generator C and on the distribution of U^*. If U^* has a density, it is simple to compute the density of $W = (U, V)$. We give an example.

Example 14.24. Suppose U^* is standard normal. Recall that U^* is the distribution of U given $T = 0$. What is the distribution of U_t, the vector U given $V = v(t)$? We choose the most difficult case: V has a *Pareto distribution* with tail exponent $\lambda > 0$, and the generator is the second of the three affine generators above. Observe that $(U_t, v(t))$ is distributed like $\gamma^t(U^*, 0)$, and hence U_t is normal. We compute e^{Ct} and write down the equation $\gamma^t w(0) = w(t)$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ t & 1 & 0 & 0 \\ 0 & 0 & e^{\tau t} & 0 \\ 0 & 0 & te^{\tau t} & e^{\tau t} \end{pmatrix} \begin{pmatrix} 1 \\ u_1 \\ 1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 + u_1 \\ e^{\tau t} \\ te^{\tau t} + e^{\tau t} u_2 \end{pmatrix}.$$

So $v(t) = e^{\tau t}$. Together with $\mathbb{P}\{T > t\} = \mathbb{P}\{V > v(t)\}$ this gives $e^{-t} = 1/(e^{\tau t})^\lambda$, and hence $\tau = 1/\lambda > 0$. So $U_t = (t + U_1^*, e^{t/\theta}(t + U_2^*))$ is normal, with mean $t(1, e^{\tau t})$, and with diagonal *covariance* matrix $\Sigma = \mathrm{diag}(1, e^{2t/\theta})$. By the same arguments, if U^* has density $f_0(u)$ then W has density

$$g(u, v) = g(v) f_v(u) = \frac{\theta}{v^{\theta+1}} f_0(u_1 - t, (u_2 - t)/v) \frac{1}{v}, \quad t = \theta \log v, \ v \geq 1.$$

This also is the density of the excess measure, but now for $v > 0$. ◊

14.10 Projections. The measure $\tilde{\rho} = \eta(\rho)$ on the interval (j_*, j^*) in (14.2) is the projection of the *excess measure* by the vertical coordinate η. This is an excess measure for the one-dimensional group of positive affine transformations $\tilde{\gamma}^t$ on $\mathbb{R}$, and (j_*, j^*) is the orbit $\tilde{\gamma}^t(0)$, $t \in \mathbb{R}$. Actually there is a host of projections of ρ which are excess measures. This is clear if one writes C in the *Jordan canonical form*. Section 18.12 gives the necessary information about the Jordan form of a square matrix. One may delete any Jordan block, except those containing the virtual coordinate ζ_0 and the vertical coordinate η. Blocks may also be reduced by deleting some of the higher entries. The reduced matrix is the generator of a one-parameter group $\bar{\gamma}^t$ of affine transformations on a lower dimensional space, and the image $\bar{\rho}$ is an excess measure with symmetries $\bar{\gamma}^t$. Proposition 18.62 shows that any admissible projection has this form in appropriate coordinates. However it is convenient to have a description of such projections which does not depend on the form of the matrix of C.

Example 14.25. Let $(X, Y) \in \mathcal{D}^h(U, V)$ where $(U, V) \in \mathbb{R}^3$ is a high risk limit vector with one of the generators C of the previous subsection. Let

$$W' = (U', V) \in \mathbb{R}^2, \quad U' = U_1 + U_2 + V + 1.$$

For which C is W' a high risk limit vector? Does $Z' \in \mathcal{D}^h(W')$ hold for the vector $Z' = (X_1 + X_2 + Y + 1, Y)$? $\diamondsuit$

With an $(1+m)$-dimensional subspace M of the dual L^* of the linear space $L = \mathbb{R}^{1+d}$ is associated a natural linear quotient map $Q: L \to M^*$ so that $\xi Q \in M$ for all $\xi \in M^{**}$. Call M *admissible* for the generator C on L if M contains the virtual coordinate ξ_0 and the vertical coordinate η, and if

$$\xi \in M \Rightarrow \xi C \in M.$$

If M is admissible there is a *generator* $\bar{C}$ on M^* such that

$$QC = \bar{C}Q. \tag{14.12}$$

Proposition 14.26. *Let ρ be an excess measure for the generator C. If MßL^* is admissible for C then $\bar{\rho} = Q\rho$ is an excess measure for the generator $\bar{C}$ in* (14.12).

Admissible projections Q need not transform domains of attraction into domains of attraction. If the normalizations respect the coordinates associated with the Jordan form, then domains of attraction are mapped into domains of attraction. This occurs for coordinatewise maxima where the normalizations are *CAT*s which preserve coordinates. Projections then just delete a number of components of the vector Z. By the Spectral Decomposition Theorem this also occurs if C is diagonal with distinct real eigenvalues. The *SDT* allows us to restrict attention to generators C whose complex eigenvalues all have the same real part. Below we look at excess measures which are highly symmetric.

Theorem 14.27 (Projection Theorem). *Suppose $(X, Y) \in \mathcal{D}^h(U, V)$ where (U, V) has a cylinder symmetric distribution. Let $1 \le m < d$. Set $U' = (U_1, \ldots, U_{m-1})$. Then (U', V) has a cylinder symmetric distribution. Let $A: \mathbb{R}^{h+1} \to \mathbb{R}^m$ be a linear surjection such that $\eta A(x, y) = y$ for a linear function η on $\mathbb{R}^m$. Then $A(X, Y) \in \mathcal{D}^h(U', V)$.*

The proof follows in $d-m$ steps from

Lemma 14.28. *Let $Z_n = (T_n, X_n, Y_n)$ and $W = (S, U, V)$ be vectors in $\mathbb{R}^{1+k+1}$. Suppose $\alpha_n^{-1}(T_n, X_n, Y_n) \Rightarrow (S, U, V)$, where $\alpha_n \in \mathcal{A}^h(\mathbb{R}^{(1+k)+1})$ preserves the class of horizontal halfspaces. Suppose $((S, U), V)$ is cylinder symmetric. Then (U, V) is cylinder symmetric and there are $\beta_n \in \mathcal{A}^h(\mathbb{R}^{k+1})$ such that $\beta_n^{-1}(X_n, Y_n) \Rightarrow (U, V)$.*

Proof. The proof is pure linear algebra. Cylinder symmetry of (U, V) follows since $\mathrm{O}(k)$ is a subgroup of $\mathrm{O}(1+k)$. Let $p(s, u) = u$ be the projection from $\mathbb{R}^{1+k}$ onto $\mathbb{R}^k$. Write $\alpha_n(w) = A_n w + a_n$. There exist rotations $R_n \in \mathrm{O}(1+k)$ such that $A_n R_n e_1 = r_n e_1$ for some $r_n > 0$. Let $\mathcal{A}^{h1}$ be the subgroup of $\mathcal{A}^h$ preserving the class of vertical lines. Then $\gamma_n = \alpha_n \circ R_n \in \mathcal{A}^{h1}$, and so too γ_n^{-1}. Observe that

$$\gamma_n^{-1}(t, x, y) = p_n + (r_n t + b_n^T x + s_n y, E_n x + f_n y, c_n y), \quad p_n \in \mathbb{R}^{1+k+1}$$

This implies $p \circ \gamma_n^{-1} = \beta_n^{-1} \circ p$ where $\beta_n^{-1}(x, y) = p(p_n) + (E_n x + f_n y, c_n y)$. $\square$

14.11 Sturdy measures and steady distributions. The first question we ask is: What can one say about the mass $\rho(J)$ of halfspaces J close to the horizontal halfspace J_0 on which the limit vector W lives? Note that $\rho(J_0) = 1$ and $\rho(\partial J_0) = 0$ since univariate GPDs are continuous. The measure ρ is *sturdy* if $\rho(J)$ is finite for all halfspaces J sufficiently close to J_0; it is *flimsy* if there is a sequence of halfspaces $J_n \to J_0$ such that $\rho(J_n) = \infty$ for $n \geq 1$. It turns out that some of the weirder generators in the examples above only allow flimsy excess measures.

In terms of the *intrusion cone* $\Delta = \Delta(\rho)$ introduced in Section 5.4 the excess measure ρ is sturdy precisely if the vertical coordinate η lies in Δ. Similarly convergence $\rho_n = \alpha_n^{-1}(\pi)/p_n \to \rho$ is *steady* if ρ lies in the *convergence cone* Γ.

Proposition 14.29. *If an excess measure is sturdy, then so are its admissible projections.*

Proof. Let $\bar{\rho} = p(\rho)$ on $\mathbb{R}^{m+1}$ be a projection of the excess measure ρ on $\mathbb{R}^{h+1}$. Halfspaces J in $\mathbb{R}^{m+1}$ correspond to halfspaces $H = p^{-1}(J)$ in $\mathbb{R}^{h+1}$, and $\bar{\rho}(J) = \rho(H)$. The horizontal halfspace H_0 in $\mathbb{R}^{h+1}$ with measure $\rho(H_0) = 1$ corresponds to a horizontal halfspace J_0 in $\mathbb{R}^{m+1}$ with mass $\bar{\rho}(J_0) = 1$. If $\bar{\rho}$ is flimsy there is a sequence $J_n \to J_0$ with $\bar{\rho}(J_n) = \infty$, and ρ is flimsy. $\square$

Let us now take a closer look at sturdy excess measures ρ, and their generators C. We distinguish three cases according to the sign of the Pareto parameter τ. We may reduce to a lower dimensional case by projection. In first instance we shall look at elementary excess measures. These live on *orbits* $\Gamma(t) = \gamma^t(z_0)$, $t \in \mathbb{R}$, of the one-parameter symmetry group $\gamma^t = e^{tC}$. For a detailed discussion of the form of such orbits we refer to Section 18.9.

1) ($\tau = 0$) The vertical component of ρ is exponential.

The measure ρ is sturdy if and only if there exists an $\varepsilon > 0$ such that the complement of the cone $C_\varepsilon = \{v < -\varepsilon\|u\|\}$ has finite measure. We shall call ρ *completely sturdy* if $\rho(H)$ is finite for all halfspaces $H = \{v \geq c^T u + c_0\}$, $c \in \mathbb{R}^h$, $c_0 \in \mathbb{R}$. Let U^* be the random vector in $\mathbb{R}^h$ with spectral distribution ρ^*.

Proposition 14.30. *Suppose $\tau = 0$. If ρ is sturdy, then the generator in Jordan form has the form*

$$\begin{pmatrix} J & 0 \\ 0 & C^* \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{14.13}$$

where J has size two, and C^* is a matrix of size h in Jordan form describing the effect of the symmetry group on the horizontal coordinates.

Proof. By Proposition 14.29 it suffices to take $d = 2$ and show that the group $(u, v) \mapsto (u + tv + t^2/2, v + t)$ only has flimsy excess measures. The orbit of the point $(a, 0)$ is the curve $\Gamma(t) = (a + t^2/2, t)$, $t \in \mathbb{R}$. It will lie above the cone C_ε eventually for $t \to -\infty$. If $\mathbb{P}\{U^* = a\} = p > 0$ then $\rho(\Gamma[t_1, t_2]) = p\tilde{\rho}(t_1, t_2]$, and hence $\rho(\Gamma(-\infty, t]) = \infty$ for all t. Hence ρ is flimsy if U^* has an atom in a. The same argument shows that ρ is flimsy if $\mathbb{P}\{U^* \geq a\} > 0$. □

Remark 14.31. For $C^* = 0$ there is a close link to the multivariate *Laplace transform* of U^*:

$$\rho\{v \geq c^T u\} = \iint_{c^T u}^{\infty} e^{-v} dv d\rho^*(u) = \int e^{-c^T u} d\rho^*(u) = \mathbb{E} e^{-c^T U^*}.$$

The excess measure ρ is sturdy if and only if the *Laplace transform* of the spectral measure ρ^* is finite on a neighbourhood of the origin. The excess measure is completely sturdy if the Laplace transform is an entire function.

If C^* has an eigenvalue $\lambda < 0$, or a complex eigenvalue λ in $\Re < 0$, then ρ is flimsy. If λ is real, take $d = 2$. Orbits have the form $\Gamma(t) = (e^{\lambda t} a, t)$ and $\Gamma(-\infty, t]$ has infinite mass for all t. Since the orbit grows exponentially in the horizontal direction, $\rho(C_\varepsilon^c) = \infty$ for all ε if U^* charges $a \neq 0$. This also holds if $|U^*| \geq \delta$ has positive probability for some $\delta > 0$. If λ is non-real take $d = 3$. The orbit Γ will spiral, but the same argument applies.

Assume C^* is a Jordan block with zeros on the diagonal and ones below. Then orbits have the form $\Gamma(t) = (a_1, a_2 + ta_1, a_3 + ta_2 + t^2 a_1/2, \ldots)$. If $U_1^* = 0$ a.s. then ρ lives on the hyperplane $\{u_1 = 0\}$ and is degenerate. If $a_1 \neq 0$ and if the size of C^* is 3 or larger then ρ is flimsy. If the size of C^* is 2 then ρ is flimsy if U_1^* is unbounded. If U_1^* is bounded, then ρ is sturdy, but not completely sturdy. The same results apply if C^* is a Jordan block corresponding to an imaginary eigenvalue.

Conclusion. Let ρ be a sturdy excess measure for exceedances over horizontal thresholds. Assume $\tilde{\rho}$ is exponential. Then the *generator* has the form (14.13) and the eigenvalues of C^* lie in $\Re \geq 0$. For eigenvalues on the imaginary axis there are strong conditions on the tail behaviour of U^* in the direction of the eigencoordinates. In dimension $d = 3$ this leaves only $0 + 0 + 4 + 2 + 2 = 8$ qualitatively different sturdy excess measures ρ for $\tau = 0$. ◊

2) ($\tau > 0$) The univariate measure $\tilde{\rho}$ is Pareto on $(0, \infty)$ with exponent $1/\tau > 0$.

If ρ is sturdy, then, by the argument above, orbits converge to the origin for $t \to -\infty$, or at least remain bounded. In the first case γ^t is a linear expansion group, and all complex eigenvalues of the generator C lie in $\Re > 0$. Moreover $\rho(\varepsilon B^c)$ is finite for all $\varepsilon > 0$. In the second case there may be eigenvalues on the imaginary axis, but the corresponding part of the complex Jordan form is diagonal. See Section 18.9 for details. If $\xi_1, \ldots, \xi_m$ are the real coordinates corresponding to this part of the Jordan matrix, then the variables $\xi_i(U^*)$, $i = 1, \ldots, m$, are bounded.

3) ($\tau < 0$) The univariate measure $\tilde{\rho}$ is a power law: $\rho[-t, \infty) = ct^\lambda$ for $t > 0$ with $\lambda = 1/|\tau|$.

If ρ is sturdy, the diagonal elements of the real Jordan form satisfy $c_{ii} \geq \tau$.

If one imposes cylinder symmetry on the limit distribution, the number of classes is reduced to $2 + 4 + 4$ in every dimension $d > 2$.

For the cases of practical interest, $\tau \geq 0$, steady convergence may be handled by the theorem below. Recall that V^+ is a cone in the dual space, defined in (5.9). It consists of all linear functionals ξ for which $\xi(V)$ is bounded above.

Theorem 14.32. *Let $\tau \geq 0$. Let the excess measure ρ for exceedances over horizontal thresholds be sturdy. Then there is a compact set Kß$\{v = j_*\}$ such that ρ is a Radon measure on $O = \mathbb{R}^d \setminus K$. Assume $\rho(H_+) = 1$. Let V be an open cone, such that KßV, and $V \cap \{v > j_*\}$ is bounded. Assume $\rho(V^c)$ is finite, and $\rho(\partial V) = 0$. Let π be a probability measure on $\mathbb{R}^d$ and $\alpha_n \in \mathcal{A}^h$ such that $p_n = \pi(\alpha_n(H_+)) \to 0$ and $\alpha_n^{-1}(\pi)/p_n \to \rho$ vaguely on O, and weakly on H_+. Set $V_n = \alpha_n(V)$. If $\pi(V_n^c)/p_n \to \rho(V^c)$ then $\alpha_n^{-1}(\pi)/p_n \to \rho$ weakly on V^c, and*

$$\alpha_n^{-1}(Z^{H_n}) \Rightarrow W_J \tag{14.14}$$

for any halfspace $J = \{\xi \geq c\}$ßO with ξ an interior point of V^+, $\rho(J) > 0$ and $\rho(\partial J) = 0$, and any sequence $H_n = \alpha_n(J_n)$ with $J_n \to J$. Here W_J is the vector with distribution $d\rho^J = 1_J d\rho/\rho(J)$.

Proof. The existence of K follows from the analysis of sturdy measures above. The limit relation (14.14) holds by the theory developed in Chapter I: Weak convergence on $F = V^c$ follows from Theorem 4.21. Hence $\Lambda = \text{int}(V^+)$ lies in the convergence cone Γ. The result now follows from Propositions 5.20 and 5.21. □

14.12 Spectral stability. There is a more algebraic robustness condition for excess measures.

Definition. An excess measure ρ for exceedances over horizontal thresholds satisfies the condition of *spectral stability* if ρ also has a spectral measure on hyperplanes close to the horizontal plane ∂J_0: Every halfspace J in a neighbourhood of J_0 has the property that ρ-a.e. orbit $t \mapsto \gamma^t(w)$ intersects ∂J in one point.

Let us describe the class $\mathcal{R}_\tau$ of sturdy excess measures with Pareto parameter τ, which are spectrally stable.

If $\tau = 0$ then the $\tilde{\gamma}^t$ are vertical translations and the complex eigenvalues of the generator C lie on the imaginary axis. In the complex *Jordan form* of C only blocks of size one and two occur. For any base vector e_i the corresponding univariate spectral measure $\xi_i(\rho^*)$ has a moment generating function which converges on a neighbourhood of the origin. If $U^* \in \mathbb{R}^h$ has distribution ρ^* then $\mathbb{E}e^{\varepsilon|U_i^*|}$ is finite for some $\varepsilon > 0$. For certain base vectors the components U_i^* are bounded.

If τ is positive one may assume that $\tilde{\rho}$ lives on $(0, \infty)$. The vertical marginal $\tilde{\rho}$ has tail function

$$\widetilde{R}(t) = \tilde{\rho}[t, \infty) = \rho(\mathbb{R}^h \times [t, \infty)) = 1/t^{1/\tau}, \quad t > 0.$$

The symmetries are linear. The orbits $\gamma^t(z)$ converge to the origin for $t \to -\infty$, or remain bounded with probability one. The eigenvalues θ of the generator lie in the strip $0 \leq \Re\theta \leq \tau$ in $\mathbb{C}$. For $\Re\theta = \tau$ the corresponding Jordan blocks are diagonal. For $\rho \in \mathcal{R}_\tau$ the *spectral measure* lives on $\mathbb{R}^h \times \{1\}$ and has bounded support.

For $\tau < 0$ the description of $\mathcal{R}_\tau$ is similar.

Theorem 14.33. *The excess measures ρ in $\mathcal{R}_\tau$ are characterized as follows. The generator of the symmetry group has Jordan form*

$$C = \begin{pmatrix} J_\tau & 0 \\ 0 & C^* \end{pmatrix}, \quad J_0 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad J_\tau = \begin{pmatrix} 0 & 0 \\ 0 & \tau \end{pmatrix}, \ \tau \neq 0. \tag{14.15}$$

For $\tau \neq 0$ the real parts of the eigenvalues θ of C^* lie in the closed interval with endpoints τ and zero, if $\Re\theta = \tau$ the corresponding Jordan blocks are diagonal. For $\tau = 0$ the eigenvalues lie on the imaginary axis and only complex Jordan blocks of size one and two occur. For $\tau \neq 0$ the spectral measure ρ^* has bounded support; for $\tau = 0$ the spectral measure has exponential tails: $\int e^{\varepsilon\|u\|} d\rho^*(u) < \infty$ for some $\varepsilon > 0$ and the image of ρ^* under the quotient map onto $\mathbb{R}^h/\text{ker}C^*$ has bounded support.

Let us describe the limit Poisson point process N with mean measure ρ. The vertical component $\widetilde{N}$ is a Poisson point process on $\mathbb{R}$ with mean measure $\tilde{\rho}$ and points $V_1 > V_2 > \cdots$. The points $W_n = (U_n, V_n)$ of N may be ordered by the vertical coordinate. Assume $\rho \in \mathcal{R}_\tau$. For simplicity assume the generator C has the form (14.15) with $C^* = \operatorname{diag}(\theta_1, \ldots, \theta_h)$, and all eigenvalues real. Let $U_1^*, U_2^*, \ldots$ be independent observations from the spectral measure ρ^*, and assume the sequence (U_n^*) is independent of the sequence (V_n). Then $p_k = \theta_k/\tau \in [0, 1]$ for $k = 1, \ldots, h$, for $\tau \neq 0$, and N has points

$$W_n = (V_n^{p_1} U_{n1}^*, \ldots, V_n^{p_h} U_{nh}^*, V_n), \quad n = 1, 2, \ldots.$$

For $\tau = 0$ the eigenvalues are zero, and $W_n = (U_n^*, V_n)$.

14.13 Excess measures for horizontal thresholds. We now have a good understanding of excess measures for exceedances over horizontal thresholds. For such a measure there exists a unique horizontal halfspace $J_0 = \mathbb{R}^h \times [j_0, \infty)$ such that $\rho(J_0) = 1$.

The vertical coordinate of ρ is a univariate excess measure $\tilde{\rho}$. It is an infinite Radon measure on an unbounded interval (j_*, j^*) with a continuous density. This interval is the orbit $\{\tilde{\gamma}^t(j_0) \mid t \in \mathbb{R}\}$. The restriction to $[j_0, \infty)$ is a GPD. By definition the parameter τ of this distribution is the Pareto parameter of the excess measure ρ. We may choose coordinates such that (j_*, j^*) is the positive halfline, the negative halfline or the whole real line, depending on the sign of τ. By replacing ρ by $c\rho$ with $c > 0$ we may arrange that $j_0 = \operatorname{sign}(\tau) \in \{-1, 0, 1\}$. The tail function of $\tilde{\rho}$ then is a power function or an exponential function. In these coordinates it is obvious that $\tilde{\rho}$ has a unique group of symmetries $\tilde{\gamma}^t$, positive affine transformations on $\mathbb{R}$, such that $\tilde{\gamma}^t(\tilde{\rho}) = e^t \tilde{\rho}$. The symmetries $\tilde{\gamma}^t$ are multiplications $v \mapsto e^{\tau t} v$ or translations $v \mapsto v + t$. This is just the univariate theory.

The excess measure ρ lives on $\mathbb{R}^h \times (j_*, j^*)$ and satisfies $\gamma^t(\rho) = e^t \rho$, $t \in \mathbb{R}$, for a one-parameter group $\gamma^t = e^{tC}$ in $\mathcal{A}^h$. The matrices are

$$C = \begin{pmatrix} 0 & 0 & 0 \\ p_0 & C^* & q_0 \\ b_0 & 0 & \tau \end{pmatrix}, \quad \gamma^t = \begin{pmatrix} 1 & 0 & 0 \\ p(t) & A^t & q(t) \\ b(t) & 0 & a^t \end{pmatrix}, \quad A^t = e^{tC^*},\ a^t = e^{t\tau}.$$

The entries in the bottom row describe the positive affine transformations $\tilde{\gamma}(t)\colon v \mapsto a^t v + b(t)$. If the generator has *Jordan form*, then $b_0 = 0$, or $\tau = 0$ and $b_0 = 1$, corresponding to $v \mapsto e^{\tau t} v$ and $v \mapsto v + t$. Moreover $q_0 = 0$ and hence $q(t) \equiv 0$. In addition to the one-parameter group γ^t in $\mathcal{A}^h$ there is a probability measure on $\mathbb{R}^h$, the spectral distribution ρ_0^*. Any triple (C, ρ_0^*, j_0) determines an excess measure ρ, provided $b_0 + \tau j_0 > 0$, to ensure that $\gamma^t(J_0)$ßJ_0 for $t > 0$. If ρ_0^* has a density then one may construct the density of ρ as in Section 14.9, or by using (14) in the Preview, where t may be expressed in terms of $v = \tilde{\gamma}^t(j_0)$.

Conditions on the excess measure ρ, that it is sturdy, or spectrally stable, are easily translated into conditions on the Jordan form of the generator C, and the form of the spectral measure ρ^*. This is done by reducing dimension by a suitable projection, as described in Section 14.10. For the present we shall not impose such extra conditions.

Proposition 14.34. *Let ρ be an excess measure with symmetries $\gamma^t = e^{tC} \in \mathcal{A}^h$. Suppose $\rho(J_0) = 1$ for $J_0 = \mathbb{R}^h \times [j_0, \infty)$. Let $d\rho_0 = 1_{J_0} d\rho$. If ρ is full then ρ_0 is a limit distribution for exceedances over horizontal thresholds, and $\mathcal{D}^h(\rho_0)$ is non-empty.*

Proof. If ρ_0 lives on a hyperplane then so does ρ, since the hyperplane is invariant under γ^t for $t \geq 0$, and hence for all t. This shows that ρ_0 is non-degenerate. Now observe that $\rho_0 \in \mathcal{D}^h(\rho)$ since $e^t \gamma^{-t}(\rho_0) = \rho$ on $\{v \geq v(-t)\}$ where $v(-t) = \tilde{\gamma}^{-t}(j_0) \to j_*$ for $t \to \infty$. □

Example 14.35. It is possible that ρ is full and that ρ^* is concentrated in one point. Take $\gamma^t(u_1, u_2, v) = (e^{-t}u_1, e^t u_2, v + t)$, and let ρ^* be concentrated in $(1, 1, 0)$. ◊

The vector W has the property that for $v \in [j_0, j^*)$ the conditional distributions of U given $V = v$ all are of the same type. This is a consequence of the Representation Theorem: the limit vector W is distributed like $\gamma^T(U^*, j_0)$, with T standard exponential, and independent of U^*. Conditional on $V = v$ the horizontal part U is distributed like $A^t U^* + m(t)$ where $m(t)$ is a vector in $\mathbb{R}^h$, A^t is the central matrix in γ^t above, and $t = t(v)$ is determined by the equation $e^{-t} = \mathbb{P}\{V \geq v\}$. If U^* is standard normal then all these conditional distributions are normal; if U^* is uniformly distributed on the open unit ball in $\mathbb{R}^h$, then for each v the vector U conditional on $V = v$ is uniformly distributed on an open ellipsoid centered in $m(t)$; if U^* is uniformly distributed on the centered cube $(-1, 1)^h$ in $\mathbb{R}^h$ then U, conditional on $V = v$ is distributed on an open parallelopiped centered in $m(t)$.

14.14 Normalizing curves and typical distributions. The global behaviour of the distribution of a vector $Z = (X, Y)$ in $\mathcal{D}^h(\rho)$ is determined by the excess measure ρ and by the curve of normalizations in $\mathcal{A}^h$. In this section we want to describe how one may construct a random vector $Z \in \mathcal{D}^h(\rho)$ given the excess measure and a curve of normalizations.

If the excess measure ρ for exceedances over horizontal thresholds has a continuous density g, then

$$g(u, v) = g_v(u)\tilde{g}(v),$$

where $\tilde{g}$ is the density of the corresponding *GPD* G_τ, or rather of the measure $\tilde{\rho}$ on (j_*, j^*) extending this GPD. If the density g^* of the spectral measure ρ^* is Gaussian, then g_v is a Gaussian density for each v; if g^* is the uniform density on the unit ball in $\mathbb{R}^h$, then g_v is the uniform density on an ellipsoid E_v in $\mathbb{R}^h$. In general the conditional densities g_v all have the same shape.

So if we want to construct distributions π in the domain of ρ, the obvious thing to do is to look at densities f of the form

$$f(x, y) = f_y(x)\tilde{f}(y)$$

where all densities f_y have the same shape. If ρ^* is Gaussian, we take densities f_y which are Gaussian; if ρ^* is uniformly distributed on the unit ball B in $\mathbb{R}^h$, then f_y is the density of the *uniform distribution* on an ellipsoid E_y in $\mathbb{R}^h$.

Definition. Let ρ be an excess measure for exceedances over horizontal thresholds with Pareto parameter τ, and with spectral distribution ρ_0^*. A probability measure π is *typical* if

$$\pi(dx, dy) = \pi_y(dx)\widetilde{F}(dy), \tag{14.16}$$

where the probability distributions π_y depend continuously on y, and all are of the same type as ρ_0^*.

Often one assumes that ρ has a continuous density on $\mathbb{R}^h \times (j_*, j^*)$, or a density g which is continuous and positive on an open subset of $\mathbb{R}^h \times (j_*, j^*)$, and vanishes outside this open set. The spectral density g^* then has similar properties on $\mathbb{R}^h$. In that case we may want to assume that Y has a density $\tilde{f}$ which is continuous and positive on $[y_0, y_\infty)$, and vanishes outside this interval, and that $\tilde{f}$ varies regularly in y_∞ in case $\tau \neq 0$, and is a *von Mises function* in case $\tau = 0$. If $Z = (X, Y) \in \mathcal{D}(\rho)$ then Y has a df $\widetilde{F} \in \mathcal{D}^+(\tau)$, where τ is the Pareto parameter of ρ. By the univariate theory there is a df F tail asymptotic to $\widetilde{F}$, whose density has the desired properties. The normalized density $\tilde{g}_t(v)$ will converge to the density $\tilde{g}$ of $\tilde{\rho}$ on (j_*, j^*) uniformly on compact subsets, and in $\mathbf{L}^1([v, \infty))$ for any $v > j_*$, by the univariate theory. The dfs F and $\widetilde{F}$ have the same normalizations. Since one may replace the df $\widetilde{F}$ in (14.16) by any df F which is tail asymptotic to $\widetilde{F}$ without affecting convergence, one may choose π typical such that the density $f(x, y) = f^*(x, y)\tilde{f}(y)$ is continuous on $\mathbb{R}^h \times [y_0, y_\infty)$, or continuous and positive on an open subset of this space, and zero elsewhere.

We shall use these typical distributions and densities to explore the domains of attraction of excess measures for exceedances over horizontal thresholds. There are two approaches.

1) In the geometric approach, which we shall use in the next section, one starts with a df $\widetilde{F} \in \mathcal{D}^+(\tau)$ with continuous density $\tilde{f}$, a simple spectral distribution, Gaussian say, or uniform on a ball or a cube, and a simple one-parameter group of affine transformations γ^t in $\mathcal{A}^h$, vertical translations, or scalar expansions or contractions, and one asks for typical densities which lie in the domain of the associated excess measure. For the Gaussian spectral measure we have to determine continuous curves $\mu(y)$ in $\mathbb{R}^h$, and positive quadratic forms Σ_y, depending continuously

on y, such that the vector with density $f^*(x, y)\tilde{f}(y)$ lies in the domain of an excess measure with Gaussian spectral measure, and the given symmetries. Here, for each y the function $f^*(\cdot, y)$ is a Gaussian density on $\mathbb{R}^h$ with expectation $\mu(y)$ and *covariance* Σ_y. Similarly, if ρ^* is uniformly distributed on the unit ball in $\mathbb{R}^h$, then (X, Y) has density $1_T(x, y)h(y)$, where $T ß \mathbb{R}^h \times [y_0, y_\infty)$ is a curvaceous tube with horizontal sections which are ellipsoids $\mu(y) + E_y$ in $\mathbb{R}^h$, and $h(y) = \tilde{f}(y)/|E_y|$.

2) In the algebraic approach, which we follow here, one is given a continuous curve $\beta \colon [0, \infty) \to \mathcal{A}^h$, which varies like γ^t, and one assumes that the typical distribution π satisfies the basic limit relation $e^t \beta(t)^{-1}(\pi) \to \rho$. We shall prove that for any full excess measure ρ there exists a typical distribution $\pi \in \mathcal{D}^h(\rho)$, such that $\rho_t = e^t \beta(t)^{-1}(\pi) \to \rho$ weakly on $\mathbb{R}^h \times [v, \infty)$ for $t \to \infty$ for any $v > j_*$. If ρ has a continuous density g we may choose π to have a continuous density, and the densities of ρ_t will converge to g uniformly on compact sets.

The algebraic approach below is rather formal. It is advisable to take a look at the more geometric approach in Section 15.2 before proceeding. Now we proceed in three steps:

1) We first want to regularize the normalization curve $\beta \colon [0, \infty) \to \mathcal{A}^h$ so that it is continuous, and so that the vertical coordinate $t \mapsto y(t) = \tilde{\beta}(t)(j_0)$ is a homeomorphism from $[0, \infty)$ onto $[y_0, y_\infty)$. This determines a df $\tilde{F}(y)$ for the vertical coordinate by $1 - \tilde{F}(y(t)) = e^{-t}$.

By the general theory on multivariate regular variation in Sections 18.1 and 18.2 for any curve $\alpha \colon [0, \infty) \to \mathcal{A}$ which *varies like* γ^t, one may choose $\beta \colon [0, \infty) \to \mathcal{A}$ continuous such that $\beta(t) \sim \alpha(t)$ for $t \to \infty$. The curve β will automatically vary like γ^t. Moreover if $\alpha(t) \in \mathcal{G}$ for all $t \geq 0$ for a closed subgroup $\mathcal{G}$ of $\mathcal{A}$, then by construction this also holds for $\beta(t)$. By Lemma 15.24 we are free to alter the normalizations for the vertical coordinate. If $\delta \colon [0, \infty) \to \mathcal{A}^+$ is continuous and asymptotic to $\tilde{\beta}$, then there exists a continuous curve $\alpha \colon [0, \infty) \to \mathcal{A}^h$ such that $\tilde{\alpha} = \delta$ and $\alpha(t) \sim \beta(t)$ for $t \to \infty$.

2) How does one define the conditional distributions π_y? Suppose $t \mapsto y(t) = \tilde{\beta}(t)(j_0)$ is a homeomorphism from $[0, \infty)$ onto $[y_0, y_\infty)$. We shall define the conditional distribution of X given $Y = y$ to be that of $\mu(y) + A(y)U^*$ where

$$\beta(t)(u, j_0) = (A(y)u + \mu(y), y), \quad u \in \mathbb{R}^h, \ y = y(t) \in [y_0, y_\infty).$$

This defines a continuous family of probability distributions π_y on $\mathbb{R}^h$, since $y \mapsto \mu(y)$ and $y \mapsto A(y)$ are continuous. Let $\tilde{F}$ be a df such that $e^t \tilde{\beta}(t)^{-1}(\tilde{F}(dy)) \to \tilde{\rho}(dv)$ weakly on $[c, \infty)$ for any $c > j_*$ for $t \to \infty$. Then y_∞ is the upper endpoint of $\tilde{F}$. Set $\pi_y = \pi_{y_0}$ for $y < y_0$. We now have a vector $Z = (X, Y)$ with distribution $\pi(dx, dy) = \pi_y(dx)\tilde{F}(dy)$. Define $\rho_t = e^t \beta(t)^{-1}(\pi)$. Then $\tilde{\rho}_t \to \tilde{\rho}$ weakly on $[v, \infty)$ for any $v > j_*$ by the univariate theory in Section 18.3. In order to prove $\rho_t \to \rho$ weakly on $\mathbb{R}^h \times [v, \infty)$ it suffices to prove weak convergence of the conditional distributions of the horizontal component, given the vertical component. This follows from

Proposition 14.36. *Let $\mathcal{X}$ and $\mathcal{Y}$ be locally compact separable metric spaces, and let $\mathcal{P}(\mathcal{X})$ denote the Polish space of probability measures on $\mathcal{X}$. Let $\pi_n^*\colon \mathcal{Y} \to \mathcal{P}(\mathcal{X})$ be continuous for $n = 0, 1, \ldots$, and suppose*

$$\pi_n^*(\cdot, y_n) \to \pi_0^*(\cdot, y_0), \quad y_n \to y_0,\ y_0 \in \mathcal{Y}.$$

Let μ_n be finite measures on $\mathcal{Y}$ for $n \geq 0$. Set $\pi_n(dx, dy) = \pi_n^*(dx, y)\mu_n(dy)$. If $\mu_n \to \mu_0$ weakly on $\mathcal{Y}$ then $\pi_n \to \pi_0$ weakly on $\mathcal{X} \times \mathcal{Y}$.

Proof. It suffices to prove vague convergence. Let $\varphi\colon \mathcal{X}\times\mathcal{Y} \to [0, \infty)$ by continuous with compact support contained in $K_1 \times K_2$. The functions

$$\psi_n(y) = \int \varphi(x, y)\pi^*(dx, y), \quad y \in \mathcal{Y}$$

vanish outside K_2. Let $y_n \to y_0$. Then $\varphi(\cdot, y_n) \to \varphi(\cdot, y_0)$ uniformly on $\mathcal{X}$, and

$$\int \varphi(x, y_0)\pi_n^*(dx, y_n) \to \int \varphi(x, y_0)\pi_0^*(dx, y_0)$$

implies $\psi_n(y_n) \to \psi_0(y_0)$. Hence $\psi_n \to \psi_0$ uniformly on $\mathcal{Y}$, and $\int \psi_n d\mu_n \to \int \psi_0 d\mu_0$. □

3) *Weak convergence* of the conditional distributions is a simple matter, since these all are of the same type. We are concerned with maps of the form $\varphi(x, y) = \varphi_y(x)\tilde{\varphi}(y) = (u, v)$. The composition of two such maps has a similar form, and so has the inverse if it exists. Measures transform as follows:

$$\varphi(\pi_y(dx)\tilde{\pi}(dy)) = \mu_v(du)\tilde{\mu}(dv), \quad (\varphi_y(\pi_y), \tilde{\varphi}(\tilde{\pi})) = (\mu_v, \tilde{\mu}), \quad v = \tilde{\varphi}(y).$$

Write $\varphi(\pi_y, y_0) = (\mu_v, v_0)$ with $v = \tilde{\varphi}(y)$ and $v_0 = \tilde{\varphi}(u_0)$ where (π_y, y_0) denotes the measure π_y on the hyperplane $y = y_0$. Then

$$\beta(t_n)^{-1}\beta(t_n + s_n)(\rho_0^*, j_0) =: \gamma_n(\rho_0^*, j_0) \to \gamma^s(\rho_0^*, j_0), \quad t_n \to \infty,\ s_n \to s.$$

This gives:

Theorem 14.37. *Let ρ be an excess measure with symmetries, $\gamma^t(\rho) = e^t\rho$, in $\mathcal{A}^h$. Suppose $\rho(J_0) = 1$, where $J_0 = \mathbb{R}^h \times [j_0, \infty)$. Let ρ_0^* be the associated spectral distribution on $\mathbb{R}^h$. Let $\beta\colon [0, \infty) \to \mathcal{A}^h$ vary like γ^t. Assume β is continuous, and $t \mapsto y(t) = \tilde{\beta}(t)(j_0)$ is a homeomorphism from $[0, \infty)$ onto $[y_0, y_\infty)$. Let*

$F = 1 - T$ be a df such that $T(y(t)) \sim e^{-t}$ for $t \to \infty$. Let μ be the typical distribution associated with ρ_0^*, β and F:

$$\mu(dx, dy) = \mu_y(dx)F(dy), \quad (\mu_y(dx), y) = \begin{cases} \beta(t)(\rho_0^*, j_0) & y = y(t) \in [y_0, y_\infty) \\ \beta(0)(\rho_0^*, j_0) & y < y_0. \end{cases}$$

Then $\rho_t := e^t \beta(t)^{-1}(\mu) \to \rho$ weakly on $\mathbb{R}^h \times [v, \infty)$ as $t \to \infty$, for $v > j_* =$ $\inf \tilde{\gamma}^{-n}(j_0)$.

Under appropriate conditions the densities will also converge. It suffices that the densities of the vertical component converge. Convergence $\alpha_n(U^*) \Rightarrow \alpha_0(U^*)$ implies convergence of the densities if U^* has a continuous density. See Lemma 15.9 below.

14.15 Approximation by typical distributions. We now have the following scheme:

$$\pi \longrightarrow \beta \longrightarrow \mu.$$

The question is: How close is the typical distribution μ to the original distribution π? The answer is simple.

Definition. Let μ be a Radon measure on $\mathcal{X} \times \mathbb{R}$, where $\mathcal{X}$ is a locally compact separable metric space. Let μ^* be a probability measure on $\mathcal{X}$, and let λ denote Lebesgue measure on $\mathbb{R}$. We say that μ is *asymptotic* to $\mu^* \times \lambda$, if

$$\tau^{-t}(\mu) \to \mu^* \times \lambda \text{ vaguely on } \mathcal{X} \times \mathbb{R}, \quad t \to \infty,$$

where $\tau^t, t \in \mathbb{R}$, is the one-parameter group of *vertical upward translations*: $\tau^t(x, s) = (x, t + s)$, and if

$$M(t_n + s_n) - M(t_n) \to s, \quad t_n \to \infty, \; s_n \to s > 0, \tag{14.17}$$

where $M(t) = \mu(\mathcal{X} \times [0, t])$ for $t \geq 0$.

The space $\mathcal{X}$ may be compact. In the theory of heavy-tailed coordinatewise maxima on $[0, \infty)^d$ it is a simplex; in the theory of exceedances over elliptic thresholds it is the unit sphere ∂B. In our case $\mathcal{X} = \mathbb{R}^h$. Set

$$d\mu_1(x, v) = e^{-t} d\mu(x, t), \quad d\nu(x, t) = \mu^*(x) e^{-t} dx dt.$$

By the lemma below

$$\nu_t := e^t \tau^{-t}(\mu_1) \to \nu \text{ weakly on } J = \mathbb{R}^h \times [c, \infty), \quad t \to \infty, \; c \in \mathbb{R}.$$

Convergence need not be steady. It is possible that $\nu_t(J) \to \infty$ whenever the halfspace H is not horizontal, even when μ^* is standard Gaussian. Compare Example 9.14.

Lemma 14.38. *Let F be a df on $\mathbb{R}$ and set $dM(y) = e^y dF(y)$. Then $1-F(y) \sim e^{-y}$ for $y \to \infty$ if and only if* (14.17) *holds.*

Proof. By partial integration. □

We fix an excess measure ρ with symmetries $\gamma^t \in \mathcal{A}^h$, $\gamma^t(\rho) = e^t\rho$, for $t \in \mathbb{R}$. Let $\rho(J_0) = 1$ where $J_0 = \mathbb{R}^h \times [j_0, \infty)$. Let U^* have distribution ρ^*.

Suppose $\beta\colon [0,\infty) \to \mathcal{A}^h$ is continuous and varies like γ^t, and $t \mapsto y(t) := \tilde{\beta}(t)(j_0)$ is a homeomorphism of $[0,\infty)$ onto $[y_0, y_\infty)$. Let μ be the associated typical distribution, $\mu(dx, dy) = \mu^*(dx, y)F(dy)$ where we specify the df of the vertical coordinate by $1-F(y(t)) = e^{-t}$ for $t \geq 0$. Define $\Psi\colon \mathbb{R}^h \times [y_0, y_\infty) \to \mathbb{R}^h \times [0,\infty)$ by

$$\Psi(\beta(t)(u, j_0)) = (u, t), \quad (u,t) \in \mathbb{R}^h \times [0, \infty).$$

Then Ψ is a homeomorphism, since $t \mapsto y(t) = \tilde{\beta}(t)(j_0)$ is a homeomorphism by assumption, and $\beta(t)$ for fixed t is an (invertible) affine transformation, from $\mathbb{R}^h \times \{j_0\}$ to $\mathbb{R}^h \times \{y(t)\}$, which depends continuously on t. The image of μ is the product measure $\rho^*(du)e^{-v}dv$ on H_+. The image of $e^t\mu(dx, dy(t))$ is $\rho^*(du)dv$ on H_+. The transformation Ψ preserves horizontal hyperplanes, so it suffices to check the result for the vertical coordinate, and to check the affine transformations in the horizontal coordinate. As in the case of non-linear transformations for coordinate maxima, one may regard Ψ as the non-linear transformation which trivialises the affine normalizations.

Theorem 14.39. *Under the assumptions above*

$$\nu(dudt) := e^t\Psi^{-1}(\pi)(dudt)$$

on $\mathbb{R}^{h+1}$ is asymptotic to $\rho_0^* \times \lambda$.

Proof. The vertical translations τ^t on $\mathbb{R}^{h+1}$ are equivalent to the one-parameter group of affine transformations γ^t on $O = \mathbb{R}^h \times (j_*, j^*)$. The relation is formalized by the map $\Phi\colon (u,t) \mapsto \gamma^t(u, j_0)$. This map is the identity if γ^t is the one-parameter group of vertical translations on $\mathbb{R}^{h+1}$ and $j_0 = 0$. In the equivalence the excess measure ρ with spectral measure ρ^* corresponds to the product measure $\rho^* \times \varepsilon$, where ε is the exponential measure on $\mathbb{R}$ with density e^{-t}. So define

$$\Xi\colon (\gamma^t(u, j_0)) \mapsto \beta(t)(u, j_0) \in \mathbb{R}^\times[y_0, y_\infty).$$

The typical distribution μ is the image of $d\rho_0 = 1_{J_0}d\rho$ under the map Ξ. Observe that

$$\beta(t_n)^{-1}(\Xi(\gamma^{t_n}(w_n)) \to w, \quad t_n \to \infty,\ w_n \to w.$$

So $S_n := \beta(t_n)^{-1} \circ \Xi \circ \gamma^{t_n} \to \text{id}$ uniformly on compact subsets of O. Define ν_0 on $O \cap \{v \geq j_0\}$ by $\pi = \Xi(\nu_0)$, and set $\nu_t = e^t\gamma^{-t}(\nu_0)$. Then

$$S_n(\nu_{t_n}) = \beta(t_n)^{-1}(\Xi(\gamma^{t_n}(\nu_{t_n}))) = e^{t_n}\beta(t_n)^{-1}(\Xi(\nu_0)) = e^{t_n}\beta(t_n)^{-1}(\pi) \to \rho$$

vaguely on O implies that $\nu_{t_n} \to \rho$ vaguely on O. As usual let ˜ denote the vertical component. Then $\widetilde{\Xi}(\tilde{\gamma}^t(j_0)) = \tilde{\beta}(t)(j_0) = y(t)$. Hence $\tilde{\nu}_{t_n} = e^{t_n}\tilde{\gamma}^{-t_n}(\tilde{\nu}_0) \to \tilde{\rho}$ weakly on $[c, \infty)$ for each $c > j_*$. □

We now have a simple recipe for creating the distributions $\pi \in \mathcal{D}^h(\rho)$ for a given excess measure ρ. Choose β to vary like γ^t as in Section 15.5. Construct Ψ as above, and choose a measure μ asymptotic to $\rho^* \times \lambda$. The restriction ν of the exponential transform $e^{-t}d\mu(u,t)$ to H_+ is a finite measure. Let π be a probability measure on $\mathbb{R}^d$ which agrees with $\Psi(\nu)$ on some horizontal halfspace of positive mass.

15 Horizontal thresholds – examples

15.1 Domains for exceedances over horizontal thresholds. Not much is known about the domains of attraction $\mathcal{D}^h(\rho)$ for exceedances over horizontal thresholds for the various symmetry groups. The following pages may be compared to a 17th century map of Africa. We give examples and some specific results. Each subsection presents a number of unrelated examples. Our aim is to kindle the interest of the reader. We restrict ourselves to three simple groups: Vertical translations, and scalar contractions and expansions. The final subsection treats regular variation for matrices in $\mathcal{A}^h$.

15.2 Vertical translations. Assume the symmetries of the excess measure are vertical translations. The generator C is zero except for the entry $C_{10} = 1$. The translation $(u, v) \mapsto (u, v + t)$ maps ρ into $e^t\rho$. Orbits are vertical lines. Assume $\rho(H_+) = 1$. Limit vectors $W = (U, V)$ have a simple form: V is standard exponential and independent of U. The distribution of the horizontal component U is the spectral measure ρ^*. Spectral stability holds if ρ^* has exponential tails. For any halfspace $H = \{v \geq c^T u\}$ of finite mass $\rho(H)$, one may define the spectral measure ρ_c^* on $\mathbb{R}^h$ by

$$\rho_c^*(A) = \rho(H \cap (A \times \mathbb{R})) = \int_A \int_{c^T u}^{\infty} e^{-t} d\rho^*(u) = \int_A e^{-c^T u} d\rho^*(u), \quad A ß \mathbb{R}^h.$$

The probability measure $\rho_c^*/\rho_c^*(H)$ with direction c is the *Esscher transform* of ρ^*. The symmetry of the Gauss-exponential high risk limit law is reflected in the stability of Gaussian distributions under exponential tilting.

Proposition 15.1. *The family of probability measures $\mu_c = \rho_c^*/\rho_c^*(H)$, $c \in \mathbb{R}^h$, $\rho(H) < \infty$, is the natural exponential family generated by ρ^*.*

In order to explore the domain of attraction we start with simple distributions. Let f be a density of the form

$$f(x, y) = f_y(x)\tilde{f}(y)$$

where $\tilde{f}$ lies in the domain $\mathcal{D}^+(0)$ of the univariate exponential law, and the conditional densities f_y all have the same shape. One could take $\tilde{f}$ standard exponential, and f_y Gaussian. From a geometric point of view it is simpler to assume that U is uniformly distributed on the unit ball in $\mathbb{R}^h$. Then for each $y \geq 0$ the density f_y describes the *uniform distribution* on an ellipsoid E_y in $\mathbb{R}^h$.

1) Assume that U is uniformly distributed on the unit disk in $\mathbb{R}^h$. What does the distribution of $Z = (X, Y) \in \mathcal{D}^h(W)$ look like? For simplicity assume Y is standard exponential, and, conditional on $Y = y \geq 0$, the component X is uniformly distributed on a translate $\mu(y) + B$ of the unit disk. The random vector Z lives on a curved tube T, whose sections are unit disks centered on the curve $(\mu(t), t)$, $t \geq 0$. We assume that this curve is continuous. What behaviour is allowed if we want the high risk scenarios Z^y, properly normalized, to converge?

Example 15.2. Assume $d = 3$. Curves $\mu(t)$ which converge to a point $\mu_\infty \in \mathbb{R}^2$ for $t \to \infty$ are allowed, as are curves which are asymptotically vertical, $\dot{\mu}(t) \to 0$. But also curves with any other limit direction, $\dot{\mu}(t) \to q$ for $q \in \mathbb{R}^2$. In affine coordinates the vertical axis need not be perpendicular to the horizontal coordinate plane. Now consider the curve

$$\mu(t) = t^{3/2}(\cos(t^{1/5}), \sin(t^{1/5})).$$

It spirals off to infinity; the velocity increases without bound. However, the exponents in the formula on the right have been chosen so that the second derivative vanishes in infinity: $\ddot{\mu}(t) \to 0$. ◊

Proposition 15.3. *Let $W = (U, V)$ have density $h(u, v) = g(u)e^{-v}$ on H_+. Let Y be standard exponential and let $Z = (X, Y)$ have density $f(x, y) = f_y(x)e^{-y}$ on H_+. Let $\mu\colon [0, \infty) \to \mathbb{R}^h$ be a C^2 curve. Suppose $g_y(u) = f_y(\mu(y) + u) \to g(u)$ a.e. on $\mathbb{R}^h$. If $\ddot{\mu}$ vanishes in ∞ then $Z \in \mathcal{D}^h(W)$ with normalizations*

$$\alpha_y(u, v) = (u + \mu(y) + v\dot{\mu}(y), v + y) \quad y \geq 0.$$

Proof. The normalized high risk scenario $W_y = \alpha_y^{-1}(Z^y)$ has density

$$\begin{aligned} h_y(u, v) &= f(\alpha_y(u, v))/e^{-y} \\ &= f(u + \mu(y) + v\dot{\mu}(y), v + y)/e^{-y} \\ &= g_{v+y}(u + \mu(y) + v\dot{\mu}(y) - \mu(v + y))e^{-v} \to g(u)e^{-v} \end{aligned}$$

since $\ddot{\mu} \to 0$ implies $\mu(y + v) - \mu(y) - v\dot{\mu}(y) \to 0$ by an appropriate Taylor expansion. The f_y are densities on $\mathbb{R}^h$. Hence so are the g_y, and $g_y \to g$ in $\mathbf{L}^1$ by

Scheffé's Theorem. It follows that $g_t(\gamma_t(u)) \to g$ in $\mathbf{L}^1$ for affine transformations $\gamma_t \to \mathrm{id}$. Hence $h_y \to h$ in $\mathbf{L}^1(H_+)$. Now apply Proposition 14.36. □

2) The vertical coordinate of a vector $Z = (X, Y)$ in the domain of W need not be exponential. It suffices that $Y \in \mathcal{D}^+(0)$. Then $\mathbb{P}\{Y \geq y\} \sim e^{-\psi(y)}$ for a function ψ which satisfies (6.4), and the random variable $\tilde{Y} = \psi(Y)$ has a tail which is asymptotically standard exponential, $\mathbb{P}\{\tilde{Y} \geq y\} \sim e^{-y}$. The transformation $(X, Y) \mapsto (X, \tilde{Y})$ does not affect the conditional distributions of X given y, it only rearranges them. Since we are only concerned with horizontal halfspaces we could check that $\tilde{Z} = (X, \tilde{Y}) \in \mathcal{D}^h(W)$ and then conclude that $Z \in \mathcal{D}^h(W)$, as in the case of coordinatewise maxima, see Theorem 7.16. The example below shows that this procedure does not work. It is possible that $\tilde{Z} \in \mathcal{D}^h(W)$ and Z not.

Example 15.4. Given $\tilde{Y} = t$ let X be normal $N(\mu(t), 1)$ where $\mu(t) = 4t^{7/4}/7$. Then $\ddot{\mu}(t) = 3/(4t^{1/4}) \to 0$ implies $\tilde{Z} \in \mathcal{D}^h(W)$. Choose $\psi(y) = (\log y)^2$ for $y \geq y_0$. Then $a(y) = y/(2\log y)$ and $a'(y) \sim 1/(2\log y) \to 0$. Since X is normal $N(\mu(\psi(y)), 1)$ given $Y = y$, the high risk scenarios Z^y may be normalized to converge in distribution to the Gauss-exponential vector W if the curve $y \mapsto p(y) = \mu(\psi(y))$ is asymptotically linear over intervals of the order of $a(y)$. Write

$$p(y + va(y)) - p(y) - va(y)p'(y) = v^2a^2(y)p''(y + \omega a(y))/2.$$

By straightforward calculations the right hand side does not vanish in y_∞. Indeed $a^2p'' = \ddot{\mu} - \dot{\mu}a'$ and $\dot{\mu}a' \sim \sqrt{\log y}/2$. Since $a(y+\omega a(y)) \sim a(y)$ and $\ddot{\mu}(\psi(y)) \to 0$ the quadratic term can not be neglected. ◊

3) Again assume U is *uniformly distributed* on the unit disk B in $\mathbb{R}^h$. Then W lives on the vertical tube $B \times [0, \infty)$. What can one say about the distribution of $Z = (X, Y) \in \mathcal{D}^h(W)$?

Let Y have df $F \in \mathcal{D}^+(0)$ with upper endpoint y_∞. There exists a C^1 function $e\colon [y_0, y_\infty) \to (0, \infty)$ whose derivative vanishes in y_∞, and which also vanishes itself if y_∞ is finite. This is the scale function associated with the tail function $R = 1 - F$ in Section 6. It is determined up to asymptotic equality by

$$R(y_n + s_n e(y_n))/R(y_n) \to e^{-s}, \quad y_n \to \infty,\ s_n \to s \in \mathbb{R}. \tag{15.18}$$

Assume X, conditionally on $Y = y$, is uniformly distributed on an open ellipsoid $\mu(y) + E_y ß\mathbb{R}^h$ centered in $\mu(y)$. We assume that $\mu(y)$ and E_y are continuous in $y \in [y_0, y_\infty)$, and that $Y \in [y_0, y_\infty)$ a.s. If Y has a continuous density $\tilde{f}(y)$, then Z lives on the curvaceous tube $Tß\mathbb{R}^h \times [y_0, y_\infty)$ with horizontal sections $\mu(y) + E_y$ at level y, and Z has density

$$f(x, y) = 1_T(x, y)h(y), \quad h(y) = \tilde{f}(y)/|E_y|.$$

First assume $\mu \equiv 0$. The function $e(y)$ is a scale function for the vertical coordinate. In the same way E_y is a scale for the horizontal coordinate. If we want T to converge to the standard tube $B \times \mathbb{R}$ under affine normalizations which map $\mu(y) + E_y$ onto B and vertical intervals $[y, y + e(y))$ into intervals of length one asymptotically, then we have to impose the condition

$$E_{y'_n} \sim E_{y_n}, \quad y'_n = y_n + s_n e(y_n),\ y_n \to y_\infty,\ s_n \to s \geq 0. \tag{15.19}$$

The function $y \mapsto |E_y|$ is flat for the scale function e. If $\tilde{f}$ is asymptotic to a von Mises function, then so is h, since the quotient is flat.

In a more geometric vein one might start with the functions $e(y)$ and E_y, or just a sequence of positive numbers e_n and centered open ellipsoids E^n in $\mathbb{R}^h$, with the condition that $e_{n+1} \sim e_n$, and $E^{n+1} \sim E^n$. Define y_n such that $y_{n+1} - y_n \sim e_n$, and then define the functions $e(y)$ and E_y by interpolation such that $e(y_n) = e_n$ and $E_{y_n} = E^n$. The scale sequences e_n and E_n may be chosen independently, The curve $\mu(y)$ depends on both the vertical and the horizontal scaling as we shall see below. That explains why a normalization of the vertical coordinate may throw a distribution π out of the domain $\mathcal{D}^h(\rho)$, as we saw in Example 15.4 above.

In order to link this geometric set-up to the more algebraic set-up of affine normalizations and regular variation, we need to write the continuous curve of ellipsoids as $E_y = A_y(B)$ for a continuous curve of linear normalizations. We also want $A_{y'_n} \sim A_{y_n}$ to hold whenever $y_n \to y_\infty$, and $y'_n = y_n + s_n e(y_n)$, with $s_n \to s \geq 0$. This is possible by the lemma below.

Lemma 15.5. *Suppose $E(t)$, $t \geq 0$, are open centered ellipsoids in $\mathbb{R}^d$, depending continuously on t, such that*

$$E(t_n + s_n) \sim E(t_n), \quad t_n \to \infty,\ s_n \to s \in \mathbb{R}.$$

There exists a continuous curve $A\colon [0, \infty) \to \mathrm{GL}(d)$ which varies like the identity such that $E(t) = A(t)(B)$.

Proof. The correspondence $E = A(B)$ between open centered ellipsoids E and symmetric matrices A is a homeomorphism. (It is clear that $A_n \to A_0$ implies $E_n \to E_0$. Conversely $E_n \to E_0$ implies that (A_n) is relatively compact. Any limit point A satisfies $E_0 = A(B)$. Since this equation determines A there is only one limit point, and $A_n \to A$.) One could write $E(t) = A(t)(B)$ with $A(t)$ symmetric. The curve A then is continuous, but it is not clear whether it varies slowly. So we shall construct the curve A piecewise. Choose $A(0)$ symmetric such that $A(0)(B) = E(0)$. If $A(t)$ has been constructed for $t \leq n$, define $A(n+\theta) = A(n)Q_n(\theta)$ for $0 \leq \theta \leq 1$ with $Q_n(\theta)$ symmetric such that $F_n(\theta) := A(n)^{-1}(E(n+\theta)) = Q_n(\theta)(B)$. Since $F_n(\theta_n) \to B$ for $\theta_n \in [0, 1]$, by continuity $Q_n(\theta_n) \to I$. In particular $\theta_n = 1$ gives $A(n)^{-1}A(n+1) \to I$. For $t_n \to \infty$, and $|s_n| \leq m$, the quotient $A(t_n)^{-1}A(t_n + s_n)$ tends to I as a product of $m + 2$ factors which tend to I. □

Suppose $\mu \equiv 0$. Suppose $F \in \mathcal{D}^+(0)$, with upper endpoint $y_\infty \leq \infty$, and finite lower endpoint y_0, has density $\tilde{f}$ and rate function $e\colon [y_0, y_\infty) \to (0, \infty)$. Suppose the ellipsoids E_y, depend continuously on $y \in [y_0, y_\infty)$, and satisfy (15.19). Let T denote the curvaceous tube with horizontal cross sections E_y, and let $Z = (X, Y)$ have density $f(x, y) = \tilde{f}(y)1_T(x, y)/|E_y|$. Geometrically it is clear that one may introduce local coordinates centered in $(0, y)$ such that T in these coordinates looks like the standard tube $B \times \mathbb{R}$ in $\mathbb{R}^{h+1}$. Write this as $\alpha_y^{-1}(T) \to B \times \mathbb{R}$, where $\alpha_y(u, v) = (A_y u, y + ve(y))$, and $A_y(B) = E_y$. The condition $F \in \mathcal{D}^+(0)$, with rate function $e(y)$, then ensures that $\alpha_y^{-1}(\pi)/(1 - F(y)) \to \rho$ vaguely for $y \to y_\infty$ where ρ has density e^{-v} on $B \times \mathbb{R}$.

What conditions should we impose on the curve μ? The normalization α_y^{-1} should map $\mu(y)$ into the origin. The normalized curve should be close to the vertical axis in the neighbourhood of the origin. This means that the difference between $\mu(y + e(y)) - \mu(y)$ and $\mu(y + 2e(y)) - \mu(y + e(y))$ should be small in relation to the ellipsoid E_y. A simple condition is

$$e(y)^2 \mu''(y) = o(E_y), \quad y \to y_\infty. \tag{15.20}$$

Here we write $x_n = o(E_n)$ for a sequence of open ellipsoids if for any $\varepsilon > 0$ eventually $x_n \in \varepsilon E_n$.

Proposition 15.6. *Suppose* (15.20) *holds. The probability distribution π with the density f introduced above lies in the domain $\mathcal{D}^h(\rho)$ of the excess measure ρ with density $g(u, v) = 1_B(u)e^{-v}/|B|$ on $\mathbb{R}^{h+1}$.*

Proof. Let $y_n \to y_\infty$. Let $A_n \in \mathrm{GL}(h)$ such that $A_n(B) = E_{y_n}$. Let $\mu_n = \mu(y_n)$, $\sigma_n = \mu'(y_n)$, and $e_n = e(y_n)$. Define $\alpha_n \in \mathcal{A}$ by

$$\alpha_n(u, v) = (A_n u + \mu_n + \sigma_n v, y_n + e_n v), \quad (u, v) \in \mathbb{R}^{h+1}.$$

Let g_n be the density of $\rho_n = e^{t_n}\alpha_n^{-1}(\pi)$, where t_n satisfies $\mathbb{P}\{Y \geq y_n\} = e^{-t_n}$. Convergence of the vertical component holds by assumption. Consider the horizontal component: $g_n^*(\cdot, v)$ is uniformly distributed on an open ellipsoid $p + E$ in $\mathbb{R}^h$. It suffices to show that for any sequence $v_n \to v$ the ellipsoids $p_n + E^n$ tend to the unit ball B in $\mathbb{R}^h$. Now $E^n = A_n^{-1}(E_{y_n'})$ where $y_n' = y_n + v_n e_n$. Asymptotic equality of convex sets does not depend on coordinates. So $E_{y_n'} \sim E_{y_n}$ implies $E^n \sim B$. It remains to show that $p_n = o(E_n)$, or, equivalently, $\mu(y_n') - \mu_n - v_n\sigma_n = o(E_n)$. This follows by a second order Taylor–McLaurin expansion. □

Suppose, conditionally on $Y = y$, the vector X lives on the boundary of the ellipsoid $\mu(y) + E_y$, and is uniformly distributed. (This distribution is defined as the image of the *uniform distribution* on the unit sphere ∂B under the affine transformation which maps B onto $\mu(y) + E_y$.) In that case $e^{t_n}\alpha_n^{-1}(\pi) \to \rho$ where now ρ is the product of the uniform distribution on ∂B and the exponential measure with density e^{-v}. The vertical component Y may have any df $F = 1 - R \in \mathcal{D}^+(0)$, provided (15.18) holds. If ρ^* is a spherically symmetric probability distribution on $\mathbb{R}^h$, and ρ is the product of ρ^* and the exponential measure on $\mathbb{R}$, and if Y has df $F = 1-R$, as above, and X conditionally on $Y = y$ is distributed like $\mu(y)+A_y(U^*)$ where U^* has distribution ρ^*, and $A_y(B) = E_y$, then $Z \in \mathcal{D}^h(\rho)$. More precisely:

Theorem 15.7. *Let $U^* \in \mathbb{R}^h$ have a spherically symmetric probability distribution ρ^*. Let $\rho = \rho^* \times \varepsilon$ where ε is the measure with density e^{-v} on $\mathbb{R}$. Let $Z = (X, Y)$ be a random vector in $\mathbb{R}^{h+1}$ with distribution π. Let Y have df $F = 1 - R \in \mathcal{D}^+(0)$ with upper endpoint $y_\infty \le \infty$, and $F(y_0) = 0$. Let $y_n \to y_\infty$ such that $R(y_{n+1})/R(y_n) \to 1/e$. There exists a C^1 function $f : [y_0, y_\infty) \to (0,\infty)$ whose derivative vanishes in y_∞, and which vanishes in y_∞ itself if y_∞ is finite, such that $f(y_n) \sim y_{n+1} - y_n$. Let $f(y)$ be any such function. Suppose X conditionally on $Y = y$ for $y \in [y_0, y_\infty)$ is distributed like $\mu(y) + A_y U^*$. If the ellipsoids $E_y = A_y(B)$ in $\mathbb{R}^h$ satisfy (15.19), and if μ is C^2 and $f^2(y)\mu''(y) = o(E_y)$ for $y \to y_\infty$, then $\pi \in \mathcal{D}^h(\rho)$.*

Now assume that ρ has a density. Suppose ρ^* has a Borel measurable density g^* on $\mathbb{R}^h$. *Spherical symmetry* implies that $g^*(u) = g_0^*(\|u\|)$ for some non-negative Borel function g_0^* on $[0,\infty)$, which satisfies the integrability condition

$$hB(h)\int_0^\infty g_0^*(s)s^{h-1}ds = 1, \quad b(h) = |B|/h. \tag{15.21}$$

Here $B(h)$ is the volume of the unit ball in $\mathbb{R}^h$.

We want to describe the density of $Z = (X, Y)$ in a similar fashion.

Definition. A function $\varphi\colon \mathbb{R}^h \times [y_0, y_\infty) \to [0,\infty)$ is *homogeneous-elliptic*, and we write $\varphi \in \mathcal{HE}$, if φ is continuous, and

$$\begin{aligned} &\varphi(rx, y) = r\varphi(x, y), \quad x \in \mathbb{R}^h,\ r \ge 0,\ y \in [y_0, y_\infty) \\ &E_y = \{x \in \mathbb{R}^h \mid \varphi(x, y) < 1\} \in \mathcal{E}(h), \quad y \in [y_0, y_\infty). \end{aligned} \tag{15.22}$$

Here $\mathcal{E}(h)$ is the set of all centered open ellipsoids in $\mathbb{R}^h$. We write $\varphi \in \mathcal{HE}_0$ if in addition

$$E_{y_n'} \sim E_{y_n}, \quad y_n \to y_\infty,\ y_n' - y_n = O(f(y_n))$$

for a given C^1 function $f : [y_0, y_\infty) \to (0,\infty)$ whose derivative vanishes in y_∞, and which vanishes in y_∞ itself if y_∞ is finite.

Typically $\varphi(x, y) = \|A_{\psi(y)}x\|$ where $A\colon [0,\infty) \to \mathrm{GL}(h)$ varies like the identity and $\psi\colon [y_0, y_\infty) \to [0,\infty)$ is a C^2 homeomorphism with a positive derivative $\psi'(y) \sim 1/f(y)$. Actually the function φ is completely determined by the curvaceous tube with horizontal sections E_y.

Theorem 15.8. *Let ρ be an excess measure with density $g(u,v) = g^*(\|u\|)e^{-v}$ where $0 < \int g^*(r)r^{h-1}dr < \infty$. Let $\psi\colon [y_0, y_\infty) \to [0,\infty)$ be a C^2-diffeomorphism with $f'(y) \to 0$ for $y \to y_\infty$, where $f(y) = 1/\psi'(y)$. Let $L\colon [y_0, y_\infty) \to (0,\infty)$ be flat for the scale function f, and let $\varphi \in \mathcal{H}\mathcal{E}_0$ for the function $f(y)$. Let $\mu\colon [y_0, y_\infty) \to \mathbb{R}^h$ be C^2 such that*

$$\varphi(f(y)^2\mu''(y), y) \to 0, \quad y \to y_\infty.$$

Let $Z = (X, Y) \in \mathbb{R}^{h+1}$ have distribution π on $\mathbb{R}^h \times [y_0, y_\infty)$ with density

$$f(x,y) = g^*(\varphi(x - \mu(y), y))L(y)e^{-\psi(y)}dxdy/C, \quad y \geq y_0.$$

Then $\pi \in \mathcal{D}^h(\rho)$. If g is continuous, the normalized densities g_y tend to g uniformly on compact sets.

Proof. Let $E_y = \{x \mid \varphi(x,y) < 1\}$. If $g^* = 1_{[0,1)}$ then $f(x,y) = 1_T(x,y)\tilde{f}(y)$ with T the curvaceous tube constructed above, and $\tilde{f}$ the density of a df $\tilde{F} = 1 - R$ where R satisfies (15.18). So $\pi \in \mathcal{D}^h(\rho)$ by the previous proposition. The last statement follows from the lemma below. □

Lemma 15.9. *Let $\alpha_n^{-1}(\pi) \to \pi$ weakly. Suppose π has a density f which is continuous and positive on the open set O and vanishes outside O. Let f_n be the density of $\alpha_n^{-1}(\pi)$. Then*

$$f_n(z_n) \to f(z), \quad z_n \to z,\ z \notin \partial O.$$

Proof. By the Convergence of Types Theorem we may choose symmetries σ_n of π such that $\beta_n = \alpha_n\sigma_n \to \mathrm{id}$. Then $f_n = |\det\beta_n| f \circ \beta_n$. Set $w_n = \beta_n^{-1}(z_n)$. Then $w_n \to z$, and $f_n(z_n) = c_n f(w_n) \to f(w)$ since $c_n = |\det\beta_n| \to 1$. □

Example 15.10. Suppose the vertical coordinate Y has a standard *Gaussian distribution*. Then $Y \in \mathcal{D}^+(0)$. One may write the density of Y as $g(y) = L(y)e^{-\psi(y)}$ where $e^{-\psi}$ is a *von Mises function*, and L is flat. The scale function on the vertical coordinate is $f(y) = 1/\psi'(y) = 1/y$. Take $[y_0, y_\infty) = [1,\infty)$. Let $d = 3$ and assume that the ellipses E_y are centered disks of radius $r(y)$. First assume $r(y) \equiv 1$. What are the conditions on the curve μ? It is possible that $\|\mu(y)\|$ grows faster than y^3 for $y \to \infty$. Now vary $r(y)$. How fast can the ellipses E_y grow? It is possible that $r(y) = e^y$ since this function satisfies $r(y_n')/r(y_n)$ for $y_n \to \infty$, $y_n' - y_n = O(1/y_n)$. If the disks E_y expand at this rate, the center $\mu(y)$ may grow even faster. The curve $\mu(y) = ye^y(\cos y, \sin y)$ is allowed. ◊

15.3 Cones and vertices. In this section we continue our geometric analysis. We study the asymptotic behaviour of the distribution π of a random vector Z in the neighbourhood of a vertex of the convex support. Bounded domains are not typical for risk analysis. However a loss function may have a pole at a boundary point. In biology and engineering a stress variable like temperature may be limited to a bounded domain. The analysis below exhibits the close relation between the algebraic condition of regular variation and geometric concepts like convex sets and cones. There are relations with min-stable distributions with exponential marginals on $(0, \infty)$, and with *copulas*.

Assume the vertex z_0 is the origin. Let $F(y) = \mathbb{P}\{Y \leq y\}$ vary regularly with exponent $\lambda > 0$ for $y \to 0+$. Here $Y = \eta(Z)$, and η denotes the vertical coordinate. Introduce the high risk scenarios

$$Z^y = Z^{H_y}, \quad H_y = \{\eta \leq y\},\ y > 0.$$

Do there exist linear transformations α_y mapping $J = \{v \leq 1\}$ onto H_y such that the normalized high risk scenarios converge, equivalently

$$\alpha_y^{-1}(\pi)/F(y) \to \rho \ \text{ vaguely on } \mathbb{R}^d, \quad y \to 0+. \tag{15.23}$$

In this section we assume that the excess measure ρ is symmetric for scalar contractions

$$\gamma^t(\rho) = e^t\rho, \quad \gamma^t w = e^{-\theta t}w, \quad t \in \mathbb{R},\ w \in \mathbb{R}^d,\ \theta = 1/\lambda = -\tau > 0. \tag{15.24}$$

These symmetries are preserved under linear transformations. The parameter $\lambda = 1/\theta$ is the exponent of regular variation of the df of the vertical component in zero, and $\tau = -1/\lambda < 0$ the Pareto parameter. The convex support C of the Radon measure ρ is a closed cone, since the support is invariant under the scalar contractions γ^t. We shall generally assume that the cone is proper, more precisely that $C \cap \{\eta \leq 1\}$ is compact. It is then of interest to know under what conditions the normalized sample clouds converge to the limiting Poisson point process on C weakly on halfspaces H which intersect the cone in a bounded non-empty set. Probability distributions for which this holds are *completely steady*.

In this section we have flipped the vertical axis. We are looking at lower halfspaces, at minima, and at exceedances below horizontal levels. This has the advantage that Z and ρ live on the positive open halfspace $\mathbb{R}^h \times (0, \infty)$. Two basic examples to keep in mind are:

ρ is *Lebesgue measure* on the cone C, and Z is uniformly distributed. Here $\lambda = d$.

ρ is the *exponent measure* of a min-stable df on $[0, \infty)^d$ with exponential marginals, compare Section 7.4. Now $\eta = z_1 + \cdots + z_d$, Cß$[0, \infty)^d$, and $\lambda = 1$.

1) Take $d = 2$ and let ρ be *Lebesgue measure* on $[0,\infty)^2$. Let Z be uniformly distributed on the open set

$$U = \{-\varphi_0(y) < x < \varphi_1(y), 0 < y < 1\}\text{ß}\mathbb{R} \times (0,\infty)$$

where $-\varphi_0 < \varphi_1$ are continuous functions which vanish in the origin. Let Z^y denote the high risk scenario associated with the halfspace $H^y = \{\eta \le y\}$. For simplicity assume $\varphi_0 = \varphi_1 = \varphi$. Let $\alpha_y(u,v) = (\varphi(y)u, yv)$. Then $\alpha_y^{-1}(Z^y) \Rightarrow W$ where W is uniformly distributed on the triangle $|u| < v < 1$ if φ varies regularly in $0+$ with exponent 1. The excess measure is ρ. It is possible that $\varphi(y) = o(y)$ for $y \to 0$, for instance if $\varphi(y) = y/|\log y|$. In that case U has the form of a sharp thorn. The vector Z then is not steady, since the cone $v \ge \varepsilon|u|$ does not contain $\alpha_y^{-1}(U)$ for y small. If $\varphi(y)/y \to c > 0$ then φ automatically varies regularly with exponent 1 and Z is steady; in fact Z is completely steady if U is contained in the cone $y \ge |x|/c$. If not, then there exists a halfspace $H_0 = \{y \ge x/a\}$ with $a > c$ which intersects U in a set $U_1\text{ß}\{y > \delta\}$ of positive measure. Slight variations H of H_0 yield high risk scenarios Z^H which are uniformly distributed on the disjoint union of a small neighbourhood of the origin in U, and a subset of U_1 of the same area. The distribution of Z^H is not close to the uniform distribution on a triangle. So Z is not completely steady. It is also possible that $\varphi(y)/y \to \infty$ for $y \to 0$. As long as $\varphi(y)/y$ is increasing the vector Z is completely steady. Here are two examples. In both ρ is Lebesgue measure on the cone $\{v > |u|\}$ in the plane.

Example 15.11. $\mathcal{D}^h(\rho)$ contains a vector $Z \in \mathbb{R}^2$ which is uniformly distributed on a bounded open convex set whose boundary is continuously differentiable. The vector Z is completely steady. Take $U = \{|x| < yL(y),\ 0 < y < 1\}$ where $L(y) = |\log(y - y^2)|$. ◊

Example 15.12. The domain $\mathcal{D}^h(\rho)$ contains a completely steady vector Z with a continuous density f on $\mathbb{R}^2$ such that $\{f > 0\} = \mathbb{R} \times (0,\infty)$.

Proof. Let f_0 be the density $c1_U/L(y)$, with U as above, and let f_1 agree with f_0 on U, vanish for $|x| > yL(y) + y^2$, $y \in (0,1)$, and be defined by linear interpolation on the remaining two intervals $yL(y) \le |x| \le yL(y) + y^2$. Then f_0 and $c_1 f_1$ lie in $\mathcal{D}^h(\rho)$ with the same normalizations α_y as the uniform distribution on U. The density $c_1 f_1$ is continuous on the plane, and so is $f = (c_1 f_1 + f_2)/2$ where $f_2(x,y) = y^4 e^{-y} e^{-|x|}/48$ on H_+. Let $J = \{v \le b + au\}$ with $0 \le a < 1$ and $b > 0$. The corresponding halfspaces for Z are $H_s = \{y \le bs + asL(s)x\}$. Take s so small that $2b\sqrt{s} < 1$ and $2sL(s)s^{-1/3} < \sqrt{s}$. Then H_s is contained in the union of $\{y \le \sqrt{s}\}$ and $\{x \ge s^{-1/3}\}$. Hence

$$\int_{H_s} f_1(z)dz \le s^{5/2} + e^{-1/s^{1/3}} \ll s^2 L^2(s),$$

$$\int_{U_0\cap\{y<s\}} f_2(z)dz/2 \sim s^2L^2(s), \quad s\to 0+,$$

which shows that f_2 does not affect convergence. □

Remark 15.13. Now take $d = 3$ and let ρ be *Lebesgue measure* on $C = [0,\infty)^3$. Set $y = z_1 + z_2 + z_3$. Let Z be uniformly distributed on a bounded open set Uß$\{0 < y < 1\}$, whose horizontal sections are open triangles with vertices $Q_0(y)$, $Q_1(y)$, $Q_2(y)$ which depend continuously on $y \in (0,1)$. What condition on the curves Q_i will ensure that $Z \in \mathcal{D}^h(\rho)$? What extra conditions ensure that Z is steady or completely steady? Does it suffice that U is convex? If so, can the triangles U_y revolve? Note that we have not even answered these questions for $d = 2$.

2) We first turn to the general theory. The first question is: Is the picture describing ρ as what one sees of the distribution of Z as one zooms in on a point z_0 on the boundary of the *convex support* of Z correct for all $Z \in \mathcal{D}^h(\rho)$? We know that $Y = \eta(Z)$ lies in $\mathcal{D}^+(\tau)$ with $\tau = -\theta < 0$. Hence Y has finite lower endpoint y_0 and the df F of Y varies regularly in y_0. Assume $y_0 = 0$. Then Z lives on the open halfspace $\{y > 0\}$ and charges all slices $\{0 < y < \varepsilon\}$ with $\varepsilon > 0$. It is possible that the support of Y is the whole closed upper halfspace, even if ρ lives on a proper cone, Example 15.12. Does there exist a point $z_0 = (x_0, 0)$ around which the distributions of the *high risk scenarios* cluster? If so, and if we choose coordinates such that z_0 is the origin, is it possible to normalize the high risk scenarios by linear transformations?

Theorem 15.14. *Let ρ be a Radon measure on $\mathbb{R}^d$ which lives on H_+ and satisfies* (15.24). *Let $Z \in \mathcal{D}^h(\rho)$ have distribution π. There exists a point $z_0 = (x_0, y_0)$ such that*

$$Z^s = Z^{H^s} \to z_0 \ \textit{in probability}\,, \quad s \to 0+,\ H^s = \{y \le s\}.$$

Choose coordinates such that z_0 is the origin. There exist linear transformations α_s mapping $J_0 = \{v \le 1\}$ onto H^s, and depending continuously on s such that

$$\alpha_{s_n}^{-1}\alpha_{c_n s_n}(u,v) \to (cu, cv), \quad s_n \to 0+,\ c_n \to c > 0 \tag{15.25}$$

$$\alpha_s^{-1}(\pi)/\mathbb{P}\{Y \le s\} \to \rho \ \textit{vaguely on } \mathbb{R}^d, \quad s \to 0+\,. \tag{15.26}$$

Proof. We may assume that the df F of Y has lower endpoint $y_0 = 0$. By Theorem 14.7 there exists a continuous function $t \mapsto \tilde{\alpha}_t \in \mathcal{A}(d)$ which varies like the contraction group γ^t in (15.24). A reparametrisation $\alpha_s = \beta_{t(s)}$ gives $\alpha_s^{-1}(Z^s) \Rightarrow W$ with $\tilde{\alpha}_s(v) = sv$. Regular variation of $\tilde{\alpha}_s$ in the sense of (15.25), and of F, implies that the affine transformations $\alpha_s(w) = A_s w + b_s$ satisfy the two limit relations of the theorem. The lemma below shows that one may replace α_s by $w \mapsto A_s w + z_0$. This holds for the sequence $s_n = e^{-n}$ and by interpolation for all s. □

Lemma 15.15. *Let* $\alpha_n \in \mathcal{A}$, $\alpha_n(w) = A_n w + a_n$, *and* $Q \in \mathrm{GL}$. *Assume that*

$$\alpha_n^{-1}\alpha_{n+1}(w) \to Qw, \quad n \to \infty.$$

Suppose the (complex) eigenvalues λ *of* Q *satisfy* $0 < |\lambda| < 1$. *Then* $a_n \to z_0$, *and* $\alpha_n \sim \beta_n$ *where* $\beta_n(w) = A_n w + z_0$.

Proof. Set $\alpha_n^{-1}\alpha_{n+1}(w) = Q_n w + q_n$. Then

$$\begin{aligned}\alpha_m^{-1}\alpha_{m+n}(w) &= q_{m+1} + \cdots + Q_{m+1}\dots Q_{m+n-1}q_{m+n} + Q_{m+1}\dots Q_{m+n}w \\ &= A_m^{-1}(A_{m+n}w + a_{m+n} - a_m).\end{aligned}$$

There exists $c \in (0,1)$ and $m_0 \geq 1$ such that $\|Q^m\| < c_0^m$ for $m \geq m_0$. This implies

$$\|Q_{k+1}\dots Q_{k+m_0}\| < c^{m_0}, \quad k \geq k_0.$$

Since $\|Q_k\| \to \|Q\|$, there is a constant $K \geq 1$ so large that $\|Q_k\| \leq K$ for $k \geq k_0$. Hence

$$\|Q_{k+1}\dots Q_{k+m}\| < Cc^m, \quad k \geq k_0,\ m \geq 1,\ C = (K/c)^{m_0}. \tag{15.27}$$

There is a decreasing null sequence δ_n such that $\|q_n\| \leq \delta_n$ for $n \geq 1$. This gives

$$\|A_m^{-1}(a_{m+n} - a_m)\| \leq C\delta_m/(1-c), \quad m, n \geq 1.$$

The bound is uniform in n. So we may replace a_{m+n} by a limit point z. The limit point is unique since $A_n \to 0$ by (15.27). □

Of particular interest is the situation where z_0 is a *vertex* of the convex support of Z, and Z lives on a proper cone C_0 with top z_0. In that case one may choose coordinates such that Z has non-negative coordinates. Condition (15.25) implies that α_s/s varies slowly for $s \to 0+$. It is satisfied if $\alpha_s/s \to \alpha_0 \in \mathrm{GL}(d)$. In that case a simple scalar normalization will do:

$$Z^s/s \Rightarrow \widehat{W} = \alpha_0^{-1}W, \quad \hat{\rho} = \alpha_0^{-1}(\rho).$$

3) Distribution functions are a convenient tool to investigate the asymptotic behaviour of the distribution of a vector with non-negative components in the neighbourhood of the origin, in particular if the normalizations α_s^{-1} are *scalar expansions*.

Let F be the df of a vector $Z \in [0,\infty)^d$. Suppose $F(0) = 0$ and

$$F(sz)/F(se) \to R(z) \text{ weakly on } (0,1)^d, \quad s \to 0+,\ e = (1,\dots,1) \in (0,\infty)^d.$$

Then R extends to the df of a random vector $\widetilde{W}$ on $[0,1]^d$. Let $F_0(t) = F(te)$. Then $F_0(st)/F_0(s) \to R_0(t) = R(te)$ for $s \to 0+$, $t \in (0,1)$ when te is continuity point of R. Hence convergence holds on a dense set, which implies weak convergence. From the univariate theory of exceedances it follows that $R_0(t) = t^\lambda$ for some $\lambda > 0$, or $R_0 \equiv 0$ or $R_0 \equiv 1$. In the latter cases $R \equiv 0$ or $R \equiv 1$. We exclude these two degenerate cases. Let $Z^{(s)}$ denote Z conditioned to the cube $[0,s]^d$, and let Z^s be Z conditioned to the halfspace $\{y \leq s\}$ where $y = z_1 + \cdots + z_d$. If $Z^{(s)}/s$ converges to $\tilde{W}$ then Z^s/s converges to W where W is $\tilde{W}$ conditioned to $w_1 + \cdots + w_d \leq 1$, and conversely, by a straightforward conditioning argument. By a change of coordinates we obtain the following result.

Theorem 15.16. *Suppose $Z \geq 0$ has df F, and $F(0) = 0$. Let $a, c \in (0,\infty)^d$. Define the vertical coordinate by $\xi z = c^T z$. If $F(sz)/F(sa) \to R(z)$ weakly on $(0,a)$ for $s \to 0+$, and if $0 < R(b) < 1$ for some $b \in (0,a)$, then $Z \in \mathcal{D}^h(\rho)$, where $\rho[0,tz] = t^\lambda R(z)$ for $z \in (0,a)$, $t > 0$, and $Z^{\{\xi \leq s\}}/s \Rightarrow W$ where W has distribution $1_J d\rho/\rho(J)$ with $J = \{\xi \leq 1\}$. Conversely if $Z \geq 0$, and if $Z \in \mathcal{D}^\xi(\rho)$ for an excess measure ρ on $[0,\infty)^d$ with normalizations $\alpha_s(w) = sw$ then*

$$F(sz)/F(sa) \to R(z) = \rho[0,z] \text{ pointwise}, \quad s \to 0+,\ z \in (0,\infty)^d,$$

where ρ satisfies (15.24).

The condition on F in the theorem above does not suffice for F to lie in the domain of attraction of a min-stable distribution. For that one also needs convergence of the marginals F_K,

$$F_K(sz)/c(s) \to R_K(z) \quad s \to 0+,\ z \in \mathbb{R}^K,\ \emptyset \neq K ß\{1,\dots,d\}.$$

Here $c(s) = F_1(s) + \cdots + F_d(s)$. Compare Proposition 7.9.

Example 15.17. Let Z be uniformly distributed on the unit cube. Then $Z \in \mathcal{D}^h(\rho)$ where ρ is *Lebesgue measure* on $(0,\infty)^d$, and $\eta = -(z_1 + \cdots + z_d)$ is the vertical coordinate. The vector Z also lies in the domain $\mathcal{D}^\wedge(W)$ for a *min-stable* vector W with independent exponential marginals. The exponent measure of W lives on the coordinate axes in infinity and does not charge $[0,\infty)^d$. Let $\varepsilon \in (0,1)$. If we delete all mass in the ball εB and condition Z to live on εB^c then the vector still lies in $\mathcal{D}^\wedge(W)$; if we alter the distribution outside εB the vector still lies in $\mathcal{D}^h(\rho)$.

There is a difference in interpretation. In the case of minima we ask for the conditional distribution of the other components if one of the components is small. The answer is that it is not affected (by independence). In the case of exceedances we ask for the conditional distribution of Z given the sum of the coordinates (or the maximum) is small. It is not affected, it remains uniform. ◊

15.4 Cones and heavy tails. We now turn to excess measures ρ on $\{v > 0\}$ whose symmetries are *scalar expansions*:

$$\gamma^t(\rho) = e^t \rho, \quad \gamma^t(w) = e^{\tau t} w, \quad t \in \mathbb{R},\ w = (u,v) \in \mathbb{R}^{h+1},\ \tau = 1/\lambda > 0. \tag{15.28}$$

The measure ρ is infinite on $\{v > 0\}$, and finite on $\{v \geq c\}$, for $c > 0$. The domain of attraction consists of distributions with heavy tails. Heavy tailed distributions on $\mathbb{R}^d$ form the subject of Section 16 and 17. Here we discuss exceedances over horizontal thresholds, and assume ρ lives on $\{v > 0\}$.

Example 15.18. Let N be the Poisson point process on $(0, \infty)^3$ with intensity $1/\eta^{3+\lambda}$, where $\eta = z_1 + z_2 + z_3$. The mean measure ρ satisfies (15.28). Define the curvaceous cone $\widetilde{C}$ by three continuous curves Q_1, Q_2, Q_3 emanating from the origin and diverging to infinity, such that the three points $Q_i(y)$ are the vertices of a triangle $\widetilde{C}_y$ in the plane $\{\eta = y\}$ for each $y > 0$. Assume that the normalized triangles $T_y = \widetilde{C}_y/y$, $y \geq 1$, in the plane $\{\eta = 1\}$ *vary slowly,* i.e.

$$T_{c_n y_n} \sim T_{y_n}, \quad y_n \to \infty,\ c_n \to c > 0.$$

Let $f_0 \colon [0, \infty) \to (0, \infty)$ be continuous, and vary regularly in infinity, with exponent $-(3 + \lambda)$. Then the integral c_0 of $f_0 \circ \eta$ over $\widetilde{C}$ is finite. So $f = 1_{\widetilde{C}} f_0 \circ \eta / c_0$ is a probability density which lives on $\widetilde{C}$. Let $Z_1, Z_2, \ldots$ be independent observations from this density. Let $a_n \to \infty$ such that $\mathbb{P}\{\eta(Z) \geq a_n\} \sim c/n$, where $c = \rho\{\eta \geq 1\}$, and let α_n be the matrix with columns $Q_{ni} = Q_i(a_n)$, $i = 1, 2, 3$. The normalized sample cloud N_n has points $\alpha_n^{-1}(Z_1), \ldots, \alpha_n^{-1}(Z_n)$. Weak convergence of the normalized measures implies weak convergence of the point processes:

$$N_n \Rightarrow N \;\text{ weakly on } \mathbb{R}^3 \setminus \varepsilon B, \quad \varepsilon > 0. \tag{15.29}$$

The curvaceous cone is determined asymptotically by the sections $\widetilde{C}_{y_n}$ with $y_n = e^n$. Conversely any sequence of triangles T_n such that $T_{n+1} \sim T_n$ determine a curvaceous cone $\widetilde{C}$ with sections $\widetilde{C}_{e^n} = e^n T_n$ for $n \geq 0$. One may define the curves Q_i by linear interpolation between successive points $Q_i(e^n)$, where the $Q_i(e^n)$ are the vertices of $e^n T_n$. Such a sequence of triangles T_n is a *discrete skeleton* for the curvaceous cone. The sequence may be quite wild, even if we assume that all triangles lie inside a given bounded convex set D^* in the plane $\{\eta = 1\}$. Indeed, any triangle may be transformed continuously into any other triangle. Hence for any $\varepsilon > 0$ the two triangles may be linked by a finite sequence of triangles $S_0, \ldots, S_m$ such that $|S_k \cup S_{k-1}| < e^{\varepsilon}|S_k \cap S_{k-1}|$. It follows that one may construct a sequence of triangles $T_{n+1} \sim T_n$ which is dense in the space $\mathcal{T}$ of all compact triangles in D^*. In particular, subsequences may converge to disjoint limit triangles. The corresponding sample clouds then asymptotically live on disjoint cones. Yet (15.29) holds, and the convex hulls of the sample clouds converge. $\lozenge$

There is a superficial similarity between the theory of this section and that of the previous section. In both cases the convex support of the excess measure is a closed cone C. If C has a bounded intersection with $\{\eta \leq 1\}$ one may define the compact convex set $C^*ß\mathbb{R}^h$ as in the previous section by

$$C \cap \{\eta = 1\} = C^* \times \{1\}.$$

If Z lives on $[0, \infty)^d$ then, as above we define $\eta = z_1 + \cdots + z_d$. If the high risk scenarios Z^y may be normalized by scalars, then

$$Z^y / y \Rightarrow W,$$

where now we take the limit for $y \to \infty$ instead of $y \to 0+$ as in the preceding section.

Actually the difference with the theory of the preceding section is considerable.

First observe that any Radon measure ρ on $[0, \infty)^d \setminus \{0\}$ which satisfies (15.28) may be interpreted both as an excess measure for exceedances over horizontal thresholds, $\{\eta = y\}$, and as the *exponent measure* of a max-stable distribution on $[0, \infty)^d$ with marginals $H_i(t) = e^{-c_i/t^\lambda}$, $t > 0$. Exponent measures of such max-stable distributions live on $[0, \infty)^d \setminus \{0\}$ and do not charge points in infinity. This yields a one to one correspondence between excess measures on $[0, \infty)^d \setminus \{0\}$ and exponent measures.

In the previous section any *contamination* of the distribution was blown up, and thus liable to destroy steady convergence; in the present set up the contamination is dampened and will disappear in the limit unless it is at least as heavy tailed as the excess measure.

The normalizations α_y, $y > 0$, determine a curvaceous cone $\widetilde{C}$ whose horizontal sections are affine images of C^*. In the previous section the shape of $\widetilde{C}$ determined whether Z was completely steady. From the example above we see that Z is completely steady if it lives on the curvaceous cone $\widetilde{C}$.

For distributions with heavy tails an important difference between exceedances over horizontal thresholds and *coordinatewise extremes* lies in the interpretation. For coordinatewise extremes the d coordinates Z_i have an equal status as measures of risk; for exceedances the vertical coordinate $\eta = z_1 + \cdots + z_d$ (or some other positive linear combination) measures risk; the horizontal part of Z, does not measure risk, but modulates it. Assume Z has non-negative components with heavy tails with the same tail exponent $\lambda > 0$. Exponent measures for such heavy tailed distributions on $[0, \infty)^d$ are precisely the excess measures in (15.28) which live on the cone $[0, \infty)^d$ with $v = \eta = z_1 + \cdots + z_d$. Given the excess measure ρ the asymptotic behaviour of the distribution of Z is determined by d slowly varying functions L_i given by $1 - F_i(z_i) = L_i(z_i)/z_i^\lambda$. For exceedances over horizontal thresholds the global behaviour is determined by a slowly varying function L, which describes the tail $1 - F_0(y) = L(y)/y^\lambda$ of the vertical component $Y = \eta(Z)$, and a sequence of h-dimensional simplices T_n such that $T_{n+1} \sim T_n$, which determines the regularly varying normalization for the horizontal component of the high risk vectors Z^y. If $Z \geq 0$, and the tails of the d components Z_i are asymptotically equal, the relation between coordinatewise extremes and high risk scenarios for exceedances over horizontal thresholds is particularly simple.

Theorem 15.19. *Suppose $Z \geq 0$ has df F with marginals F_i which satisfy*

$$1 - F_i(t) \sim L(t)/t^\lambda, \quad t \to \infty,\ i = 1, \dots, d.$$

Let the Radon measure ρ on $[0,\infty)^d \setminus \{0\}$ satisfy (15.28). Then $Z \in \mathcal{D}^\vee(\rho)$ if and only if $Z \in \mathcal{D}^h(\rho)$ for $\eta = z_1 + \dots + z_d$. Moreover L varies slowly in infinity.

By applying a suitable positive diagonal linear transformation to ρ we may ensure that the marginals of ρ satisfy $\rho_i[t,\infty) = 1/t^\lambda$ for $t > 0$. Then, for $t \to \infty$,

$$Z^{\{\eta \geq t\}}/t \Rightarrow W, \quad (Z^{\vee n})/a_n \Rightarrow \tilde{W} \tag{15.30}$$

$$\varepsilon^{-t}(\pi)/\pi\{\eta \geq e^t\} \to \rho \ \text{ weakly on } \{\eta \geq c\}, \quad c > 0, \tag{15.31}$$

where W has distribution $1_{\{\eta \geq 1\}} d\rho/\rho\{\eta \geq 1\}$, $\tilde{W}$ is max-stable with exponent measure ρ, $a_n > 0$ satisfies $1 - F_i(a_n) \sim 1/n$, $\varepsilon^t(w) = e^t w$, and π is the distribution of Z.

Moreover the normalized sample clouds N_n with points $Z_1/a_n, \dots, Z_n/a_n$ converge in distribution to the Poisson point process N on $O = \mathbb{R}^d \setminus \{0\}$ with mean measure ρ, weakly on $\mathbb{R}^d \setminus \varepsilon B$ for any $\varepsilon > 0$, and the convex hulls converge.

Proof. The diagonal normalizations are scalar expansions: $\alpha_n = a_n$. Set $e = (1, \dots, 1)$. Let F be the df of Z. Convergence

$$(1 - F(tz))/(1 - F(te)) \to R(z) = \rho([0,z]^c)/\rho([0,e]^c), \quad t \to \infty,\ z \in (0,\infty)^d$$

is equivalent to $Z^{\vee n}/a_n \Rightarrow \tilde{W}$ by Theorem 7.3, and to (15.30) by conditioning. The latter implies $N_n \Rightarrow N$ weakly on εB^c. Convergence of the convex hulls will be proved in Theorem 16.18. □

The assumption that the excess measure ρ lives on a cone which intersects the hyperplane $\{v = 1\}$ in a bounded set, and that the distribution π of the vector Z lives on the associated curvaceous cone allows a comparison between the theory of exceedances over horizontal thresholds and the theory of coordinatewise maxima. If we drop these assumptions a number of difficulties arise. We give two examples.

The first example shows that *vague convergence $n\alpha_n(\pi) \to \rho$* on $\{v > 0\}$ need not imply weak convergence on $\{v \geq 1\}$, even when the vertical coordinate lies in the domain of a GPD.

Example 15.20. Let ρ be the excess measure in (15.28). Assume $\rho\{v \geq 1\} = 1$. Let $W = (U, V)$ have distribution $1_{\{v \geq 1\}} d\rho$. Let $Z = (X, Y)$ have distribution π on $\{y \geq 1\}$. Assume $Y^y/y \Rightarrow V$ for $y \to \infty$ and $n\alpha_n^{-1}(\pi) \to \rho$ vaguely on $\{v > 0\}$ for a sequence of linear maps α_n mapping $\{y \geq 1\}$ onto $\{y \geq y_n\}$ with $y_{n+1} \sim y_n \to \infty$. This does not imply $\alpha_n^{-1}(Z^{y_n}) \Rightarrow W$.

We assume that $Z \in \mathbb{R}^2$ has density $(f_0 + f_1)/2$ where $f_0(x, y) = 1_S(x, y)/y^3$ with $S = \{y > 1, 0 < x < y\}$ and $f_1 = 1_T(x, y)\log(1 + y)/y^3$ where $T =$

$\{y > 1, 0 < x < y/\log(1+y)\}$. Then Y has distribution $\tilde{\pi}$ with density $1/y^2$ on $[1,\infty)$, and $Z^y/y \Rightarrow \overline{W}$ where the distribution of W is a mixture of the density f_0 and a univariate GPD on the vertical axis: dv/v^2 on $[1,\infty)$. Set $\beta_t(u,v) = (tu/\log(1+t), tv)$. Then $(2t)\beta_t^{-1}(\pi) \to \rho$ vaguely on $\{v > 0\}$, where ρ has density $1/y^3$ on $\{0 < x < y\}$, and $t\beta(t)^{-1}(\tilde{\pi}) \to \tilde{\rho}$ where $\tilde{\rho}$ is the projection of ρ on the vertical coordinate. $\Diamond$

There is a simple procedure to ensure weak convergence: the two-step conditioning which was used for heavy tailed multivariate GPDs in Section 12.1 will work here too, and will yield sufficient conditions for weak convergence. First condition on the complement of a ball or centered ellipsoid E, and then on a halfspace H supporting E. The limit theory for exceedances over elliptic thresholds is handled by the theory of regular variation developed in MS. This theory is the subject of Section 16 and 17. The final conditioning on the halfspace is trivial if $\rho\{v \geq 1\}$ is positive and $\rho(B^c)$ finite. The global behaviour of a probability distribution in the domain of attraction $\mathcal{D}^\infty(\rho)$ for exceedances over elliptic thresholds depends on a sequence of ellipsoids $2^n E_n$ with $E_{n+1} \sim E_n$, and a sequence of rotations R_n in $\mathrm{SO}(d)$, as is shown in Section 17.2. For exceedances over horizontal thresholds the associated normalizations α_n have to map horizontal halfspaces into horizontal halfspaces. Below we prove that such normalizations α_n exist.

Proposition 15.21. *Let $E_0, E_1, \ldots$ be open centered ellipsoids in $\mathbb{R}^d$ such that $E_{n+1} \sim E_n$. Let $p_n \in \partial E_n$ maximize the vertical coordinate. There exist linear transformations $\alpha_n \in \mathcal{A}^h$ such that $\alpha_n(B) = E_n$, $\alpha_n(e_d) = p_n$, with e_d the vertical unit vector, and $\alpha_n^{-1}\alpha_{n+1} \to I$. If (β_n) is another such sequence then $\beta_n = \alpha_n R_n$ with $R_n \in \mathrm{O}(h)$ and $R_{n+1} \sim R_n$.*

Proof. Let the ellipse F_nß$\mathbb{R}^h$ describe the intersection of E_n with the horizontal coordinate plane $\{v = 0\}$. Then $E_{n+1} \sim E_n$ implies $F_{n+1} \sim F_n$. There exist $A_n \in \mathrm{GL}(h)$ with $A_{n+1} \sim A_n$ such that $F_n = A_n(B^h)$ where B^h is the unit ball in $\mathbb{R}^h$. Extend A_n to a matrix α_n of size d by affixing a bottom row of h zeros, and then affixing the column vector p_n to the right side. Then $\alpha_{n+1} \sim \alpha_n$, $\alpha_n(B) = E_n$, $\alpha_n(e_d) = p_n$, and $\alpha_n \in \mathcal{A}^h$. Finally note that $\beta_n^{-1}\beta_{n+1} = R_n^{-1}\alpha_n^{-1}\alpha_{n+1}R_{n+1} \sim R_n^{-1}R_{n+1}$ since $\mathrm{O}(h)$ is compact. $\square$

Two-step conditioning is not always possible, even when Z lives on $\{y > 0\}$. In the example below the vector $Z = (X, Y) \in \mathbb{R}^2$ has heavy tails and converges under *scalar normalizations*:

$$Z^{\{y \geq t\}}/t \Rightarrow W = (U, V), \quad t \to \infty. \tag{15.32}$$

Yet the tail exponents differ:

$$\mathbb{P}\{V \geq t\} = 1/t^4, \quad \mathbb{P}\{|U| \geq t\} \asymp 1/t^2, \quad t \to \infty. \tag{15.33}$$

Example 15.22. Let the excess measure ρ on $\{v > 0\}$ß$\mathbb{R}^2$ have density $g(u,v) = 1/r^{1+\delta}v^{\mu}$ with $\delta > 0$, $\mu > 0$ and $\delta + \mu > 1$. For $t > 0$

$$\rho\{v \geq t\} = c/t^{\delta+\mu-1}, \quad c = \rho\{v \geq 1\} < \infty,$$

and

$$p_1(t) = \rho\{v \geq 1, u \geq t\} = \int_1^{\infty} J(t/v)dv/v^{\delta+\mu},$$

$$J(q) = \int_q^{\infty} ds/(1+s^2)^{(1+\delta)/2}.$$

Then $J(q) \asymp 1 \wedge 1/q^{1+\delta}$ on $(0,\infty)$ implies $p_1(t) \asymp t_+^{\mu-2}/t^{\mu+\delta-1}$ for $\mu \neq 2$. Let Z have density $g_0(x,y) = 1_{\{y \geq 1\}}g(x,y)/c$. Then $Z^{\{y \geq t\}}/t$ is distributed like Z for $t > 1$. Hence (15.32) holds with $W = Z$. Take $\mu = 4$ and $\delta = 1$ to obtain (15.33). $\diamondsuit$

15.5* Regular variation for matrices in $\mathcal{A}^h$. This section contains criteria, in terms of the matrix representations, for a curve $\beta\colon [0,\infty) \to \mathcal{A}^h$ to vary like γ^t.

The set of affine transformations on $\mathbb{R}^{h+1}$ which map horizontal halfspaces into horizontal halfspaces form a closed subgroup $\mathcal{A}^h$ of $\mathcal{A}$. Elements of $\mathcal{A}^h$ are blocked matrices of size $1 + h + 1$, see equation (2) in the Preview.

$$\alpha = \begin{pmatrix} 1 & 0 & 0 \\ p & A & q \\ b & 0 & c \end{pmatrix}, \quad \alpha^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ P & A^{-1} & Q \\ -b/c & 0 & 1/c \end{pmatrix}, \quad \begin{aligned} P &= A^{-1}(bq/c - p) \\ Q &= -A^{-1}q/c. \end{aligned} \tag{15.34}$$

The two-by-two matrices formed by the four corner entries describe the positive affine transformations $\tilde{\alpha}\colon v \mapsto cv + b$. The map $\alpha \mapsto \tilde{\alpha}$ from $\mathcal{A}^h$ to $\mathcal{A}^+$ is a homomorphism: If $\gamma = \alpha\beta$, then $\tilde{\gamma} = \tilde{\alpha}\tilde{\beta}$. Hence, if $\alpha\colon [0,\infty) \to \mathcal{A}^h$ varies like γ^t, then $\tilde{\alpha}(t)$ varies like $\tilde{\gamma}^t$. Similarly $\alpha \mapsto A$ is a homomorphism, and hence, if $\gamma^t(w) = Q^t w + q(t)$ and $\alpha\colon [0,\infty) \to \mathcal{A}^h$ varies like γ^t, then the central matrices $A(t)$ of size h in (15.34) vary like Q^t.

In multivariate asymptotics centered ellipsoids E_t describe the scale.

Definition. Write $x_t = o(E_t)$ if for any $\varepsilon > 0$ eventually $x_t \in \varepsilon E_t$ for $t \to \infty$.

For matrices with the same bottom row asymptotic equality is simple to express.

Proposition 15.23. *Let $\alpha(t)$ and $\bar{\alpha}(t)$ in $\mathcal{A}^h$ have matrix representations as above. Assume $\bar{a}(t) = a(t)$ and $\bar{b}(t) = b(t)$. Then $\bar{\alpha}(t) \sim \alpha(t)$ for $t \to \infty$ if and only if*

$$\bar{A}(t) \sim A(t), \quad \bar{p}(t) - p(t) = o(E_t), \quad \bar{q}(t) - q(t) = o(E_t), \quad t \to \infty.$$

Let us now show how to alter the bottom row, while retaining asymptotic equality.

Lemma 15.24. *Let* $\alpha\colon [0,\infty) \to \mathcal{A}^h$, *and let* $\delta(t) \sim \tilde{\alpha}(t)$ *for* $t \to \infty$. *Suppose* $\tilde{\alpha}(t)^{-1}\delta(t)(v) = c(t)v + d(t)$. *Define* $\beta(t) = \alpha(t)\eta(t)$ *where* $\eta(t)(u,v) = (u, c(t)v + d(t))$. *Then* $\beta(t) \sim \alpha(t)$ *for* $t \to \infty$, *and* $\tilde{\beta}(t) = \delta(t)$.

Proof. It suffices to observe that $\eta(t)(u,v) \to (u,v)$ implies $\beta(t) \sim \alpha(t)$, and that $\tilde{\beta}(t) = \tilde{\alpha}(t)\tilde{\eta}(t) = \delta(t)$ since $\gamma \to \tilde{\gamma}$ is a homomorphism from $\mathcal{A}^h$ to $\mathcal{A}^+$. □

Example 15.25. Simply replacing the two bottom corner entries in the matrix need not work. Consider the $\alpha_n(c) \in \mathcal{A}^h$ below

$$\alpha_n(c) = \begin{pmatrix} 1 & 0 & 0 \\ 4n^{5/2} + 5n^2 & n & 10n^{3/2} \\ n & 0 & 1/(1 + c/\sqrt{n}) \end{pmatrix}, \quad \tilde{\alpha}_n(c) = \begin{pmatrix} 1 & 0 \\ n & 1/(1 + c/\sqrt{n}) \end{pmatrix}.$$

First observe that $\tilde{\alpha}_n(1) \sim \tilde{\alpha}_n(0)$. It is tempting to replace the lower right hand entry of $\alpha_n(1)$ by 1. A straightforward calculation shows that $\alpha_n(1)^{-1}\alpha_{n+1}(1) \to \gamma\colon (u,v) \mapsto (u, v+1)$. The matrices $\alpha_n(0)$ have a different asymptotic behaviour: $\alpha_n(0)^{-1}\alpha_{n+1}(0) \to \delta\colon (u,v) \mapsto (u + 10, v + 1)$. ◊

We are interested in curves $\alpha\colon [0,\infty) \to \mathcal{A}^h$ which vary like γ^t, where $\gamma^t = e^{tC}$ is a one-parameter group in $\mathcal{A}^h$, such that $\tilde{\gamma}^t$ is the vertical translation group, $v \mapsto v + t$ on $\mathbb{R}$. The corresponding matrices are

$$C = \begin{pmatrix} 0 & 0 & 0 \\ 0 & C^* & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \gamma^t = e^{tC} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & H^t & 0 \\ t & 0 & 1 \end{pmatrix},$$
$$\alpha(t) = \begin{pmatrix} 1 & 0 & 0 \\ p(t) & A(t) & q(t) \\ b(t) & 0 & c(t) \end{pmatrix}.$$

The vectors $p(t)$ and $q(t)$ in the matrix representation of $\alpha(t)$ are non-zero in general. If C^* is in real *Jordan form*, and its diagonal elements are non-zero, then by the Spectral Decomposition Theorem one may choose coordinates and $\bar{\alpha}(t) \sim \alpha(t)$ such that $\bar{p}(t)$ and $\bar{q}(t)$ vanish. If moreover C^* is diagonal, and the diagonal entries are distinct, then, by the same theorem, one may choose these coordinates so that $\bar{A}(t)$ is diagonal. See Section 18.4 for details.

Let us first consider sequences. Suppose the α_n have the form (15.34), and

$$\alpha_n = \alpha_0\gamma_1 \ldots \gamma_n, \quad \gamma_n \to \gamma. \tag{15.35}$$

Introduce the vectors

$$f_n = q_n/c_n, \quad g_n = (p_{n+1} - p_n)/(b_{n+1} - b_n).$$

Let E_n be the ellipse $A_n(B)$ in $\mathbb{R}^h$ where B is the unit ball in $\mathbb{R}^h$. A simple computation gives

Proposition 15.26. *$\gamma_n \to \gamma$ if and only if*

$$A_n^{-1}A_{n+1} \to H, \quad c_{n+1}/c_n \to 1, \quad (b_{n+1} - b_n)/c_n \to 1$$

$$f_n - g_n = o(E_n/c_n), \quad f_{n+1} - f_n = o(E_n/c_n).$$

Proof. This follows from the matrix product below with $\alpha = \alpha_n$ and $\bar{\alpha} = \alpha_{n+1}$.

$$\alpha^{-1}\bar{\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ p & A & q \\ b & 0 & c \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ \bar{p} & \bar{A} & \bar{q} \\ \bar{b} & 0 & \bar{c} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & A^{-1} & 0 \\ 0 & 0 & \bar{c}/c \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \pi - \beta f & \bar{A} & \bar{c}\varphi \\ \beta/\bar{c} & 0 & 1 \end{pmatrix}; \quad \begin{aligned} \pi &= \bar{p} - p, \\ \varphi &= \bar{f} - f, \\ \beta &= \bar{b} - b. \end{aligned}$$ □

We can now construct A_n, b_n, c_n, p_n and q_n so that $\alpha_n^{-1}\alpha_{n+1} \to \gamma$.The right hand columns are determined by the left hand columns: $c_n = b_{n+1} - b_n$, and $q_n = p_{n+1} - p_n$.

Proposition 15.27. *Let $c_n > 0$, $\varphi_n \in \mathbb{R}^h$, and $A_n \in \mathrm{GL}(h)$. Set $E_n = A_n(B)$, $b_{n+1} = b_n + c_n$, $f_{n+1} = f_n + \varphi_n$, $q_n = c_n f_n$, $p_{n+1} = p_n + q_n$. Define α_n in terms of p_n, A_n, q_n, b_n and c_n as above. If $c_{n+1} \sim c_n$, $A_n^{-1}A_{n+1} \to H$, and $\varphi_n = o(E_n/c_n)$, then $\alpha_n^{-1}\alpha_{n+1} \to \gamma$.*

Let us imbed the sequence α_n in (15.35) in a curve $\alpha(t)$. For $t = n+\theta \in [n, n+1)$ define

$$\bar{q}(t) = q_n, \quad \bar{c}(t) = c_n, \quad \bar{b}(t) = b_n + \theta(b_{n+1} - b_n),$$
$$\bar{p}(t) = p_n + \theta(p_{n+1} - p_n), \quad \bar{A}(t) = A_n H^\theta.$$

Proposition 15.28. *Suppose $\alpha_n^{-1}\alpha_{n+1} \to \gamma$. Define $\bar{\alpha}$ on $[0, \infty)$ as above. Then $\bar{\alpha}$ varies like γ^t.*

Proof. It suffices to prove $\alpha_n^{-1}\bar{\alpha}(n + \theta_n) \to \gamma^\theta$ for $\theta_n \in (0, 1)$, $\theta_n \to \theta$. Write down the matrix product above with $\alpha = \alpha_n$ and $\bar{\alpha} = \bar{\alpha}(n + \theta_n)$, and check that it converges to $e^{\theta C}$. Comparison with the product $\alpha_n^{-1}\alpha_{n+1}$ shows that the right hand column converges to $(1, 0, \theta)^T$. □

There are simple criteria for regular variation in terms of derivatives. If the curve $\alpha\colon [0, \infty) \to \mathcal{A}$ is C^1 and $\alpha(t)^{-1}\dot{\alpha}(t){:} = C(t) \to C$ for $t \to \infty$, then $\alpha(t)$ varies like e^{tC}, by continuity of the solution of linear ODEs. The function $\varphi_n(s) = \alpha(t_n)^{-1}\alpha(t_n + s)$ satisfies $\dot{\varphi}_n(s) = \varphi_n(s)C(t_n + s)$. The limit φ exists and solves $\dot{\varphi} = \varphi C$, which gives $\varphi(s) = \gamma^s$.

Proposition 15.29. *Let* $A\colon [0,\infty) \to \mathrm{GL}(h)$ *and* $c\colon [0,\infty) \to (0,\infty)$ *be* C^1 *curves. Suppose* $A(t)^{-1}\dot{A}(t) \to C^*$ *and* $\dot{c}(t)/c(t) \to 0$. *Let* $\varphi\colon [0,\infty) \to \mathbb{R}^h$ *satisfy* $\varphi(t) = o(E_t/c(t))$ *where* $E_t = A(t)(B)$. *Let* f, q, p *and* b *satisfy*

$$\dot{f}(t) = \varphi(t), \quad q(t) = c(t)f(t), \quad \dot{p}(t) = q(t), \quad \dot{b}(t) = c(t).$$

Then $\alpha\colon [0,\infty) \to \mathcal{A}^h$ *varies like* γ^t.

Proof. We have to check that $\alpha(t)^{-1}\dot{\alpha}(t) = C(t) \to C$ for $t \to \infty$. In terms of the matrices this means $A^{-1}(\dot{p} - q) \to 0$, $A^{-1}\dot{A} \to C^*$, and $A^{-1}(\dot{q} - \dot{c}q/c) \to 0$. These relations hold by the conditions in the proposition. □

Let $y(t) = b(t)$ denote the vertical coordinate. A change of variables gives

$$p(t) = \mu(y(t)), \quad c(t) = e(y(t)), \quad E_t = F_{y(t)}.$$

The basic condition

$$c(t)(\dot{p}(t)/c(t))^{\dot{}} = o(E_t), \quad t \to \infty,$$

in the proposition above, in the new variables gives (15.20):

$$e^2(y)\mu''(y) = o(F_y), \quad y \to y_\infty.$$

16 Heavy tails and elliptic thresholds

16.1 Introduction. The random variable Y has *heavy tails* if $\mathbb{E}|Y|^m$ is infinite for some integer $m > 0$. For asymptotics one needs stronger conditions. If Y is non-negative one assumes that the distribution tail varies regularly,

$$1 - F(y) = L(y)/y^\lambda,$$

for some $\lambda > 0$ and some slowly varying function L. In general one assumes

$$\mathbb{P}\{|Y| \geq y\} = L(y)/y^\lambda$$

together with a *balance* condition

$$\mathbb{P}\{Y \geq y\}/\mathbb{P}\{|Y| \geq y\} \to q \in [0,1], \quad y \to \infty.$$

Moments of order less than λ are finite; those of order exceeding λ infinite. In financial mathematics there is evidence that daily log returns have heavy tails with λ in the order of 3 or 4. In *non-life insurance* (fire or storm) λ may be close to one. Regular variation of the upper tail characterizes the domains $\mathcal{D}^+(\tau)$, for $\tau > 0$.

In this section we look at multivariate distributions with heavy tails, and ask under what conditions there is a limit distribution for high risk scenarios when we condition Z to be large. In the past years there has been an increasing awareness that a more geometric approach to the limit theory for coordinatewise maxima for heavy-tailed non-negative vectors simplifies the subject if one uses the right notation. See Resnick [2004]. The restriction to non-negative vectors is not necessary, the reliance on multivariate dfs obscures the tail asymptotics outside the positive quadrant. The use of Poisson point processes is particularly well suited to heavy tails. The relation to multivariate regular variation has been rigorously developed by Meerschaert and Scheffler in their book Meerschaert & Scheffler [2001], which we refer to as MS. Although MS is about limit laws for sums of independent random vectors, the two chapters on regular variation for linear transformations, and for measures on $\mathbb{R}^d \setminus \{0\}$, may actually be regarded as an analysis of *high risk scenarios* for random vectors with heavy tails. For more information on the statistical analysis of multivariate heavy tails see Resnick [2006], or de Haan & Ferreira [2006].

How does one describe the asymptotic behaviour of a vector Z in the plane? One may compactify the plane by adding a point in infinity – thus obtaining the Riemann sphere for the complex plane – and condition Z on a decreasing sequence of neighbourhoods of the point in infinity. One may just as well condition Z to lie outside E_n for an increasing sequence of bounded open convex sets E_n which cover the plane. Our *Ansatz* here is that for large open convex neighbourhoods E of the origin and for certain linear expansions α the distribution of the vector Z conditioned to lie in E^c, and of the vector Z conditioned to lie in the complement of $\alpha(E)$ should have approximately the same shape. The high risk scenarios $Z^{E_n^c}$, properly normalized, should converge in distribution to a random vector W living on the complement of an open bounded convex set E. One could take E_n, and E, to be open triangles. One could also take E to be the open square $(-1, 1)^2$, and E_n to be open squares, or open coordinate rectangles, or open parallelograms. In these cases the set $E_n = \alpha_n(E)$ determines the normalization α_n up to the six or eight symmetries of E. We prefer to take $E = B$, the open unit disk, as our basic set. The sets E_n are open ellipses, and $W_n = \alpha_n^{-1}(Z^{E_n^c})$ are vectors living on B^c. For halfspaces there was a group of affine transformations mapping the upper halfspace H_+ onto itself. For *exceedances over elliptic thresholds* the orthogonal transformations map the unit ball B onto itself.

Assume the sequence of ellipses grows so that $\mathbb{P}\{Z \notin E_n\} \sim e^{-n}$. It may happen that $E_{n+1} \sim qE_n$ for some constant $q > 1$ (successive ellipses have the same shape asymptotically). It may be possible to choose coordinates such that E_n is the unit disk, and E_{n+1} has the form $(x/a_n)^2 + (y/b_n)^2 < 1$ with half axes $a_n \to a$ and $b_n \to b$ with $1 < a < b$. Decay along the major axis is slower than along the minor axis. There is a third option: Shear expansions. See Example 16.6 and Section 18.8. There are two points of view. One school argues that decay with the same exponent along both axes is such a coincidence that it may be ignored in practice; the other school claims that if the two exponents differ then asymptotically the tails in the direction of fast decay become negligible, and may be ignored. Moreover, if symmetry does occur, why should it not be simple? We shall develop the theory with special attention to scalar and diagonal non-scalar expansions. The domains of attraction in these two situations differ like sun and moon.

There is a situation of great practical interest where the d marginal tails agree. That is the case where $Z = (Z_{n+1}, \dots, Z_{n+d})$ is a finite section from a stationary sequence. Here the analytic theory of coordinatewise maxima is the most obvious candidate to investigate the tail behaviour, but the geometric extreme value theory developed in the present text may give useful insight in the structure of the tail of the vector Z. Stationarity of the sequence (Z_n) does not imply the existence of an excess measure. It does impose certain restrictions on the measure and its symmetry group if an excess measure exists.

Example 16.1. Even if the marginals $Z_1, \dots, Z_d$ have the same heavy tailed distribution, there may be linear combinations with lighter tails. Let X_n, $n \in \mathbb{Z}$, be iid with df F, and let Y_0 be independent of (X_n) with df G. Assume that these distributions are symmetric. The variables $Z_n = X_n + Y_0$, $n \in \mathbb{Z}$, form a stationary sequence. Suppose $1 - F(x) \sim 1/x^2$ and $1 - G(y) \sim 1/y$. Let $U_0 = \theta_1 Z_1 + \cdots + \theta_d Z_d$ where θ is a unit vector. If $\theta_1 + \cdots + \theta_d = 0$ then U_0 has tails with exponent two, else tails with exponent one. Despite stationarity the marginals fail to describe the tail behaviour of the vector $(Z_1, \dots, Z_d)$. $\Diamond$

A large sample cloud from the standard normal distribution π on the plane, properly scaled, will form a black disk. The measure $n\pi$ contracted by $\sqrt{2 \log n}$ converges to a measure ρ which is infinite for every non-empty open set in the unit disk and vanishes outside the closed disk. Such behaviour is typical for light tailed distributions as we saw in Section 9.5. For such distributions one has to zoom in at a boundary point to obtain an interesting local picture. For heavy tails the global and local description of the sample cloud agree. Sample clouds from distributions with heavy tails have no edge. The sample cloud does not consist of a dense body of points surrounded by a halo. There is only a halo of isolated points emanating from a center where points accumulate. The variation in the convex hull is of the same order of magnitude as the convex hull itself. For such sample clouds there is only a global theory. That theory is the subject of the present section, and the next.

Example 16.2. Consider a continuous symmetric distribution on $\mathbb{R}$ with tail $R(t) = 1 - F(t) \sim 1/t^5$. Such a tail is not very heavy. Moments of order < 5 are finite. Take ten very large samples of n points, and let T_k denote the maximum of the kth sample. The ten points T_k will flock around the value $b_n \sim n^{1/5}$ where the tail R assumes the value $1/n$. By the univariate theory each of the normalized points T_k/b_n

will behave like the maximal point V_1 from a Poisson point process on $(0, \infty)$ with tail $1/v^5$. Observe

$$\mathbb{P}\{3/4 \le V_1 \le 2\} = e^{-(3/4)^5} - e^{-2^5} \approx 0.8.$$

For large n the window $[(3b_n/4), 2b_n]$ will fail to contain the ten points T_k with probability > 0.9. The window $[-5b_n, 5b_n]$ is eight times larger. It will contain the n-point cloud for all ten samples with probability > 0.99 since $20e^{-5^5} < 0.01$. $\diamondsuit$

The example suggests that for samples $Z_1, \dots, Z_n$ from heavy tailed distributions π on $\mathbb{R}^d$ one should consider the *global behaviour* of $n\pi_n$ with $\pi_n = \alpha_n^{-1}(\pi)$ for suitable linear or affine expansions α_n.

We shall adopt the following assumptions:

$$e^t \alpha(t)^{-1}(\pi) \to \rho \ \text{ weakly on } \varepsilon B^c, \quad t \to \infty, \ \varepsilon > 0, \tag{16.1}$$

where ρ is a Radon measure on $O = \mathbb{R}^d \setminus \{0\}$, finite and positive on the complement of the unit ball B. Weak convergence in (16.1) ensures that no points of the sample clouds escape to infinity.

Definition. A one-parameter group γ^t, $t \in \mathbb{R}$, is called a *linear expansion group* if the transformations γ^t are linear, and if all eigenvalues of γ lie outside the unit circle in $\mathbb{C}$.

In addition to (16.1) we assume that $\alpha\colon [0, \infty) \to \mathcal{A}$ varies like γ^t,

$$\alpha(t_n)^{-1}\alpha(t_n + s_n) \to \gamma^s, \quad t_n \to \infty, \ s_n \to s, \ s \in \mathbb{R}, \tag{16.2}$$

where γ^t, $t \in \mathbb{R}$, is a linear expansion group.

From the general theory of regular variation as developed in MS it follows that ρ is an excess measure which satisfies

$$\gamma^t(\rho) = e^t \rho, \quad t \in \mathbb{R}. \tag{16.3}$$

The proof that (16.1) and (16.2) imply (16.3) is straightforward, see Proposition 18.23. The conditions imply that the high risk scenarios $Z^{E_t^c}$ for the ellipsoids $E_t = \alpha(t)(B)$, normalized by $\alpha(t)$, converge in distribution to the vector W on B^c with distribution $1_{B^c}\, d\rho/\rho(B^c)$. As was pointed out above, it is not the elliptic thresholds, ∂E_t, but rather the expansions α_t which matter.

Definition. Let Z be a random vector in $\mathbb{R}^d$ with distribution π, and let ρ be a Radon measure on $O = \mathbb{R}^d \setminus \{0\}$ which does not live on a linear hyperplane. We say that π and Z lie in the *domain of elliptic attraction* of ρ and write $\pi \in \mathcal{D}^\infty(\rho)$ if (16.1) and (16.2) hold.

An intuitive introduction to the conditions (16.1) and (16.2) was given in the Preview. The conditions may be formulated in terms of regular variation of measures, as developed in MS. Here is their definition, adapted to our notation:

Definition. A finite measure μ on $\mathbb{R}^d$ *varies regularly* if there is a continuous function $\beta : [0, \infty) \to \mathrm{GL}(d)$ such that (16.2) holds and

$$e^t \beta(t)^{-1}(\mu) \to \rho \text{ vaguely on } \mathbb{R}^d \setminus \{0\}$$

for a Radon measure ρ on $\mathbb{R}^d \setminus \{0\}$ which does not live on a hyperplane through the origin, and which satisfies $\rho(B^c) < \infty$. The measure μ *varies regularly at infinity* if in addition any bounded set is covered by the ellipsoids $E_t = \beta(t)(B)$ eventually, where B denotes the open unit ball.

We shall see later, in Section 16.4, that vague convergence on $\mathbb{R}^d \setminus \{0\}$ implies weak convergence on εB^c for $\varepsilon > 0$. So the domains of attraction $\mathcal{D}^\infty(\rho)$ for *exceedances over elliptic thresholds* consist precisely of those probability measures which vary regularly at infinity. In the literature on the asymptotics of heavy tails the term multivariate regular variation is often used in a more restricted sense. Only scalar normalizations are allowed. This special case will be treated in Section 17.1.

The theory presented in this section, and the examples in the next should give the reader an indication of the scope of the theory of regular variation of linear operators and measures presented in MS. Our exposition is self-contained. Our notation and terminology are in line with the previous sections, and differ somewhat from the usage in MS. In particular we shall write γ^t, $t \in \mathbb{R}$, with generator C rather than r^C, $r > 0$. Moreover we normalize probability distributions by α_n^{-1} where MS use A_n. As a consequence we write $\alpha_n \sim \beta_n$ if $\alpha_n^{-1}\beta_n \to \mathrm{id}$, whereas in MS asymptotic equality means $A_n B_n^{-1} \to I$. For very heavy tails sample sums behave like extremes. We shall briefly touch on the relation to operator stable processes and distributions and the interpretation of excess measures as Lévy measures in Section 17.6.

We start with a description of excess measures for heavy tails. We then proceed to investigate the domains of attraction $\mathcal{D}^\infty(\rho)$ of such excess measures. We show that one may choose coordinates and a continuous normalization curve α in (16.1) such that the ellipsoids $E_t = \alpha(t)(B)$ satisfy the inclusion $\mathrm{cl}(E_s)$ßE_t for $0 \le s < t$. This allows us to introduce *polar coordinates* in which the boundaries of the ellipsoids E_t function as spheres, and t as the radial coordinate. The limit relation (16.1) then describes the asymptotic independence of the radial and angular part of the vector Z in these coordinates. In Section 16.4 we prove convergence of the convex hulls, and of the empirical loss for functions whose growth is bounded by a suitable power function. Section 16.9 proves the basic step in the Spectral Decomposition Theorem.

Section 17 treats a number of cases in greater detail: scalar normalizations, scalar expansion groups, and normalization by coordinate boxes. We shall briefly discuss tails with different rates of decay in different directions. We shall describe the excess measures of maximal symmetry associated with such tail behaviour. Finally we return to elliptic thresholds: we determine the limit laws for high risk scenarios $Z^{E_n^c}$ where E_1ßE_2ß$\cdots$ is an increasing sequence of ellipsoids. The limit behaviour here is not as simple as for exceedances over horizontal thresholds. Not every non-degenerate limit law extends to an excess measure which varies regularly in infinity!

The main result of these two sections is a complete characterization of the domain of attraction, $\mathcal{D}^\infty(\rho)$, for excess measures ρ with a continuous positive density, whose generator C has complex diagonal *Jordan form*. In the exemplary case where ρ is the *Euclidean Pareto* excess measure, with density $1/\|w\|^{d+1/\tau}$ (and $C = \tau I$), any distribution function $\pi \in \mathcal{D}^\infty(\rho)$ has the form $d\pi = f d\mu$, where f is a continuous unimodal function with elliptic *level sets*: $\{f > e^{-t}\} = E_t = \alpha(t)(B)$, and μ is a roughening of Lebesgue measure. Such measures may be discrete, or have a wild density. They are characterized in terms of a partition of $\mathbb{R}^d$ into bounded sets A_n such that $\mu(A_n) \sim |A_n|$, where $|A|$ denotes the volume of A. The global tail behaviour of the probability distribution π is determined by the *typical density* f, and hence by a sequence of ellipsoids E_n which satisfy $E_{n+1} \sim e^\tau E_n$. In general the linear expansions γ^t are not scalar, and the typical density f is not constant on ∂E_t. The structure of the typical density f in the expression $d\pi = f d\mu$ depends on the generator C. This structure is determined in Section 17 for all generators C whose complex Jordan form is diagonal.

The final section of this chapter is of a more theoretical nature. It contains an introduction to multivariate regular variation, a discussion of the Meerschaert Spectral Decomposition Theorem, and a classification of all excess measures on $\mathbb{R}^d$.

16.2 The excess measure. Our first task is to find the excess measures ρ and their symmetry groups $\gamma^t, t \in \mathbb{R}$. What is an expansion? What sets are expanded? Are the ellipsoids $\gamma^t(B)$ ordered by inclusion? How is the excess measure distributed over the different orbits of the expansion group? We shall also look at densities and polar coordinates.

For scalar expansion groups, $\gamma^t(w) = e^{\tau t} w$, $t \in \mathbb{R}$, with τ a positive constant, *polar coordinates* form a convenient tool. Any continuous non-negative function g_0 on the unit sphere ∂B, determines the density g of an excess measure ρ on $\mathbb{R}^d \setminus \{0\}$ which satisfies $\gamma^t(\rho) = e^t \rho$, for the scalar expansions above. In polar coordinates

$$g(r\theta) = g_0(\theta)/q^t, \quad q = e^{1+d\tau}, \ \theta \in \partial B, \ r > 0.$$

Note that in affine geometry there are no balls, only ellipsoids. One can not even speak of centered ellipsoids.

There are several reasonable interpretations for an *expansion*. They turn out to be equivalent. Let us write $\gamma^t(w) = q_t + Q^t w$, and let C^* be the linear part of the generator C of γ^t. So $Q^t = e^{tC^*}$. The (complex) eigenvalues ζ_i of C^* lie in $\Re > 0$ if and only if the eigenvalues $e^{\zeta_i t}$ of Q^t lie outside the closed unit disk for

$t > 0$. If one of these conditions holds then we call γ an expansion. Since zero is no eigenvalue of C^* there is a unique point z_0 such that $Cz_0 = 0$. The point z_0 is fix point of the group γ^t. So γ^t, $t \in \mathbb{R}$, becomes a linear group Q^t, $t \in \mathbb{R}$, if we choose the origin in z_0. Henceforth γ^t will be linear transformations on $\mathbb{R}^d$, and the generator C is a matrix of size d.

Definition. A linear *expansion* is a linear transformation Q whose eigenvalues lie outside the unit circle in $\mathbb{C}$. An *expansion* with center z_0 is an affine transformation $\gamma\colon z \mapsto Qz + q$ where Q is a linear expansion and $\gamma(z_0) = z_0$. An *expansion group* is a one-parameter group of linear transformations Q^t, $t \in \mathbb{R}$, with Q a linear expansion.

Exercise 16.3. If γ^t, $t \in \mathbb{R}$, is an expansion group then $\|\gamma^{-t}\| \to 0$ for $t \to \infty$. For any $\varepsilon > 0$ and any $r > 1$ there exists a constant t_0 such that $\gamma^t(\varepsilon B) \supset rB$ for $t \geq t_0$. ♢

Let us show that the algebraic concept of an expansion group agrees with our geometric intuition.

Proposition 16.4. *Let E be an open centered ellipsoid and γ^t, $t \in \mathbb{R}$, a linear group. If $\gamma(E)$ contains the closure of E then γ^t, $t \in \mathbb{R}$, is an expansion group.*

Proof. The condition implies that the eigenvalues λ_i of Q satisfy $|\lambda_i| > 1$. Restrict E to a subspace spanned by a real or complex eigenvector! □

Proposition 16.5. *Let C be a matrix of size d. Write $C_0 = (C^T + C)/2$ for the symmetric part of C. If C_0 is positive definite then C generates an expansion group. Set $Q(w) = w^T C w$. The minimum value δ of Q over ∂B is positive, and*

$$e^{\delta t} B \text{ß} \gamma^t(B), \quad t \geq 0.$$

Proof. Observe that $Q(w) = w^T C_0 w$. The antisymmetric part of C does not contribute. Since ∂B is compact the minimum is attained and positive. Let $x_0 \in \partial B$, and set $x(t) = \gamma^t(x_0)$, and $\varphi(t) = x(t)^T x(t)$. Then $\dot{x}(t) = Cx(t)$, and $\dot{\varphi}(t) = 2x(t)^T Cx(t)$. By assumption $x^T Cx \geq \delta x^T x$. Hence $\dot{\varphi}(t) \geq 2\delta\varphi(t)$. Together with $\varphi(0) = 0$ it follows that $\log \varphi(t) \geq 2\delta t$. Hence $\|\gamma^t x\| \geq e^{\delta t} x$ for all $x \in \partial B$, $t \geq 0$. □

Definition. A bounded convex open set E which contains the origin is *adapted* to the generator C, or the linear expansion group $\gamma^t = e^{tC}$, if there exists $\delta > 0$ such that

$$e^{\delta t}(E) \text{ß} \gamma^t(E), \quad t \geq 0.$$

Example 16.6. The expansion groups in dimension $d = 2$ have generators

$$C = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}, \quad \begin{pmatrix} \lambda & \tau \\ -\tau & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$$

in *Jordan form*, with $\lambda, \mu > 0$ and $\tau \neq 0$, and

$$\gamma^t = \begin{pmatrix} e^{\lambda t} & 0 \\ 0 & e^{\mu t} \end{pmatrix}, \quad e^{\lambda t} \begin{pmatrix} \cos \tau t & \sin \tau t \\ -\sin \tau t & \cos \tau t \end{pmatrix}, \quad e^{\lambda t} \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}.$$

For the diagonal generator C with $\lambda \neq \mu$ the unit disk B is adapted, and so is any coordinate ellipse and any coordinate rectangle. Orbits are curves, $y = cx^\theta$, and may leave an ellipse to return and intersect it if the ellipse is elongated along the diagonal. Such ellipses are not adapted. The second group consists of expansive rotations. Ellipses (and squares) need not be adapted. The last group describes expansive *shears* along the vertical axis. The unit disk is adapted if and only if $\lambda > 1/2$. For $0 < \lambda \leq 1/2$ one may rescale the coordinates by $D = \operatorname{diag}(1, c)$

$$D(\lambda I + J)D^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/c \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ c & \lambda \end{pmatrix}. \tag{16.4}$$

For $c \in (0, 2\lambda)$ the unit disk is adapted in the new coordinates. ♢

If the centered open ellipsoid E_0 is adapted, one may choose coordinates so that $E_0 = B$ is the unit ball. Now perform an extra coordinate transformation S. If S is close to I the unit ball in the new coordinates will still be adapted since the generator in the new coordinates has the form $\widetilde{C} = S^{-1}CS$, and the minimum of $x^T \widetilde{C} x$ over the unit sphere will be positive if S is close to I. This proves one part of the next result:

Theorem 16.7. *For any linear expansion group the adapted ellipsoids form an open non-empty subset of the set of open centered ellipsoids.*

To prove that the set is non-empty, observe that the (complex) eigenvalues λ of γ^m satisfy $|\lambda| \geq e$ if m is large. If we choose coordinates such that the generator mC of γ^{mt}, $t \in \mathbb{R}$, is in Jordan form, then the unit ball is adapted for γ^{mt}, and hence also for γ^t. This follows from the lemma below.

Lemma 16.8. *Let C be in real Jordan form. If the real parts of the complex eigenvalues $c = a + ib$ satisfy $a \geq 1$, the quadratic form $x^T Cx$ is positive definite.*

Proof. First consider a real matrix $A = cI + J$ of size m with $c \geq 1$, where J has zeros except for ones just below the diagonal. Let Q be the sum of $m - 1$ terms: $Q(x) = x_1x_2 + x_2x_3 + \cdots + x_{m-1}x_m$. The inequality $2|ab| \leq a^2 + b^2$ gives

$$x^T Ax = cx^T x + Q(x) \geq x_i^2/2 > 0, \quad x \neq 0,$$

where x_i is the first non-zero entry in the vector x. Similarly if A is the real matrix of size $2m$ corresponding to the complex matrix $cI + J$ of size m, with $c = a + ib$ and $a \geq 1$, then for $w = (u_1, v_1, u_2, \dots, v_m) \neq 0$ we find $w^T A w = a u^T u + a v^T v + Q(u) + Q(v) > 0$. $\square$

If the unit ball B is adapted to the expansion group γ^t one may introduce *exponential polar coordinates* by the homeomorphism

$$\Phi\colon (p,t) \mapsto (\gamma^t p), \quad \Phi\colon \Gamma = \partial B \times \mathbb{R} \to O = \mathbb{R}^d \setminus \{0\}. \tag{16.5}$$

The group γ^t, $t \in \mathbb{R}$, on O corresponds to a vertical translation group on a cylinder:

$$\tau^t(p,s) \mapsto (p, s+t), \quad (p,s) \in \partial B \times \mathbb{R} = \Gamma. \tag{16.6}$$

Definition. An *excess measure for expansions* is a non-zero Radon measure ρ on $O = \mathbb{R}^d \setminus \{0\}$ which satisfies $\gamma^t(\rho) = e^t \rho$ for some expansion group γ^t, $t \in \mathbb{R}$.

Note that an excess measure for expansions is finite on B^c, and hence on εB^c for any $\varepsilon > 0$. Indeed, let the ellipsoid E be adapted, and set $R_0 = \gamma(E) \setminus E$ and $R_n = \gamma^n(R_0)$. Then $\rho(R_0) = c$ is finite, since ρ is a Radon measure on O, and $\rho(E^c) = c + c/e + c/e^2 + \cdots$ is finite.

The excess measure ρ on $O = \mathbb{R}^d \setminus \{0\}$ corresponds to a product measure $d\rho^* \times e^{-t} dt$ on the cylinder Γ. The measure ρ^* on ∂B is called the *spectral measure* of ρ.

The spectral measure, together with the generator C, determines the excess measure ρ. The generator determines the orbits; the spectral measure says how the excess measure is distributed over the orbits. It should be stressed that the spectral measure is not a geometric object. It depends on the coordinates. One may define a generalized spectral measure to depend on a bounded open neighbourhood U of the origin and a bounded open convex set D which is adapted:

Let $\mathcal{B}_i$ be the σ-field of invariant Borel sets $A ß O = \mathbb{R}^d \setminus \{0\}$. These sets A satisfy:

$$\gamma^t(A) = A, \quad t \in \mathbb{R}.$$

They are unions of orbits. Since D is adapted, each orbit of the group γ^t, $t \in \mathbb{R}$, intersects the boundary ∂D in precisely one point. This establishes a natural isomorphism between the σ-field $\mathcal{B}_i$ on O and the Borel σ-field on ∂D. The spectral measure $\rho^*_{U,\partial E}$ for U on ∂D may be identified with a finite measure $\bar{\rho}_U$ on $\mathcal{B}_i$ by setting

$$\rho^*_{U,\partial D}(A) = \rho(A \setminus U) = \bar{\rho}_U(A), \quad A \in \mathcal{B}_i. \tag{16.7}$$

If ρ and ρ_0 are excess measures with the same generator, and $d\rho = h d\rho_0$ then h is measurable on $\mathcal{B}_i$, $h \circ \gamma^t = h$ for all t, and $d\rho^* = h d\rho_0^*$ if $U = D$. For the sake of simplicity we assume henceforth that the unit ball is adapted to the expansion group γ^t, $t \in \mathbb{R}$, of the excess measure ρ, and that $\rho^* = \rho^*_{B,\partial B}$.

Let g_0 be a continuous non-negative function on ∂B. Define g on $\mathbb{R}^d \setminus \{0\}$ by

$$g(\gamma^t(w)) = g_0(w)/q^t, \quad w \in \partial B,\ t \in \mathbb{R};\ q = e \det \gamma.$$

Then g is continuous. It is the density of a Radon measure ρ on $\mathbb{R}^d \setminus \{0\}$, which satisfies $\gamma^t(\rho) = e^t \rho$ for $t \in \mathbb{R}$. The ring $R_0 = \gamma(B) \setminus B$ has finite measure. The rings $R_k = \gamma^k(R_0)$ fill up $\mathbb{R}^d \setminus \{0\}$ and $\rho(R_k) = e^{-k}\rho(R_0)$ implies that $\rho(B^c) < \infty$, and similarly $\rho(\varepsilon B^c) < \infty$ for any $\varepsilon > 0$, since $\gamma^{-n}(B)$ßεB eventually.

For an expansion group $\gamma^t = e^{tC}$ for which the unit ball is adapted we have two methods for constructing excess measures with a continuous density, starting from a continuous non-negative function on ∂B. We may use the procedure above to define a continuous function g on $\mathbb{R}^d \setminus \{0\}$, or we may regard the function on ∂B to be the density of the spectral measure ρ^* on ∂B. These two procedures in general yield different excess measures!

Proposition 16.9. *Let ρ be an excess measure on $\mathbb{R}^d \setminus \{0\}$, with generator C and with density g. Assume B is adapted. Let π_0 be the uniform probability distribution on ∂B. Then the spectral measure ρ^* satisfies*

$$\begin{aligned} &d\rho^*(w) = b(d)\chi(w)g(w)d\pi_0(w),\\ &w \in \partial B,\ \chi(w) = w^T C w,\ b(d) = 2\pi^{d/2}/\Gamma(d/2). \end{aligned} \tag{16.8}$$

Proof. Let $\varphi \geq 0$ be a continuous function with compact support in $\mathbb{R}^d \setminus \{0\}$ which does not vanish on ∂B. We may assume that g is continuous, else write $d\rho = h d\rho_0$ as above, where ρ_0 has a continuous positive density. Let t tend to zero from above. The width of the ring $\gamma^t(B) \setminus B$ at the point $w \in \partial B$ is asymptotic to $w^T\gamma^t(w) - 1$ and hence to $t\chi(w)$. We find

$$\int_{\gamma^t(B)\setminus B} \varphi d\rho \sim \begin{cases} (1-e^{-t}) \int_{\partial B} \varphi d\rho^*, \\ t \int_{\partial B} \varphi(w)\chi(w)d\sigma(w), \end{cases} \qquad t \to 0.$$

Here $d\sigma = b(d)d\pi_0$ it surface area on ∂B. □

Excess measures for expansions have large supports.

Proposition 16.10. *Let ρ be a Radon measure on $\mathbb{R}^d \setminus \{0\}$, and γ a linear expansion such that $\gamma(\rho) = \rho/q$ for some $q \in (0,1)$. If ρ lives on a strip $T = \{|\xi| \leq c\}$ for a non-zero functional ξ, then ρ lives on $\{\xi = 0\}$.*

Proof. If z lies in the support S of ρ, then $\gamma^k(z) \in S$ for all $k \in \mathbb{Z}$, and $\gamma^{-n}(z) \to 0$ implies $0 \in S$. Hence if ρ lives on a hyperplane it lives on a proper linear subspace.

There is a minimal linear subspace containing S. Assume ξ does not vanish on this subspace. We shall derive a contradiction. We may assume that $\mathbb{R}^d$ is minimal. Hence S is not contained in a hyperplane. There exist points $z_0, \dots, z_d \in S$ such that the convex hull of these points contains a ball $p + \varepsilon B$. Choose $r > 0$ so large that the strip T does not contain the ball rB. Now choose $n \geq 1$ so large that $\gamma^n(\varepsilon B) \supset rB$. Then one of the points $\gamma^n(z_0), \dots, \gamma^n(z_d)$ lies outside T. Contradiction. □

Theorem 16.11. *The group of measure preserving linear transformations of a full excess measure for linear expansions is compact.*

Proof. Let $\mathcal{S}$ denote the group and ρ the excess measure with linear symmetries $\gamma^t(\rho) = e^t\rho$. Assume $\rho(B^c) = 1$. Since ρ is full, there exists a constant $c > 0$ such that $\rho\{|\zeta| > 1\} > c$ for all unit functionals ζ. Let $E_t = \gamma^{-t}(B)$. Then $\rho\{|\xi| > 1\} > e^t c$ if $\{\xi \geq 1\}$ supports E_t. Choose t so large that $e^t c > 1$, and let rBßE_t. Then $\rho\{|\zeta| > r\} > 1$ for any unit functional ζ. Hence rBß$\sigma(B)$ for all $\sigma \in \mathcal{S}$. Equivalently $\|\tau\| \leq 1/r$ for $\tau \in \mathcal{S}$. □

16.3 Domains of elliptic attraction. The theories for exceedances over elliptic and horizontal thresholds are similar. For elliptic thresholds there is a simplification. The unit sphere is compact, and so is the space $\mathcal{P}(\partial B)$ of all spectral probability measures.

There are two tasks: Find the excess measures, and determine their domains of attraction. The first is simple. Linear expansion groups are determined by their generators. If we do not bother about coordinates the generator may be given in Jordan form. The only condition is that the diagonal elements in the real Jordan form be positive. This ensures that the eigenvalues of $\gamma = e^C$ lies outside the unit circle in $\mathbb{C}$. Given the *linear expansion group* γ^t, choose coordinates so that the unit ball B is adapted, and choose a finite non-zero measure ρ^* on ∂B. This defines a unique excess measure ρ which satisfies $\gamma^t(\rho) = e^t\rho$, and $\rho(B^c) < \infty$. Every excess measure ρ for linear expansions which is finite on the complement of some bounded ball rB is of this form. Details are given in the previous section.

Here we look at the domains of attraction. Recall that the domain of attraction $\mathcal{D}^\infty(\rho)$ of an excess measure ρ which does not live on a proper linear subspace is the set of all probability measures π which satisfy (16.1). We shall first prove that the normalizations $\alpha(t)$ may be assumed linear. Moreover one may choose the curve $\alpha\colon [0, \infty) \to \mathrm{GL}$ to be continuous, and so that the ellipsoids $E_t = \alpha(t)(B)$ satisfy

$$\mathrm{cl}(E_s)\text{ß}E_t, \quad 0 \leq s < t. \tag{16.9}$$

This establishes a homeomorphism between the half-cylinder $\partial B \times [0, \infty)$ and $\mathbb{R}^d \setminus E_0$ mapping (w, t) into $\alpha(t)w$. In these polar coordinates it is simple to characterize $\mathcal{D}^\infty(\rho)$.

In the previous section it was shown that one may assume the symmetry group γ^t of the limit measure ρ to be linear; here we show that one may choose linear normalizations.

We return to the condition (16.3). It implies $\mathbb{P}\{\alpha(t)^{-1}(Z) \in \varepsilon B^c\} \to 0$, and hence

$$\alpha(t)^{-1}(Z) \to 0 \text{ in probability}, \quad t \to \infty. \tag{16.10}$$

Let the eigenvalues of the matrix γ all lie inside the circle of radius r_1 in $\mathbb{C}$ (and hence inside a circle with radius $r_0 < r_1$). The bound $\|\gamma_{n+1} \dots \gamma_{n+m}\|/r_0^m < M$, $n, m \geq 0$, below, holds for any sequence $\gamma_n \to \gamma$. It implies that partial products of γ_n/r_1 vanish at an exponential rate.

Lemma 16.12. *Suppose $\beta_n \to \beta$ in $\mathrm{GL}(d)$ where β is a contraction (all eigenvalues lie inside the unit circle in $\mathbb{C}$). There exists $M > 1$ such that*

$$\|\beta_{n+1} \dots \beta_{n+m}\| < M, \quad n, m \geq 0. \tag{16.11}$$

Proof. Write β in Jordan form to see that $\|\beta^n\| \to 0$. Hence $\|\beta^k\| < 1/3$ for some $k \geq 1$, and, by continuity, $\|\beta_{n+1} \dots \beta_{n+k}\| < 1/2$ for $n \geq n_0$. Since $\|\beta_n\|$ is bounded by some constant $C_0 \geq 1$ we have

$$\|\beta_{n+1} \dots \beta_{n+m}\| < C_0^{n_0} C_0^k / 2^j, \quad jk \leq m < jk + k, \; n \geq 0.$$

This gives (16.11). □

Lemma 16.13. *Let $\alpha_n(w) = A_n w + a_n$. Suppose $\alpha_n^{-1}\alpha_{n+1}(w) \to Qw$ where Q is a linear expansion. Then $A_n^{-1}\alpha_n \to \mathrm{id}$.*

Proof. Write $\alpha_{n-1}^{-1}\alpha_n(w) = Q_n w + q_n$, with $\alpha_0 = \mathrm{id}$. Then $q_n \to 0$ and

$$\alpha_n(w) = q_1 + Q_1 q_2 + \cdots + Q_1 \dots Q_{n-1} q_n + Q_1 \dots Q_n w = a_n + A_n w.$$

Hence $A_n^{-1} a_n = b_1 + \cdots + b_n$ with $b_k = (Q_k \dots Q_n)^{-1} q_k$. Now observe that $\|Q^m\| < 1/4$ for some $m \geq 1$ implies $\|(Q_{n+1} \dots Q_{n+m})^{-1}\| < 1/2$ for $n \geq n_0$ and hence $\|(Q_{n+1} \dots Q_{n+k})^{-1}\| < M e^{-\varepsilon k}$ for $k, n \geq 1$. Since $q_n \to 0$ we have an exponential bound for $b_k = (Q_k \dots Q_n)^{-1} q_k$, independent of n, given by $\|b_k\| \leq M N e^{-\varepsilon k}$. Since $(Q_{n+1-i} \dots Q_n)^{-1} q_{n-i} \to 0$ for $i = 0, 1, 2, \dots$ we find $A_n^{-1} a_n \to 0$. □

The next result makes clear why we are so keen on unit balls that are adapted.

Theorem 16.14. *Let $\alpha \colon [0, \infty) \to \mathcal{A}$ vary like γ^t for the linear expansion group $\gamma^t = e^{tC}$. Suppose $w^T C w > \delta_0 > 0$ for $w \in \partial B$. Then there exists a continuous curve $Q \colon [0, \infty) \to \mathrm{GL}$ such that $Q(t) \sim \alpha(t)$ for $t \to \infty$, and such that the ellipsoids $E_t = Q(t)(B)$ satisfy $e^{\delta_0 s} E_r \Subset E_{r+s}$ for $r, s \geq 0$. In particular (16.9) holds.*

Proof. Set $\alpha(t)w = A(t)w + a(t)$. There exist $C_n \to C$ such that $A(n)^{-1}A(n+1) = e^{C_{n+1}}$ for $n \geq n_0$. See Section 18.2 for details. Let $\delta > 0$ be the minimum of $w^T C w - \delta_0$ on ∂B. Choose $n_1 \geq n_0$ such that $\|C_n - C\| < \delta$ for $n \geq n_1$. Define Q on $[n_1, \infty)$ by interpolation, $Q(t) = Q(n)e^{(t-n)C_{n+1}}$ for $n \leq t \leq n+1, n \geq n_1$. For $t \in [0, n_1]$ set $Q(t) = Q(n_1)\gamma^{t-n_1}$. Then $Q(t) \sim \alpha(t)$ by Proposition 18.6. Set $C_n = C$ for $n \leq n_1$. Then $w^T C_n w \geq \delta_0$ for $w \in \partial B$ and $n \geq 1$. This implies $e^{s\delta_0}(B) ß e^{sC_{n+1}}(B)$ for $s \geq 0$ and $n \geq 0$, and hence $e^{s\delta_0} E_t ß E_{t+s}$ for $n \leq t \leq t+s \leq n+1, n \geq 0$, since $Q(t+s) = Q(t)e^{sC_{n+1}}$ for these values. Now use the implication

$$e^{s\delta_0} E_r ß E_{r+s} \text{ and } e^{t\delta_0} E_{r+s} ß E_{r+s+t} \quad \Rightarrow \quad e^{(r+s)\delta_0} ß E_{r+s+t}$$

to obtain the results in the theorem. □

Henceforth we assume that the normalizations α_t like the expansions γ^t are linear transformations.

Let us now turn to the polar representation. The inclusion (16.9) allows us to map the part of π outside E_0 to the half-cylinder $\Gamma_+ = \partial B \times [0, \infty)$. The limit relation (16.1) then is equivalent to asymptotic equality of this image measure to the measure $\tilde{\rho} = \Phi(\rho)$ introduced in (16.5) under vertical translation.

Assume (16.9). We define a homeomorphism Ψ between the upper cylinder $\Gamma_+ = \partial B \times [0, \infty)$ and the complement of the ellipsoid E_0 by

$$\Psi\colon (p, t) \mapsto \alpha(t)p, \quad \Psi\colon \Gamma_+ = \partial B \times [0, \infty) \to \mathbb{R}^d \setminus E_0. \tag{16.12}$$

The image of $1_{E_0^c} d\pi$ under Ψ^{-1} is a finite measure μ on Γ. The limit relation (16.1) is equivalent to the relation

$$e^t \tau^{-t}(\mu) \to \tilde{\rho} \text{ weakly on } \Gamma_+, \quad t \to \infty \tag{16.13}$$

where $\tilde{\rho} = \Phi(\rho)$, see (16.6), and $\tau^t(\theta, s) = (\theta, s+t)$ for $(\theta, s) \in \Gamma$.

One may use the non-linear transformation Ψ to describe the domain $\mathcal{D}^\infty(\rho)$.

Theorem 16.15 (Polar Representation). *Let $\alpha\colon [0, \infty) \to \mathrm{GL}(d)$ be continuous and vary like the linear expansion group γ^t, $t \in \mathbb{R}$. Assume the unit ball B is adapted, and (16.9) holds for the ellipsoids $E_t = \alpha(t)(B)$. Let π be a probability measure on $\mathbb{R}^d$ with $\pi(E_0) < 1$, and let μ on Γ be the image of $1_{E_0^c} d\pi$ under Ψ^{-1}, see (16.12). Then $\pi \in \mathcal{D}^\infty(\rho)$ if and only if (16.13) holds.*

Proof. The basic idea is that $\mu_n \to \mu$ weakly on a compact metric space $\mathcal{X}$ implies $\rho_n = \Phi_n(\mu_n) \to \rho = \Phi(\mu)$ weakly on the compact metric space $\mathcal{Y}$ if Φ_n and Φ are homeomorphisms from $\mathcal{X}$ onto $\mathcal{Y}$ and $\Phi_n \to \Phi$ uniformly. The last condition implies $\Phi_n^{-1} \to \Phi^{-1}$ uniformly, and hence the converse implication also holds. (If

$y_n = \Phi_n(x_n) \to y = \Phi(x)$ then $\Phi(x_0) = y$ for any limit point x_0 of the sequence (x_n), and hence $x_n \to x$.)

Apply this with $\Phi_n = \beta_{t_n}^{-1}\Psi\tau^{t_n}$ where $t_n \to \infty$, and $\mathcal{X} = \Gamma_+ \cup \{\infty\}$ and $\mathcal{Y} = (\mathbb{R}^d \cup \{\infty\}) \setminus E_0$. If $(\zeta_n, s_n) \to (\zeta, s) \in \Gamma_+$ then

$$\Phi_n(\zeta_n, s_n) = \beta_{t_n}^{-1}\Psi(\zeta_n, s_n + t_n) = \beta_{t_n}^{-1}\beta_{t_n+s_n}(\zeta_n) \to \gamma^s(\zeta) = \Phi(\zeta, s).$$

If $s_n \to \infty$ then for any $r > 0$ eventually $\Phi_n(\zeta_n, s_n) \notin \gamma^r(B)$ since $\beta_{t_n}^{-1}\beta_{t_n+s}(B)$ß$\gamma^r(B)$ eventually for given $0 < r < s$. So $\Phi_n \to \Phi$ uniformly. Set $\mu = \Psi^{-1}(\pi)$ and $\mu_t = e^t\tau^{-t}(\mu)$. Then

$$\rho_{t_n} = e^{t_n}\beta_{t_n}^{-1}(\pi) = e^{t_n}\beta_{t_n}^{-1}\Psi\tau^{t_n}(e^{t_n}\mu_{t_n}) = \Phi(\mu_{t_n}).$$

Let $d\rho_n = 1_{E_0^c}d\rho_{t_n}$, $d\mu_n = 1_{\Gamma_+}d\mu_{t_n}$, $d\rho_\infty = 1_{E_0^c}d\rho_0$, $d\mu_\infty = 1_{\Gamma_+}d\Phi^{-1}(\rho_0)$. Then $\rho_n \to \rho_\infty$ if and only if $\mu_n \to \mu_\infty$. The same arguments hold with E_0 replaced by an ellipsoid E_t and Γ_+ by $\partial B \times [-c, \infty)$. □

The non-linear transformation $\pi \mapsto \mu = \Psi^{-1}(\pi)$ reveals an asymptotic product measure which satisfies: $e^t\tau^{-t}d\mu \to d\rho_0^* \times e^{-t}dt$. If one starts with a product measure $d\mu = d\rho^* \times e^{-t}dt$ on $\partial B \times [0, \infty)$, where ρ^* is a probability measure on ∂B, then $\pi = \Psi(\mu)$ will be called the *typical distribution* in $\mathcal{D}^\infty(\rho)$ associated with the spectral probability measure ρ^*, and the normalizing curve $\beta\colon [0, \infty) \to \mathrm{GL}(d)$.

16.4 Convex hulls and convergence. For heavy tailed vectors the convex hulls of the normalized sample clouds converge. It does not matter what the convex support of the distribution looks like. The orbits of the symmetry group of the excess measure are curved in general, except for scalar expansions. Hence, even if the excess measure has a continuous density h, the domain $\{h > 0\}$ need not be convex. The convex hull of the sample cloud need not give a good indication of the shape of the domain!

The normalized sample clouds N_n converge in distribution to the Poisson point process N with mean measure ρ:

$$N_n \Rightarrow N \quad \text{vaguely on } \mathbb{R}^d \setminus \{0\}.$$

Weak convergence holds on εB^c for any $\varepsilon > 0$ and on all halfspaces J which do not contain the origin. No mass escapes to infinity. Let us check that ρ does not charge the boundary of εB or of J. Observe that ρ may charge hyperplanes through the origin, since there are proper linear subspaces which are invariant.

Lemma 16.16. *Let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be analytic and vanish in the origin. Let ρ be an excess measure for expansions. Then*

$$\rho\{\varphi = c\} = 0, \quad c \neq 0.$$

Proof. Let $T = \{\gamma^t(w_0) \mid t \in \mathbb{R}\}$ be the orbit containing the point $w_0 \neq 0$. The function $t \mapsto \varphi_0(t) = \varphi(\gamma^t(w_0))$ is analytic on $\mathbb{R}$ (since the Jordan coordinates of $\gamma^t(w_0)$ are analytic in t). If $\varphi_0(t_n) = c$ for a sequence $t_n \to t$ with $t_n \neq t$ then $\varphi_0 \equiv c$ by a power series development in the point $t \in \mathbb{R}$. Hence $\varphi \equiv c$ holds on the orbit T, and since $\gamma^{-n}(w_0) \to 0$ the origin lies in the closed set $\{\varphi = c\}$. Hence $c = 0$. Now assume $c \neq 0$, and let $\Sigma = \Phi^{-1}\{\varphi = c\}$ be the image of the level surface $\{\varphi = c\}$ in the tube Γ, see (16.5), and $\mu = \Phi^{-1}(\rho)$. By the argument above the intersection of the level surface with any orbit T is countable. Hence so is the intersection of Σ with any vertical line $\{\theta\} \times \mathbb{R}$ in Γ. By Fubini $\mu(\Sigma) = 0$. Hence $\rho\{\varphi = c\} = 0$. □

Suppose ρ is an excess measure on $\mathbb{R}^d \setminus \{0\}$ for the expansion group γ^t. Let π be a probability measure, and let $\rho_t = e^t\alpha(t)^{-1}(\pi) \to \rho$ vaguely on $\mathbb{R}^d \setminus \{0\}$, where $\alpha\colon [0,\infty) \to \mathrm{GL}(d)$ varies like γ^t. We shall show that *vague convergence* implies weak convergence on the complement of any ball εB with $\varepsilon > 0$.

Proposition 16.17. *Let $\alpha\colon [0,\infty) \to \mathrm{GL}$ vary like the expansion group $\gamma^t = e^{tC}$. Let π be a probability measure on $\mathbb{R}^d$. Vague convergence $\rho_t = e^t\alpha(t)^{-1}(\pi) \to \rho$ on $\mathbb{R}^d \setminus \{0\}$ implies weak convergence on εB^c for $\varepsilon > 0$. Let τ_0 be the maximum of the real parts of the eigenvalues of the generator C above. Then*

$$\int \varphi d\rho_t \to \int \varphi d\rho, \quad t \to \infty \tag{16.14}$$

for an arbitrary continuous function φ which vanishes on a neighbourhood of 0*, if $\varphi(w)/(1 + \|w\|^\delta)$ is bounded for some $\delta < 1/\tau_0$.*

Proof. Assume B is adapted. Let $R = \gamma(B) \setminus B$, $E_t = \beta_t(B)$ and $R_t = E_{t+1} \setminus E_t$. Let $\varepsilon > 0$. Since $\rho(\partial\gamma^t B) = 0$ by the lemma above

$$e^t\pi(R_t) = \rho_t\alpha(t)^{-1}(E_{t+1} \setminus E_t) \to \rho(R), \quad t \to \infty.$$

Eventually $e^t\pi(R_t) < 2\rho(R)$. This implies

$$\rho_t(\alpha(t)^{-1}(R_{t+m})) = e^t\pi(R_{t+m}) \leq 2e^{-m}\rho(R), \quad t \geq t_0$$

and hence by summing $\rho_t(\alpha(t)^{-1}(E^c_{t+m}) \leq Ce^{-m}$ for $t \geq t_0$, $m \geq 0$. Choose m so large that $Ce^{-m} < \varepsilon$. Choose $r > 1$ so large that $rB \supset \mathrm{cl}(\gamma^m(B))$. Then $\alpha(t)^{-1}(E_{t+m}) \to \gamma^m(B)$ implies $\alpha(t)^{-1}(E^c_{t+m}) \supset rB^c$ eventually and hence $\rho_t(\alpha(t)^{-1}(rB^c)) \leq Ce^{-m} < \varepsilon$. This proves weak convergence on B^c. The proof of the second part is similar. □

Theorem 16.18. *Let ρ be an excess measure for an expansion group with generator C. Let $\tau > 0$. Assume $\Re\theta_i > \tau$ for all eigenvalues θ_i of C. Let $\pi \in \mathfrak{D}^\infty(\rho)$, with* *normalizations $\alpha(t)$, and let $\alpha_n = \alpha(t_n)$ with $e^{t_n} \sim n$. One may choose n-point sample clouds N_n from the distribution $\alpha_n^{-1}(\pi)$, and a Poisson point process N on $\mathbb{R}^d \setminus \{0\}$ with mean measure ρ, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that*

1) *the convex hull of $N_n(\omega)$ converges to the convex hull of $N(\omega)$ for all $\omega \in \Omega$;*

2) *for any ρ-a.e. continuous function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ which vanishes on a neighbourhood of the origin*

$$\int \varphi dN_n \to \int \varphi dN \quad a.s. \text{ and in } \mathbf{L}^1(\Omega, \mathcal{F}, \mathbb{P})$$

if φ/ψ is bounded for $\psi(w) = 1 + \|w\|^{1/\tau}$.

Proof. See Section 5.7. Weak convergence $\rho_t = \alpha(t)^{-1}(\pi)/\pi(\alpha(t)(B^c)) \to \rho$ holds on halfspaces which do not contain the origin. Moreover $\rho_t(\varepsilon B) \to \infty$ for all $\varepsilon > 0$. Hence the convex hull of the normalized sample cloud N_n will intersect εB with probability $p_n \to 1$. In terms of the theory developed in Section 5.7 the intrusion cone Δ and the convergence cone Γ coincide with the dual space of $\mathbb{R}^d$. The boundary condition (S) in (5.12) holds for $S = \{0\}$. Hence the convex hulls of the sample clouds converge to the convex hull of the limiting Poisson point process by Theorem 5.26. The second statement follows from the proposition above, see Section 5.7. □

Exercise 16.19. How do the normalized sample clouds behave for exceedances over horizontal thresholds? Convergence is steady if the excess measure ρ is sturdy. The excess measure is sturdy if it charges the open upper halfspace. ◊

16.5 Typical densities. In this section we assume that the excess measure ρ has a continuous positive density g on $\mathbb{R}^d \setminus \{0\}$. The measure ρ satisfies $\gamma^t(\rho) = e^t\rho$, $t \in \mathbb{R}$, and α is a normalization curve which varies like γ^t. We consider two questions:

1) Does there exist a probability measure π with a continuous density, such that $e^t\alpha(t)^{-1}(\pi) \to \rho$? How does one construct such a probability measure π? Do the densities converge too?

2) Suppose we have a continuous positive probability density f, and the quotients $f \circ \alpha(t)/f(\alpha(t)(a_0))$ converge to the corresponding quotient $g/g(a_0)$ of the excess density. Do the probability distributions converge? Do the densities converge?

Definition. Let γ^t, $t \in \mathbb{R}$, be a linear expansion group, ρ an excess measure such that $\gamma^t(\rho) = e^t\rho$ for $t \in \mathbb{R}$, and let $\alpha\colon [0,\infty) \to \mathrm{GL}(d)$ vary like γ^t. A probability density f_0 is a *typical density* for α and ρ if f_0 is continuous and positive on $\mathbb{R}^d$, if ρ has a continuous positive density g on $\mathbb{R}^d \setminus \{0\}$, and if

$$e^{t_n}|\det\alpha(t_n)|f_0(\alpha(t_n)(w_n)) \to g(w), \quad t_n \to \infty,\ w_n \to w \neq 0. \tag{16.15}$$

Remark 16.20. If the density f is continuous and positive, and asymptotic to f_0 in ∞, then f also is a typical density. If $\beta(t) \sim \alpha(t)$ for $t \to \infty$, then f_0 also is a typical density for β and ρ.

Proposition 16.21. *If the density f of the probability distribution π on $\mathbb{R}^d$ agrees with a typical density f_0 for α and ρ outside a bounded set then*

$$e^t \alpha(t)^{-1}(\pi) \to \rho \text{ vaguely on } \mathbb{R}^d \setminus \{0\}.$$

Proof. It suffices to prove that the density of the left hand side converges to the density of the right hand side uniformly on compact sets. The density of the left is $e^t |\det \alpha(t)| f \circ \alpha(t)$; on the right g. So let $w_n \to w \neq 0$ and $t_n \to \infty$. We have to prove (16.15) for f. This follows since for $\varepsilon > 0$ any bounded set is contained in $\alpha(t_n)(\varepsilon B)$ eventually, and hence $f(\alpha(t_n)(w_n)) = f_0(\alpha(t_n)(w_n))$ eventually. □

Proposition 16.22. *Let $g_0 \colon \partial B \to (0, \infty)$ be continuous. Let $\alpha \colon [0, \infty) \to \mathrm{GL}$ be continuous and vary like γ^t, and let the ellipsoids $E_t = \alpha(t)(B)$ satisfy $\mathrm{cl}(E_s) \text{ß} E_t$ for $0 \leq s < t$. Define g on $\mathbb{R}^d \setminus \{0\}$ and f_0 on $\mathbb{R}^d \setminus E_0$ by*

$$g(\gamma^t(w)) = g_0(w)/q^t, \quad t \in \mathbb{R},\ w \in \partial B,\ q = e \det \gamma$$
$$|\det \alpha(t)| f_0(\alpha(t)(w)) = g_0(w)/e^t, \quad t \geq 0,\ w \in \partial B.$$

Then g is continuous and positive on $\mathbb{R}^d \setminus \{0\}$, and f_0 is continuous and positive on E_0^c. The function g is the density of an excess measure ρ which satisfies $\gamma^t(\rho) = e^t \rho$ for $t \in \mathbb{R}$, and which is finite on εB^c for all $\varepsilon > 0$, and f_0 is integrable over E_0^c. Any positive continuous probability density f on $\mathbb{R}^d$ which agrees with f_0 outside a bounded set is typical for α and ρ.

Proof. Let $t_n \to \infty$ and $w_n \to w_0 \neq 0$. Write

$$\alpha(t_n)(w_n) = \alpha(t_n + s_n)(a_n), \quad a_n \in \partial B,\ s_n \in \mathbb{R},\ n \geq n_0.$$

Then $w_n = \alpha(t_n)^{-1}\alpha(t_n + s_n)(a_n) \to \gamma^{s_0}(a_0)$ implies $s_n \to s_0$ and $a_n \to a_0$. (If $s_n \leq s < s_0 - \delta$ infinitely often then

$$w_n \in \alpha(t_n)^{-1}\alpha(t_n + s_n)(\mathrm{cl}(B)) \text{ß} \alpha(t_n)^{-1}\alpha(t_n + s)(\mathrm{cl}(B)) \text{ß} \gamma^{s_0 - \delta}(B)$$

infinitely often. Hence w_n cannot converge to a point in $\gamma^{s_0}(\partial B)$. So $s_n > s_0 - \delta$ eventually. Similarly one proves that $s_n < s_0 + \delta$ eventually. Hence $s_n \to s_0$. This implies $a_n \to a_0$.) By definition of f_0 one may write

$$e^{t_n} |\det \alpha(t_n)| f_0(\alpha(t_n)(w_n)) = e^{t_n} |\det \alpha(t_n)| f_0(\alpha(t_n + s_n)(a_n))$$
$$= e^{-s_n} c_n g_0(a_n) \to g(w_0)$$

with $c_n = |\det \alpha(t_n + s_n)^{-1}\alpha(t_n)|$. Convergence holds since $c_n \to \det \gamma^{-s_0}$ and $e^{-s_0} g_0(a_0)/\det \gamma^{s_0} = g(w_0)$. □

Now suppose π has a continuous positive density f for which the *quotients* converge

$$\frac{f(\alpha(t_n)(w_n))}{f(\alpha(t_n)(a_0))} \to \frac{g(w_0)}{g(a_0)}, \quad t_n \to \infty,\ w_n \to w_0,\ w_0 \neq 0 \tag{16.16}$$

for some point $a_0 \in \partial B$.

Proposition 16.23. *Under the conditions of Proposition 16.22 above there is a C^1 function $\lambda\colon [0,\infty) \to \mathbb{R}$ with $\lambda(0) = 0$, and with a positive derivative which tends to 1 in infinity, such that f is typical for $\beta(t) = \alpha(\lambda(t))$ and ρ.*

Proof. Let $a_0 \in \partial B$. Define

$$F(t) := e^t |\det\alpha(t)| f(\alpha(t)(a_0)), \quad t \geq 0.$$

Let $t_n \to \infty$, $s_n \to s \in \mathbb{R}$. Then $\alpha(t_n + s_n)(a_0) = \alpha(t_n)(w_n)$ with $w_n = \gamma^{r_n}(a_n)$ for $a_n \in \partial B$. Convergence of $w_n = \alpha(t_n)^{-1}\alpha(t_n + s_n)(a_0)$ to $\gamma^s(a_0)$ implies $r_n \to s$ and $a_n \to a_0$ by continuity of the polar coordinates map Φ^{-1}. Observe that $c_n = |\det\alpha(t_n + s_n)|/|\det\alpha(t_n)| \to \det\gamma^s$. Hence

$$\begin{aligned} F(t_n + s_n)/F(t_n) &= e^{s_n} c_n f(\alpha(t_n)(w_n))/f(\alpha(t_n)(a_0)) \\ &\to e^s \det\gamma^s g(\gamma^s(a_0))/g(a_0) = 1. \end{aligned}$$

Hence F varies like 1. The function F is asymptotic to a function e^{φ} where φ is C^1 and $\dot\varphi$ vanishes in ∞. Then $e^{t+\varphi(t)}|\det\alpha(t)| f_0(\alpha(t)(a_0)) \equiv g(a_0)$, and (16.16) gives

$$e^{t_n + \varphi(t_n)} |\det\alpha(t_n)| f(\alpha(t_n)(w_n)) \to g(w_0), \quad t_n \to \infty,\ w_n \to w_0 \neq 0.$$

We may alter φ on a finite interval so that $\varphi(0) = 0$ and $\dot\varphi(t) > -1$. Now let λ be the inverse of $t \mapsto t + \varphi(t)$. □

The *time changed* curve $\beta(t) = \alpha(\lambda(t))$ above has the same properties as α: it varies like γ^t, is continuous, and has the same family of ellipsoids: $\beta(t)(B) = E_{\lambda(t)}$.

16.6 Roughening and vague convergence. Let $\alpha\colon [0,\infty) \to \mathrm{GL}$ be continuous and vary like γ^t for a group of linear expansions. Assume that the ellipsoids $E_t = \alpha(t)(B)$ satisfy $\mathrm{cl}(E_s)$ßE_t for $0 \leq s < t$. Introduce the continuous family of ellipsoids

$$F_z = \begin{cases} E_t/3 & \text{if } z \in \partial E_t \text{ for some } t \geq 0, \\ E_0/3 & \text{if } z \in E_0. \end{cases} \tag{16.17}$$

Recall that a Radon measure μ is a *roughening of Lebesgue measure* for the ellipsoids $z + F_z$, $z \in \mathbb{R}^d$, if there exists a partition of $\mathbb{R}^d$ in bounded Borel sets A_n of positive volume, $|A_n| > 0$, such that any bounded set intersects only finitely many sets A_n, such that $\mu(A_n) \sim |A_n|$, and such that for any $\varepsilon > 0$ eventually

$$A_n \text{ß} z + \varepsilon F_z, \quad z \in A_n.$$

This section is devoted to the proof of the following result:

Theorem 16.24. *For a Radon measure μ on $\mathbb{R}^d$ the following statements are equivalent:*

1) *μ is a roughening of Lebesgue measure with respect to the ellipsoids F_z;*
2) *$\mu_t = \alpha(t)^{-1}(\mu)/|\det \alpha(t)| \to \lambda$ vaguely on $\mathbb{R}^d$ for $t \to \infty$.*

Proof. Here we shall prove that $\mu_t \to \lambda$ vaguely if μ is a roughening. For the converse we need some extra material. So assume μ is a roughening of Lebesgue measure, and $A_0, A_1, \ldots$ the corresponding partition. Let φ be continuous with compact support. We have to show that $\int \varphi d\mu_t \to \int \varphi d\lambda$. We may and shall assume $0 \le \varphi \le 1$. Introduce a function χ which is constant on each atom A_n, with the value $\mu A_n/\lambda A_n$. Let $\nu = \lambda$ on the atoms where χ vanishes, and $\nu = \mu/\chi$ on the remaining atoms. Then $\nu A_n = \lambda A_n$ for all n, and $|\mu - \nu|(A_n) = o(\lambda A_n)$. We shall prove that

$$\int \varphi d\nu_t \to \int \varphi d\lambda; \quad \int \varphi d|\mu - \nu|_t \to 0. \tag{16.18}$$

The first limit is standard. For any $t \ge 0$ let $\delta(t)$ denote the maximum of the diameters of the sets $\alpha(t)^{-1}(A_n)$ which intersect the support S of φ. It suffices to prove that $\delta(t) \to 0$. The integrals then converge by uniform continuity of φ, with the arguments used for Riemann integrals. Here are the details for showing that $\delta(t) \to 0$:

Choose $s_0 \ge 0$ so large that $S \text{ß} \gamma^{s_0}(B)$. For each n let z_n be a point in A_n, and set $A_n^0 = A_n - z_n$. Let $\varepsilon > 0$. There exists t_0 such that for $t \ge t_0$:

1) $S_t := \alpha(t)(S) \text{ß} E_{t+s_0}$;
2) $\alpha(t)^{-1}(E_{t+s_0}) \text{ß} \gamma^{s_0+1}(B)$;
3) $A_n^0 \text{ß} \varepsilon E_t$ if A_n intersects E_t.

Let $t \ge t_0$. Suppose $\alpha(t)^{-1}(A_n)$ intersects S. Then A_n intersects S_t, hence E_{t+s_0} by 1). Hence $A_n^0 \text{ß} \varepsilon E_{t+s_0}$ by 3). Hence $\alpha(t)^{-1}(A_n^0) \text{ß} \varepsilon \gamma^{s_0+1}(B)$ by 2). Hence $\operatorname{diam}(\alpha(t)^{-1}(A_n)) \le \varepsilon \operatorname{diam}(\gamma^{s_0+1}(B))$. This proves the first limit relation in (16.18).

Now consider the second limit. Let $\eta > 0$ be small. The set $E = \{|\chi - 1| > \eta\}$ is a finite union of atoms. We have the inequalities:

$$\int_E (\varphi \circ \alpha(t))|\chi - 1| d\nu \le \int_E \chi + 1 d\nu = \int_E \chi d\lambda + \lambda(E) = C(\eta),$$
$$\int_{E^c} (\varphi \circ \alpha(t))|\chi - 1| d\nu \le \eta \int \varphi \circ \alpha(t) d\nu = |\det \alpha(t)| \eta \int \varphi d\nu_t.$$

So $\int \varphi d|\nu - \mu|_t \leq \eta \int \varphi d\nu_t + C(\eta)/|\det \alpha(t)|$. The first term may be made small by choosing η small. The second term may then be made small by choosing t large. □

In order to prove that μ is a roughening if $\mu_t \to \lambda$ vaguely, we have to do some extra work. We shall need the following property of a sequence of Radon measures μ_n which converges vaguely to Lebesgue measure: If D_n are Borel sets contained in the ball rB, each D_n the difference of two convex Borel sets, and if there exists a constant $\delta > 0$ such that $\lambda(D_n) > \delta$ eventually, then $\mu_n(D_n) \sim \lambda(D_n)$. This follows from the proposition below.

Proposition 16.25. *Let μ_n be Radon measures on the open set $O\text{ß}\mathbb{R}^d$ which converge vaguely to Lebesgue measure λ on O. Let $K\text{ß}O$ be compact, and let $C_n\text{ß}K$ be convex Borel sets. Then $\mu_n(C_n) - \lambda(C_n) \to 0$.*

Proof. We may assume that $\mu_n(C_n) \to c_0$ and $\lambda(C_n) \to c_1$. We claim that $c_0 = c_1$.

First suppose $C_n \to C$, and $|C| > 0$. We may assume C is closed. It has an interior point p_0. Let C^r be the set C blown up by a factor $r > 0$ from the center p_0. So $C^r = p_0 + r(C - p_0)$. Vague convergence implies $\mu_n(C^r) \to \lambda(C^r)$ whenever $C^r\text{ß}O$. Since $C^{1-\varepsilon}\text{ß}C_n\text{ß}C^{1+\varepsilon}$ eventually for any $\varepsilon > 0$ we conclude that $\mu_n(C_n) \to \lambda(C)$.

The remainder of the proof is a compactness argument. We spell it out. If there is a proper affine subspace M such that $C_n\text{ß}M + \varepsilon B$ holds infinitely often for each $\varepsilon > 0$, then $c_1 = c_0 = 0$.

Henceforth assume $c_0 + c_1 > 0$. Choose a point in each set C_n. This sequence (or a subsequence) converges to a point z_0. We may find a set of $d + 1$ limit points $z_0, \ldots, z_d$ in K whose convex hull is a simplex Σ with interior points. By taking appropriate subsequences we may assume that each of these $d + 1$ points is limit of a sequence of points in C_n. This then holds for all points in Σ. Any interior point of Σ lies in C_n eventually.

Let $A\text{ß}K$ be countable and dense. By a diagonal argument we may extract a subsequence (k_n) such that $F_n(a) := 1_{C_{k_n}}(a)$ converges for each $a \in A$. Let F be the closure of the set of $a \in A$ for which the limit is one, and let U be the interior of F. Then U contains the interior of the simplex Σ above. Each point in U lies in C_{k_n} eventually. Hence U is convex. If $z \in C_{k_n}$ infinitely often, then this also holds for all $a \in A$ in the interior of the convex hull of z and U. Hence these a lie in U. Hence $z \in F$. So $C_n \to F \in \mathcal{C}$, and we may use the opening argument to conclude that $c_0 = c_1$. □

In order to prove that a Radon measure μ is a roughening if $\mu_t \to \lambda$ vaguely, we need to construct a partition. The atoms A of our partition will have the form $A = E_0$, or $A = R \cap K$ where R is a ring, $R = E_t \setminus E_s$ with $s < t$, and K a convex cone. So let us start by partitioning $\mathbb{R}^d \setminus \{0\}$ into convex cones.

Let q be a positive integer. We want to define a partition $\mathcal{C}_q$ of $\mathbb{R}^d \setminus \{0\}$ into cones. The surface of the cube $(-1, 1)^d$ consists of $2d$ squares. Partition each into subsquares of side length $1/2^q$. There are $2d2^h2^{hq}$ such squares on the boundary of the cube $(-1, 1)^d$. Let $\mathcal{C}_q$ consist of the cones generated by these squares.

We still have to say a few words about the boundary points of the squares. The large square $(-1, 1]^h$ can easily be subdivided into 2^{h+qh} subsquares congruent to $(0, 1/2^q]^h$. So we try to partition the boundary of the cube $(-1, 1)^d$ into $2d$ squares congruent to $(-1, 1]^h$. For $d = 3$ it is not possible to write the boundary of the cube as the disjoint union of six squares $(-1, 1]^2$. There are eight vertices! However, one can divide the boundary of $(-1, 1)^d$ into the disjoint sets

$$S_{ik} = (-1, 1)^{i-1} \times \{k\} \times [-1, 1]^{d-i}, \quad i = 1, \dots, d, \ k = \pm 1.$$

Divide the open interval $(-1, 1)$ into $2^{q+1} - 1$ disjoint intervals $(a, b]$ of length $1/2^q$, and one final open interval $(1 - 1/2^q, 1)$. There is a similar partition for the closed interval $[-1, 1]$. Now use the product partition on the squares S_{ik} to obtain $\mathcal{C}_q$.

We shall now introduce a finite partition $\mathcal{A}(\gamma)$ of the ring $\gamma(B) \setminus B$. The *atoms* of this partition are cells of the form $A = R_k \cap K$ where K is one of the $2d2^{h+qh}$ cones above, and R_k the ring $\gamma^{(k+1)/2^q}(B) \setminus \gamma^{k/2^q}(B)$ for $k = 0, \dots, 2^q - 1$. We need a lower bound for the volume V of the atom A, and an upper bound for the diameter D.

Lemma 16.26. *Suppose $\gamma^t = e^{tC}$, $t \in \mathbb{R}$. Define*

$$\delta_0 = \min_{w \in \partial B} w^T C w, \quad \eta_0 = \max_{w \in \partial B} w^T C w. \tag{16.19}$$

Then the volume V and the diameter D of the atoms of the partition above satisfy

$$V \geq (\delta_0/2)(2^{-dq}/d^{d/2}), \quad D \leq (\eta_0 + \sqrt{h})\|\gamma\|/2^q. \tag{16.20}$$

Proof. Let $0 \leq s_0 < s_1 \leq 1$. Let $t_i w \in \partial E_{s_i}$ for $i = 0, 1$ for a vector $w \in \partial B$. Then

$$(s_1 - s_0)\delta_0 \leq t_1 - t_0 \leq \|\gamma\|(s_1 - s_0)\eta_0. \tag{16.21}$$

The volume V of $K \cap (\gamma^{s_1}(B) \setminus \gamma^{s_0}(B))$ may be expressed as an integral

$$V = \frac{1}{d} \int_S t_1(p)^h - t_0(p)^h dp$$

where S is the square generating K and $t_i(p)p \in \partial \gamma^{s_i}(B)$. Since $\|p\| \leq \sqrt{d}$ the inequality (16.21) gives $t_1(p) - t_0(p) \geq (s_1 - s_0)\delta_0/\sqrt{d}$, and $t_0 \geq 1/\sqrt{d}$. Hence $V \geq (h/d)|S|(s_1 - s_0)\delta_0/d^{d/2}$. This gives the first inequality in (16.20). Now take two points $p_0 = r_0 w_0$ and $p_1 = r_1 w_1$ in the cell A, where the w_i are unit vectors. Since the vectors w_i lie in $K \cap \partial B$, and Bß$(-1,1)^d$, it follows that $\|w_1 - w_0\| \leq \sqrt{h}/2^q$, the diameter of the square S. We find

$$\begin{aligned}\|p_1 - p_0\| &\leq \|p_1 - r_0 w_1\| + \|r_0 w_1 - r_0 w_0\| \\ &= r_1 - r_0 + r_0\|w_1 - w_0\| \leq \eta_0\|\gamma\|/2^q + \|\gamma\|\sqrt{h}/2^q.\end{aligned}$$

This yields the second inequality in (16.20). □

Proof of Theorem 16.24, *part* 2). We shall now prove that a Radon measure μ is a roughening of Lebesgue measure if $\mu_t \to \lambda$ vaguely. We first formulate some extra conditions on the normalizations $\alpha(t)$. See Theorem 16.14. Let $\alpha_n = \alpha(n)$. Then $\alpha_n = \alpha_0\gamma_1 \ldots \gamma_n$, and $\gamma_n = e^{C_n}$. We assume that $\alpha(t) = \alpha_n \gamma_{n+1}^{t-n}$ for $n \leq t \leq n+1$, $n \geq 0$, and that $\|\gamma_n\| \leq 2\|\gamma\|$, and all C_n are close to C. Define δ_n and η_n as in the lemma above, but now for C_n. We assume that $\delta_n \geq \delta_0/2$ for $n \geq 1$, and $\eta_n \leq 2\eta_0$. These conditions do not affect the asymptotic behaviour of $\alpha(t)$ for $t \to \infty$.

Let the partition $\mathcal{A}_0$ consist of E_0 and the rings $E_{n+1} \setminus E_n$. For $q \geq 1$ define $\mathcal{A}_q$ to consist of E_0 and the sets $\alpha_n(D)$, $n \geq 0$, where Dß$\gamma_n(B) \setminus B$ is an atom in $\mathcal{A}(\gamma_n)$. The atoms A of $\mathcal{A}_q$ are intersections of convex cones $\alpha_n(K)$, with $K \in \mathcal{C}_q$, and rings $E_{(m+1)/2^q} \setminus E_{m/2^q}$, $m \geq 0$. We claim that $\mu(A_n) \sim \lambda(A_n)$. Let $D_n = \alpha_{k_n}^{-1}(A_n)$ be the corresponding atom in the partition of $\gamma_n(B) \setminus B$. By the lemma above the volume of D_n is bounded below by $(\delta_0/4)(2^{-dq}/d^{d/2})$ since $\delta_n \geq \delta_0/2$. Proposition 16.25 gives $\mu(A_n)/\lambda(A_n) = \mu_{k_n}(D_n)/\lambda(D_n) \to 1$ since $\mu_{k_n} \to \lambda$ vaguely on $\mathbb{R}^d$, and the sets D_n are differences of convex sets and contained in the compact closure of $2\|\gamma\|B$.

Note that $\mathcal{A}_{q+1}$ is a refinement of $\mathcal{A}_q$. Define $\mathcal{A}$ as the partition which agrees with $\mathcal{A}_{q_n}$ on the ring $E_{n+1} \setminus E_n$, where q_n increases to ∞ so slowly that the asymptotic equality $\mu(A_n)/\lambda(A_n) \to 1$ holds for the atoms of $\mathcal{A}$. The diameter of the corresponding sets $D_n = \alpha(k_n)^{-1}(A_n)$ß$\gamma_{k_n}(B) \setminus B$ tends to zero since $q_n \to \infty$. Choose $z_n \in A_n$, and let $A_n^0 = A_n - z_n$. Then $\alpha_{k_n}^{-1}(A_n^0)$ß$\varepsilon_n B$ for a sequence $\varepsilon_n \to 0$. Hence A_n^0ß$\varepsilon_n E_{k_n}$. Now observe that E_{k_n}ßF_z for $z \in A_n$. So the condition $\mu_t \to \lambda$ vaguely implies that μ is a roughening of Lebesgue measure for the ellipsoids F_z and the partition $\mathcal{A}$. □

16.7 A characterization. This section contains a simple characterization of the domain of attraction of excess measures which have a continuous positive density.

We start with a linear expansion group γ^t, $t \in \mathbb{R}$. Choose coordinates such that the unit ball is *adapted*. Let $\alpha\colon [0,\infty) \to \mathrm{GL}$ vary like γ^t. By Theorem 16.14 we may assume that α is continuous and that the ellipsoids $E_t = \alpha(t)(B)$ satisfy $\mathrm{cl}(E_s)$ßE_t for $0 \leq s < t$. Let ρ be an excess measure with a continuous positive density g on $\mathbb{R}^d \setminus \{0\}$. The symmetry relations $\gamma^t(\rho) = e^t\rho$ hold for $t \in \mathbb{R}$. Equivalently, by (14) in the Preview

$$g(\gamma^t(w)) = g(w)/q^t, \quad w \neq 0,\ t \in \mathbb{R},\ q = e \det \gamma.$$

Define the function f_0 on E_0^c by

$$|\det \alpha(t)|\, f_0(\alpha(t)(w)) = g(w)/e^t, \quad w \in \partial B,\ t \geq 0.$$

By Proposition 16.22 a *typical density* for α and ρ is any continuous positive probability density f which is asymptotic to f_0 in ∞. Such densities satisfy

$$e^{t_n} |\det \alpha(t_n)| f(a(t_n)(w_n)) \to g(w), \quad t_n \to \infty,\ w_n \to w \neq 0. \tag{16.22}$$

The corresponding probability distribution π satisfies the basic limit relation:

$$e^t \alpha(t)^{-1}(\pi) \to \rho \text{ weakly on } \varepsilon B^c, \quad t \to \infty,\ \varepsilon > 0. \tag{16.23}$$

Let $z + F_z$ be a continuous family of ellipsoids, centered in $z \in \mathbb{R}^d$, which describes the scale around the point z. We assume that the ellipsoids diverge for $z \to \infty$. As in Section 16.6 we define a Radon measure μ to be a *roughening of Lebesgue measure* for the ellipsoids $z + F_z$ if there exists a partition of $\mathbb{R}^d$ in bounded Borel sets A_n of positive volume, $|A_n| > 0$, such that any bounded set intersects only finitely many sets A_n, such that $\mu(A_n) \sim |A_n|$, and such that for any $\varepsilon > 0$ eventually

$$A_n \text{ß} z + \varepsilon F_z, \quad z \in A_n.$$

We can characterize the domain of attraction for excess measures with positive continuous densities: Any probability distribution in the domain $\mathcal{D}^\infty(\rho)$ of such an excess measure may be expressed as $f d\mu$ where f is a typical density and μ is a roughening of Lebesgue measure for the ellipsoids determined by the normalizations.

Theorem 16.27. *Let γ^t, $t \in \mathbb{R}$, $\alpha \colon [0, \infty) \to \mathrm{GL}$ and ρ be as above. Let f be a typical density for α and ρ. Any probability distribution π which satisfies* (16.23) *has the form $d\pi = f d\mu$ for a Radon measure μ on $\mathbb{R}^d$ which is a roughening of Lebesgue measure with respect to the family of ellipsoids F_z in* (16.17)*. If the Radon measure μ on $\mathbb{R}^d$ is a roughening of Lebesgue measure with respect to the ellipsoids F_z above, then any probability measure π which agrees with $f d\mu$ outside a bounded set will satisfy* (16.23).

Before giving the proof, let us discuss the conditions.

The conditions on the normalizations $\alpha(t)$ enable us to give a straightforward definition of the ellipsoids F_z, since each $z \in E_0^c$ lies on the boundary of a unique ellipsoid E_t. The factor 3 in the definition of the ellipsoids is not strictly necessary. It ensures that $F_{z'}$ is comparable to F_z for all $z' \in z + F_z$. If we leave it out then the origin lies in the closure of $z + F_z$ for all z, and it is difficult to interpret the ellipsoid $z + F_z$ as the *local scale*.

As in Chapter III one may think of the family of ellipsoids $z + F_z$ as a *geometry*. The family determines a Riemannian metric in terms of the quadratic functions Q_z, defined by $F_z = \{Q_z < 1\}$. Since Q depends continuously on z it enables us to define the length of curves given by piecewise smooth functions $\varphi\colon [0,1] \to \mathbb{R}^d$:

$$L(\Gamma_\varphi) = \int_0^1 \sqrt{\dot{\varphi}^T Q_{\varphi(t)} \dot{\varphi}(t)}\,dt, \quad \Gamma_\varphi = \{\varphi(t) \mid 0 \le t \le 1\}.$$

The length does not depend on the parametrisation. One introduces a metric by defining $d_0(z_0, z_1)$ as the infimum of $L(\Gamma)$ over all curves Γ leading from z_0 to z_1. Continuity of $z \mapsto Q_z$ ensures that d_0 satisfies the axioms of a metric, and agrees with the topology. We refer to Section 11.4 for details. Here we only want to observe that the metric does not depend on the coordinates. Focus on a point $z \in \partial E_t$ with t large, and regard E_t as the unit ball. In these coordinates E_{t+s} is the ellipsoid $\alpha(t)^{-1}\alpha(t+s)(B) \approx \gamma^s(B)$. The Riemannian metric associated with the renormalized ellipsoids $\alpha(t)^{-1}E_{t+s}$ is the image of d_0 under the linear map $\alpha(t)$. It is defined by

$$d_t(w_0, w_1) = d_0(\alpha_t(w_0), \alpha_t(w_1)), \quad w_0, w_1 \in \mathbb{R}^d.$$

Convergence of the ellipsoids $\alpha(t)^{-1}\alpha(t+s)(B)$ to $\gamma^s(B)$ for $t \to \infty$ may be shown to imply convergence of the metrics $d_t \to d_\infty$ where d_∞ on $\mathbb{R}^d \setminus \{0\}$ is the Riemannian metric associated with the family $w + G_w$, and $G_w = \gamma^s(B)/3$ for $w \in \gamma^s(\partial B)$. The metric d_∞ is invariant under γ^t:

$$\gamma(w + G_w) = z + G_z, \quad z = \gamma^t(w).$$

The ellipsoid F_z determines the scale around the point z. The ellipsoid $z + F_z$ is not the unit ball around z in the Riemannian metric d_0, but one gets a good estimate of the distance between z_0 and z by taking $\varepsilon > 0$ small, and counting the number of points $z_0, \dots, z_m = z$ which are needed to ensure that the ellipsoids $z_i + \varepsilon F_{z_i}$ form a chain linking z_0 to z. The Riemannian distance, $d_0(z_0, z)$, is approximately $2\varepsilon m$. In particular, if we intersect the elliptic surface ∂E_t with a two-dimensional linear subspace, we obtain a closed curve in ∂E_t. The length of this curve in the Riemannian metric d_0 is 6π, since one may choose coordinates such that E_t is the unit ball, and the sets $z + \varepsilon F_z$, $z \in \partial E_t$, then are balls of radius $\varepsilon/3$ centered in points on the surface of the unit ball.

In Section 16.2 we introduced the polar representation, a homeomorphism $\Phi\colon (w,t) \mapsto \gamma^t(w)$ from the cylinder $\partial B \times \mathbb{R}$ to $\mathbb{R}^d \setminus \{0\}$. Similarly, the homeomorphism $\Psi\colon (t, w) \mapsto \alpha(t)(w)$ in Section 16.3 maps the half-cylinder $\partial B \times [0, \infty)$ onto $\mathbb{R}^d \setminus E_0$. One can transfer the Euclidean metric from the half-cylinder to a metric d

on $\mathbb{R}^d \setminus E_0$, by setting $d(z_0, z_1) = \|(w_0, t_0) - (w_1, t_1)\|$ for $z_i = \Psi(w_i, t_i)$. The metric d agrees with $3d_0$ on the surfaces ∂E_t, but the metrics d and $3d_0$ are not identical on E_0^c. If the ellipsoids E_t are balls the complex eigenvalues of γ all have the same absolute value. The metric d_0 is determined by the balls; the metric d will also depends on the imaginary part of the eigenvalues of C.

We prefer to work with the family of ellipsoids $z + F_z$, rather than the metric d or d_0. We are only interested in the asymptotics. For our purpose one may replace F_z by any family of ellipsoids F'_z which is asymptotic to F_z for $\|z\| \to \infty$. For many applications it suffices that there exists a constant $M > 1$ such that

$$(1/M)F_z \text{ß} F'_z \text{ß} M F_z, \quad \|z\| \geq M. \tag{16.24}$$

The geometry $z + F_z$, $z \in \mathbb{R}^d$, allows us to introduce flat functions and roughenings of Lebesgue measure.

Recall from Section 11.1 that a function $L\colon \mathbb{R}^d \to (0, \infty)$ is *flat* if it is asymptotically constant on the ellipsoids $z + F_z$:

$$L(z'_n)/L(z_n) \to 1, \quad \|z_n\| \to \infty, \ z'_n \in z_n + F_{z_n}.$$

Here we do need the constant $3 > 1$ in the definition of the ellipsoids F_z in (16.17). The rings $R_t = E_{t+1} \setminus E_t$ may be covered by a finite number of ellipsoids $z + F_z$ with $z \in R_t$, and this number has a bound which does not depend on t. It follows that flat functions are asymptotically constant on such rings. Hence for any flat function L one may introduce a function $L_0 = e^{\lambda_0}\colon [0, \infty) \to (0, \infty)$, such that

$$L(\alpha(t_n)w_n) \sim L_0(t_n), \quad t_n \to \infty, \ w_n \in \partial B.$$

We may choose L_0 such that λ_0 is C^1 with derivative $\lambda'_0(t)$ which vanishes for $t \to \infty$. See Section 11.2.

If we modify the *typical density* f_1 by a flat function L, the new density $f_2 = Lf_1$ satisfies

$$e^{t_n - \lambda_0(t_n)} |\det \alpha(t_n)| f_2(\alpha(t_n)(w_n)) \to g(w), \quad t_n \to \infty, \ w_n \to w \neq 0.$$

We may then introduce normalizations $\beta(t) = \alpha(\tau(t))$ related by a *time change*, such that (16.22) holds for f_2 with α replaced by β. Flat functions do not affect convergence, but may affect the normalization curve via a time change.

Now turn to roughenings of Lebesgue measure. The conditions imply that on the scale of the ellipsoids $z + F_z$ for large z there is not much difference between the roughening μ and Lebesgue measure λ.

Example 16.28. Suppose μ is a roughening of Lebesgue measure, and (A_n) an associated partition. If μ has a density h, the only condition on h is that the average of h over an atom A_n should tend to one for $n \to \infty$. Given the partition (A_n) one may also define a roughening μ by picking a point z_n in each set A_n, and putting mass $|A_n|$ into this point z_n.

Given an enumeration $p_0, p_1, \dots$ of the points with integer coordinates, one may choose μ to be the counting measure of the set $\mathbb{Z}^d$, and A_n the box $p_n + (-1/2, 1/2]^d$. Since E_t eventually covers any bounded set, this partition satisfies the conditions for a roughening for the ellipsoids F_z. If f is a typical density for α and ρ then $\sum_{k \in \mathbb{Z}^d} f(k)$ converges, with sum $c \in (0, \infty)$, and the random vector Z with integer valued components, which assumes the value $k \in \mathbb{Z}^d$ with probability $p_k = f(k)/c$ lies in $\mathcal{D}^\infty(\rho)$ with the normalizations $\alpha(t)$. ◇

After this digression on flat functions and roughenings we now come to the proof of Theorem 16.27.

Proof. First assume π lies in $\mathcal{D}^\infty(\rho)$. The measure $(1/f_1)d\pi$ is finite on compact sets (since $1/f_1$ is continuous on such sets), and agrees with μ outside a bounded set. Hence we may assume that $d\mu = (1/f_1)d\pi$. Then

$$d\mu_t := \alpha(t)^{-1}(d\mu)/|\det \alpha(t)| = (1/g_t)e^t \alpha(t)^{-1}(d\pi), \quad g_t = e^t |\det \alpha(t)| f \circ \alpha(t).$$

By Proposition 16.22 above $g_t \to g$ uniformly on compact subsets of $\mathbb{R}^d \setminus \{0\}$ for $t \to \infty$. This implies $1/g_t \to 1/g$ uniformly on compact subsets of $\mathbb{R}^d \setminus \{0\}$. The second factor $\rho_t = e^t \alpha(t)^{-1}(\pi)$ converges vaguely to ρ on $\mathbb{R}^d \setminus \{0\}$. It follows that $(1/g_t)d\rho_t \to (1/g)d\rho = d\lambda$ vaguely on $\mathbb{R}^d \setminus \{0\}$. Here λ denotes Lebesgue measure on $\mathbb{R}^d$. So $\mu_t \to \lambda$ vaguely on $\mathbb{R}^d \setminus \{0\}$.

We need convergence on $\mathbb{R}^d$. Observe that $\mu_t(B) = \mu(E_t)/|\det \alpha(t)|$. The condition on μ implies that $\mu(E_t) \sim |E_t|$ for $t \to \infty$, since for $m \geq m_0$ there exist finite unions of atoms A_k, U_m and V_m such that $E_{m-1}\text{ß}U_m\text{ß}E_m\text{ß}V_m\text{ß}E_{m+1}$, and $\mu(A_n) \sim |A_n|$ implies $\mu(U_m) \sim |U_m|$ and $\mu(V_m) \sim |V_m|$ provided $|U_m| \to \infty$ and $|V_m| \to \infty$ for $m \to \infty$. Thus $\mu_t(B) \to |B|$. The same relation holds for εB for any $\varepsilon > 0$. It follows that $\mu_t \to \lambda$ vaguely on $\mathbb{R}^d$. Finally apply the implication 2) $\Rightarrow$ 1) in Theorem 16.24.

Now suppose μ is a roughening of Lebesgue measure, and f_1 is typical. We may assume that $d\pi = f_1 d\mu$ by altering μ on a bounded set. Then

$$e^t \alpha(t)^{-1}(\pi) = e^t |\det \alpha(t)| f(\alpha(t)) d(\alpha(t)^{-1}(\mu))/|\det \alpha(t)| = g_t d\mu_t.$$

Since $\mu_t \to \lambda$ vaguely (by Theorem 16.24) and $g_t \to g$ uniformly on compact subsets of $\mathbb{R}^d \setminus \{0\}$ the right side tends to $g d\lambda$ vaguely for $t \to \infty$. □

For the *Euclidean Pareto* excess measure ρ_τ with density $1/\|w\|^{d+1/\tau}$ it is possible to give a geometric description of $\mathcal{D}^\infty(\rho_\tau)$. As observed in the Preview, it suffices to specify a sequence of ellipsoids E_n such that $E_{n+1} \sim 2E_n$. Such sequences determine the *typical densities*, and the geometry, F_z, $z \in \mathbb{R}^d$.

16.8* Interpolation of ellipsoids, and twisting. If $\alpha\colon [0,\infty) \to \mathrm{GL}$ varies like γ^t, then it is possible to recover the curve, up to asymptotic equality, from the sequence $(\alpha(n))$ by a simple interpolation procedure. For ellipsoids it is not clear how such a procedure should be defined. Even if one has the complete family of ellipsoids E_t, $t \geq 0$, there may be many continuous curves $\alpha\colon [0,\infty) \to \mathrm{GL}$, such that $\alpha(t)(B) = E_t$, since one may replace $\alpha(t)$ by $\alpha(t)S(t)$ where the $S(t)$ map the unit ball onto itself. Such a transformation of the curve α will be called twisting.

We shall first discuss *twisting*. Replace $\alpha(t)$ by $\beta(t) = \alpha(t) \circ S(t)$ where $S\colon [0,\infty) \to \mathrm{O}(d)$ is continuous and varies like I

$$S_{t_n}^{-1} S_{t_n+s_n} \to I, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}.$$

Choose $t_n \to \infty$ such that $S(t_n) \to S$ for some S in the compact group $\mathrm{O}(d)$. Then $S(t_n + s_n) \to S$ if s_n is bounded and

$$\begin{aligned}\beta(t_n)^{-1}\beta(t_n+s_n) &= S(t_n)^{-1}\alpha(t_n)^{-1}\alpha(t_n+s_n)S(t_n)(S(t_n)^{-1}S(t_n+s_n))\\ &\to S^{-1}\gamma^s S, \quad t_n \to \infty,\ s_n \to s \in \mathbb{R}.\end{aligned}$$

Let Σ be the set of limit points of $S(t)$ for $t \to \infty$. This is a closed connected subset of $\mathrm{O}(d)$. The curve β varies like γ^t if

$$S^{-1}\gamma^s S = \gamma^s, \quad S \in \Sigma,\ s \in \mathbb{R}. \tag{16.25}$$

The condition is satisfied for scalar expansion groups.

Now let us look at rotation expansions. Suppose $\gamma^t\colon w \mapsto e^{\tau t}w$, and $S\colon [0,\infty) \to \mathrm{O}(d)$ varies like the rotation group R^t, $t \in \mathbb{R}$. The argument above applied to $\beta(t) = \alpha(t)S(t)$ gives

$$\beta(t_n)^{-1}\beta(t_n+s_n) \to \gamma^s R^s.$$

So β varies like the rotation expansion $e^{\tau t}R^t$, $t \in \mathbb{R}$. If ρ has density h, and scalar symmetries $\gamma^t = e^{\tau t}$, then the measure $\bar{\rho}$ with density $\bar{h}(e^{\tau t}\theta) = h(e^{\tau t}R^{-t}\theta)$ is an excess measure with symmetries $e^{\tau t}R^t$ and the densities $\bar{h}$ and h agree on ∂B:

$$\bar{h}(e^{\tau t}R^t\theta) = h(e^{\tau t}\theta) = e^{-t}h(\theta), \quad \theta \in \partial B.$$

The normalizations $\alpha(t)$ and $\beta(t)$ define the same family of ellipsoids E_t. One obtains the typical density in $\mathcal{D}^\infty(\bar{\rho})$ from the density in $\mathcal{D}^\infty(\rho)$ by rotating along the boundaries of these ellipsoids.

We now turn to the problem of *interpolation*. Given two centered ellipsoids E_0 and E_1 with $\mathrm{cl}(E_0)$ßE_1 is there a natural interpolation? Does there exist a one-parameter group of linear transformations α^t such that $E_1 = \alpha(E_0)$ and E_0 is *adapted*? The ellipsoids $E_t = \alpha^t(E_0)$ then satisfy $\mathrm{cl}(E_s)$ßE_t for $s < t$.

Choose coordinates such that E_0 is the unit ball and

$$E_1 = \left\{ \frac{z_1^2}{a_1^2} + \cdots + \frac{z_d^2}{a_d^2} < 1 \right\}, \quad 1 < a_1 \leq \cdots \leq a_d. \tag{16.26}$$

Take $\alpha^t = \mathrm{diag}(a_1^t, \ldots, a_d^t)$, $t \in \mathbb{R}$. The ellipsoids

$$E_t = \alpha^t(B) = \left\{ \frac{z_1^2}{a_1^{2t}} + \cdots + \frac{z_d^2}{a_d^{2t}} < 1 \right\}$$

satisfy the requirements. This procedure was used in Section 12.1 to embed a growing sequence of ellipsoids with $E_{n+1} \sim cE_n$ into a continuous family of ellipsoids E_t, $t \geq 0$, such that

$$\mathrm{cl}(E_s)\text{ß}E_t, \quad 0 \leq s < t; \quad E_{t_n+s_n} \sim c^s E_{t_n}, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}.$$

The interpolation recipe above is geometric. In dimension $d = 2$ the numbers $a_1 \leq a_2$ are determined by the inclusions

$$a_1 E_0 \text{ß} E_1 \text{ß} a_2 E_0$$

with a_1 maximal and a_2 minimal (since this holds in the special coordinates considered above).

There are other interpolations. We restrict attention to the plane. If E_0 and E_1 have the same Euclidean shape then $E_1 = cRE_0$ for a rotation R, and one may use rotational expansions $c^t R^t$ to define E_t. One may also use *shear* expansions. For simplicity assume the centered ellipses E_0 and E_1 in the plane have equal area. Consider a line tangent to both. Assume this line is vertical. Then E_0 and E_1 both fit in a common vertical strip $\{|x| < b\}$ and the two ellipses are related by the one-parameter group of shears along the vertical axis. Now introduce an expansion factor $c > 1$. If it is large enough the map $t \mapsto c^t E_t$ will be increasing.

Example 16.29. Let $\alpha = rA = r\,\mathrm{diag}(1, a)$ with $r, a > 1$, and let R be a rotation over the angle θ. Then

$$\beta = rAR = r \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} = r \begin{pmatrix} c & s \\ -as & ac \end{pmatrix}, \quad c = \cos\theta, s = \sin\theta.$$

The nature of the group β^t is determined by the characteristic polynomial of AR

$$P(\lambda) = \det(AR - \lambda I) = (c - \lambda)(ac - \lambda) + as^2 = \lambda^2 - c(1 - a)\lambda + a.$$

We distinguish four cases:

1) The characteristic polynomial has two distinct positive zeros $\lambda_1 < \lambda_2$. There are two independent eigenvectors e_1 and e_2 and $(AR)^t e_i = \lambda_i^t e_i$. Now β^t is a family

of diagonal matrices $\mathrm{diag}(\lambda_1^t, \lambda_2^t)$ with respect to the base (e_1, e_2). As θ moves away from $\theta_0 = 0$ the eigenvalues of β move closer together and the eigenvectors move away from the positive quadrant until for $\cos\theta = 2\sqrt{a}/(1+a)$ both lie along the line $y = -\sqrt{a}x$ and the eigenvalues coincide.

2) There is a double zero $\lambda_0 = \sqrt{a}$. In this case there is only one eigenvector $e_0 = (1, -\sqrt{a})$, and β^t is a family of shear expansions along the line $\mathbb{R}e_0$.

3) When $\cos\theta$ decreases below the value $2\sqrt{a}/(1+a)$ there is a pair of conjugate complex zeros $\lambda \pm i\mu$ with $\mu \neq 0$. There is a base of complex eigenvectors $e \pm if$. In coordinates where e and f are orthonormal β^t is an expansive rotation group.

4) For $\cos\theta \leq -2\sqrt{a}/(1+a)$ the roots λ_1 and λ_2 of the characteristic polynomial are negative. There is no one-parameter group β^t such that $\beta = \alpha R$. ♢

16.9 Spectral decomposition, the basic result. This section contains the proof of a basic result in the Spectral Decomposition Theorem. For a discussion of the theorem, a formulation in terms of regular variation, and an extension to affine transformations, we refer to Section 18.4.

If the invertible matrix Q has no eigenvalues on the unit circle in $\mathbb{C}$, then $\mathbb{R}^d$ is the sum of two invariant linear subspaces, U and V, of dimension d_U and d_V, with $d_U + d_V = d$, such that Q is an expansion on U and a contraction on V. We may write $Q = Q_U \otimes Q_V$ where $Q(u, v) = (Q_U u, Q_V v)$, identifying $U \times V$ with $\mathbb{R}^d$. The eigenvalues of Q_U lie outside the unit circle, those of Q_V inside.

The behaviour of the trajectories $w, Qw, Q^2w, Q^3w, \dots$ is simple. If $w \in V$ then $Q^n w \to 0$; if $w \in V^c$ then $\|Q^n w\| \to \infty$. We shall see that the trajectories z, $A_1 z$, $A_2 z, \dots$ show a similar dichotomy in their behaviour for sequences of linear maps $A_n = Q_n \dots Q_1 A_0$ when $Q_n \to Q$.

Below we give a proof of this basic Spectral Decomposition Theorem. The result is linear algebra. The notation in this section differs from that in the remainder of the book. We work with linear transformations, and it is convenient to write $Q = \gamma^{-1}$ and $A_n = \alpha_n^{-1}$. Thus $\alpha_n = \alpha_0\gamma_1 \dots \gamma_n$ becomes $A_n = Q_n \dots Q_1 A_0$, and $\alpha_n^{-1}\alpha_{n+1} \to \gamma$ becomes $A_{n+1}A_n^{-1} \to Q$.

The reader may want to keep the two-dimensional case in mind and think in terms of matrices of size two. Any matrix of size two with two distinct positive eigenvalues is the direct sum of an expansion and a contraction after multiplication by a suitable constant $r > 0$. We choose r so that the eigenvalues of the new matrix satisfy $0 < r_V < 1 < r_U$. Actually the d-dimensional case is not more complicated than the two-dimensional case.

It is convenient to assume that A_0 is an invertible linear map from an inner product space L onto $\mathbb{R}^d$. The Q_n are invertible matrices of size d, and hence $A_n \colon L \to \mathbb{R}^d$ is invertible for each n. We shall prove that there is a decomposition $L = X + Y$ into orthogonal subspaces, such that $\dim(X) = \dim(U) = d_U$ and $\dim(Y) =$

$\dim(V) = d_V$, and such that A_n may be replaced by block diagonal matrices $B_n \otimes E_n$ where $B_n : X \to U$ and $E_n : Y \to V$ satisfy the same relations as A_n, but on lower dimensional spaces:

$$B_{n+1}B_n^{-1} \to Q_U, \quad E_{n+1}E_n^{-1} \to Q_V.$$

Theorem 16.30. *Suppose $A_0 : L \to \mathbb{R}^d$ is an invertible linear transformation from the d-dimensional inner product space L to $\mathbb{R}^d$. Let $Q_n \to Q$ with $Q = Q_U \otimes Q_V$ as above. Write*

$$A_n = Q_n \dots Q_1 A_0, \quad n \geq 0.$$

Set $Y = \{z \in L \mid A_n z \to 0\}$ and $X = Y^\perp = \{x \in L \mid x \perp Y\}$. Then $\dim(X) = \dim(U)$ *and* $\dim(Y) = \dim(V)$. *Write*

$$A_n : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} B_n x + C_n y \\ D_n x + E_n y \end{pmatrix}.$$

Set $\hat{A}_n = B_n \otimes E_n : (x, y) \mapsto (B_n x, E_n y)$. Then

$$\hat{A}_n A_n^{-1} \to I, \quad B_{n+1}B_n^{-1} \to Q_U, \quad E_{n+1}E_n^{-1} \to Q_V.$$

The proof below is due to Meerschaert & Scheffler [2001].

For simplicity we assume

$$\|Qu\| \geq 4\|u\|, \quad \|Qv\| \leq \|v\|/4. \tag{16.27}$$

Since $Q_n \to Q$ there exists n_0 such that

$$\|Q_n - Q\| =: \varepsilon_n \leq \varepsilon := 1/4, \quad n \geq n_0. \tag{16.28}$$

For any $z \in L$ we write

$$A_n z = w_n = (u_n, v_n), \quad r_n = \|u_n\|, \; s_n = \|v_n\|.$$

We shall now first investigate the limit behaviour of these sequences w_n. All results in this section derive from the two simple inequalities below:

Lemma 16.31. *For all $n \geq 0$*

$$\begin{aligned} r_{n+1} &\geq 4r_n - \varepsilon_{n+1}(r_n + s_n) \\ s_{n+1} &\leq s_n/4 + \varepsilon_{n+1}(r_n + s_n). \end{aligned}$$

Proof. The triangle inequality for the norm gives

$$\|Q_{n+1}(u_n, v_n) - Q(u_n, v_n)\| \leq \|Q_{n+1} - Q\|\|w_n\| \leq \varepsilon_{n+1}(r_n + s_n).$$

Now use (16.27). □

Corollary 16.32. *For $n \geq n_0$*

$$r_{n+1} \geq 3r_n - s_n, \quad s_{n+1} \leq s_n/2 + \varepsilon_{n+1} r_n. \tag{16.29}$$

In $\mathbb{R}^d$ we introduce the open set

$$O = \{w = (u, v) \in \mathbb{R}^d \mid \|u\| > \|v\|\}.$$

Lemma 16.33. *If $A_m z = w_m \in O$ for some $m \geq n_0$ then $w_{m+1} \in O$, and $r_{m+1} > 2r_m$.*

Proof. The assumption $w_m \in O$ implies $s_m < r_m$, and hence $r_{m+1} > 2r_m$ by (16.29). Also

$$s_{m+1} \leq s_m/2 + r_m/4 \leq r_m < r_{m+1}$$

by (16.29), the assumption $s_m < r_m$, and our inequality $r_{m+1} > 2r_m$. Hence $w_{m+1} \in O$. □

Proposition 16.34. *Let $z \in L$, and set $w_n = A_n z$. If $w_n \in O^c$ for all $n \geq n_0$ then $w_n \to 0$; if $w_m \in O$ for some $m \geq n_0$ then $w_n \in O$ for $n \geq m$ and $r_n \to \infty$ and $s_n/r_n \to 0$.*

Proof. Suppose $s_n \geq r_n$ for $n \geq n_0$. Then (16.29) gives $s_{n+1} \leq 3s_n/4$, and hence $s_n \to 0$. Because $r_n \leq s_n$ by assumption, it follows that $w_n \to 0$. If $w_m \in O$ for some $m \geq n_0$ then $w_{m+1} \in O$ by the lemma above, and $r_{m+1} \geq 2r_m$. By induction $w_n \in O$, and $r_{n+1} \geq 2r_n$ for all $n \geq m$. It remains to prove that $s_n/r_n \to 0$. Let $n = m + j \geq m$. By (16.29)

$$\begin{aligned} s_{n+1} &\leq \varepsilon_{n+1} r_n + s_n/2 \leq \varepsilon_{n+1} r_n + \varepsilon_n r_{n-1}/2 + s_{n-1}/4 \\ &\leq \cdots \leq \varepsilon_{n+1} r_n + \varepsilon_n r_{n-1}/2 + \cdots + \varepsilon_{m+1} r_m/2^j + s_m/2^{j+1}. \end{aligned}$$

Let $\eta \in (0, 1)$ be small. Choose k so large that $4^{-k} < \eta$. Then by (16.28)

$$\frac{s_{n+1}}{r_{n+1}} \leq \frac{\varepsilon_{n+1}}{2} + \cdots + \frac{\varepsilon_{n+1-k}}{2^{2k+1}} + \frac{1}{4}\left(\frac{1}{2^{2k+3}} + \cdots + \frac{1}{2^{2j+1}}\right) + \frac{1}{2^{2j+2}}.$$

The sum in brackets is less than $\eta/2$, the remaining $k + 2$ terms each vanish for $n = m + j \to \infty$. Hence $s_n/r_n < \eta$ eventually. It follows that $s_n/r_n \to 0$. □

Remark 16.35. The rate of convergence $s_n/r_n \to 0$ depends only on the sequence ε_n, and the integer m. Let $O_m = \{z \in L \mid A_m z \in O\}$ for $m \geq n_0$. Proposition 16.34 gives a uniform bound on the rate of convergence $s_n/r_n \to 0$ for $z \in O_m$.

Remark 16.36. The definition of O depends on the norm on U and V. Suppose $\delta \in (0, 1]$. Set

$$|(u, v)| = \sqrt{\|u\|^2/\delta^2 + \|v\|^2}.$$

The arguments above remain valid, but now for the set $O_\delta = \{\|u\| > \delta\|v\|\}$, and for $m \geq n_0(\delta)$.

Each map A_n induces a decomposition $L = U_n + V_n$ where $U_n = A_n^{-1}(U)$ and $V_n = A_n^{-1}(V)$. The behaviour of the linear subspaces U_n may be quite erratic, but the subspaces V_n converge, as we shall see.

Proposition 16.37. *Define the linear subset* YßL *by*

$$Y = \{z \in L \mid A_n z \to 0\}.$$

Then $\dim(Y) = \dim(V)$, *and* $V_n \to Y$. *For any* $z \in Y$ *the sequence* $w_n = A_n z$ *satisfies* $r_n/s_n \to 0$. *More precisely, there exists a sequence* $\eta_n \to 0$, *such that* $r_n \le \eta_n s_n$ *holds uniformly in* $z \in Y$.

Proof. The open sets $O_m = \{z \in L \mid A_m z \in O\}$, $m \ge n_0$, are increasing, and cover $L \setminus Y$ by Proposition 16.34. Choose an orthonormal basis $e_{m1}, \dots, e_{md_V}$ in $V_m = A_m^{-1}(V)$. Since ∂B is compact there is a subsequence $m_1 < m_2 < \cdots$ such that $e_{m_n i} \to e_i$ for $i = 1, \dots, d_V$. The linear space V_∞ spanned by the orthonormal basis $e_1, \dots, e_{d_V}$ is disjoint from O_m since V_n is disjoint from $O_n \supset O_m$ for $n \ge m$. So V_∞ß$Y = (\bigcup O_n)^c$, and $\dim(Y) \ge d_V$. The set $O_{n_0} \cup \{0\}$ contains $U_{n_0} = A_{n_0}^{-1}(U)$. Since Y is disjoint from O_{n_0} it follows that $Y \cap U_{n_0} = \{0\}$, and hence $\dim(Y) \le d - d_U = d_V$. Conclusion $\dim(Y) = d_V$. By a subsequence argument the sequence V_n converges to Y.

Let $z \in Y$. Set $A_n z = (u_n, v_n)$. Proposition 16.34 shows that $r_n \le s_n$ for $n \ge n_0$ since $A_n z \in O^c$. By the second remark above $A_n z \in O_\delta^c$ for $n \ge n_0(\delta)$, and hence $r_n \le \delta s_n$ for $n \ge n_0(\delta)$. Since this holds for any $\delta > 0$ it follows that $r_n/s_n \to 0$. The bounds $n_0(\delta)$ do not depend on z. □

We thus have the following description of the behaviour of the sequence $(u_n, v_n) = A_n z$ in terms of $r_n = \|u_n\|$ and $s_n = \|v_n\|$. If $z \in Y$ then $r_n/s_n \to 0$ uniformly in $z \in Y$; if $z \in X = Y^\perp$ then $s_n/r_n \to 0$ uniformly in $z \in X$, even uniformly for $z \in O_m$ for any given $m \ge n_0$ (by the first remark above).

Write A_n as a blocked matrix with respect to the decompositions $L = X + Y$ and $\mathbb{R}^d = U + V$:

$$A_n \colon \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} u_n \\ v_n \end{pmatrix} = \begin{pmatrix} B_n x + C_n y \\ D_n x + E_n y \end{pmatrix}.$$

The blocked diagonal matrix $\hat{A}_n = B_n \otimes E_n$ maps $x + y \in L$ into $B_n x + E_n y \in U + V = \mathbb{R}^d$.

Proposition 16.38. $A_n \hat{A}_n^{-1} \to I$.

Proof. We have to show that $D_n B_n^{-1} \to 0$ and $C_n E_n^{-1} \to 0$. Choose $u_n \in U \cap \partial B$ such that $\|D_n B_n^{-1}\| = \|D_n B_n^{-1} u_n\|$, and set $x_n = B_n^{-1} u_n \in X$. Then $A_n x_n = B_n x_n + D_n x_n = u_n + v_n$. Write $r_n = \|u_n\|$ and $s_n = \|v_n\|$. Then $r_n = 1$ by assumption, and $s_n = s_n/r_n \to 0$ by Proposition 16.37. Since $s_n = \|D_n\|$ this proves $D_n \to 0$. The proof that $C_n E_n \to 0$ is similar. □

Proof of Theorem 16.30. There is a positive integer m such that Q^m satisfies (16.27). Hence $A_{nm}\hat{A}_{nm}^{-1} \to I$. If $A_{k_n}\hat{A}_{k_n}^{-1} \to I$ then $A_{k_n+i}A_{k_n}^{-1} \to Q^i$ implies $A_{k_n+i}\hat{A}_{k_n}^{-1} \to Q^i$, and, since Q^i is block diagonal, this in turn implies $\hat{A}_{k_n+i}\hat{A}_{k_n}^{-1} \to Q^i$. Hence $A_n\hat{A}_n^{-1} \to I$, and $\hat{A}_{n+1}\hat{A}_n^{-1} \to Q$. □

We now return to the original notation. Let $\gamma \in \mathrm{GL}(d)$ and let $0 < s_1 < \cdots < s_q$ be positive numbers such that γ has no eigenvalues on the q circles with radius s_k in $\mathbb{C}$. There exist invariant linear subspaces $U_0, \dots, U_q$ of dimension $d_k = \dim(U_k) \geq 0$ with sum $d_0 + \cdots + d_q = d$, which span $\mathbb{R}^d$, and linear maps $\gamma^{(k)}: U_k \to U_k$ such that $\gamma^{(k)}(u) = \gamma(u)$ for $u \in U_k$, and such that all eigenvalues of $\gamma^{(k)}$ lie between the circles with radius s_k and s_{k+1}, where we set $s_0 = 0$ and $s_{q+1} = \infty$. Identifying $\mathbb{R}^d$ with the product $U_0 \times \cdots \times U_q$ we may write $\gamma = \gamma^{(0)} \otimes \cdots \otimes \gamma^{(q)}$. This decomposition is obvious if one writes γ in *Jordan form*.

If $\alpha_n = \alpha_0\gamma_1 \dots \gamma_n$, where $\alpha_0: L \to \mathbb{R}^d$ is an invertible linear map from the d-dimensional vector space L to $\mathbb{R}^d$, and γ_n are invertible matrices converging to γ, then there is a corresponding decomposition $L = X_0 + \cdots + X_q$ with $\dim(X_k) = \dim(U_k) = d_k$ for $k = 0, \dots, q$, and there exist $\beta_n(k): U_k \to X_k$ such that

$$\alpha_n \sim \beta_n(0) \otimes \beta_n(q) \quad \beta_n(k)^{-1}\beta_{n+1}(k) \to \gamma^{(k)}, \quad n \to \infty,\ k = 0, \dots, q.$$

If L is an inner product space one may choose the subspaces X_k to be orthogonal. Under this extra condition they are unique. Full details are given in the Spectral Decomposition Theorem below. The proof for the case $q = 1$ and $s_1 = 1$ was given above with $Q = \gamma^{-1}$ and $A_n = \alpha_n^{-1}$. The general result follows by a repeated decomposition, starting with $Q = s_k\gamma^{-1}$, for some k and then proceeding with a decomposition of Q_U or Q_V. After q steps we have the following result:

Theorem 16.39 (Discrete Spectral Decomposition Theorem). *Let $\gamma \in \mathrm{GL}(d)$ and $0 < s_1 < \cdots < s_q$. Assume γ has no eigenvalues on any of the q circles with radius s_k in $\mathbb{C}$. Define $U_0, \dots, U_q$ and $\gamma^{(0)}, \dots, \gamma^{(q)}$ as above, and assume that $d_k = \dim(U_k)$ is positive for $k = 0, \dots, q$. Let L be a d-dimensional inner product space, and $\alpha_0: \mathbb{R}^d \to L$ an invertible linear map. Let*

$$\alpha_n = \alpha_0\gamma_1 \dots \gamma_n, \quad \gamma_n \in \mathrm{GL}(d),\ \gamma_n \to \gamma.$$

We have the following results:

1) *There exist orthogonal subspaces $X_0, \dots, X_q$ which span L such that*

$$X_k + \cdots + X_q = \{z \in L \mid s_k^n\alpha_n^{-1}(z) \to 0\}, \quad k = 1, \dots, q.$$

The limit relation above together with the orthogonality determines the subspaces $X_0, \dots, X_q$.

2) *If* z *does not lie in* $X_k + \cdots + X_q$ *then* $\|s_k^n \alpha_n^{-1}(z)\| \to \infty$, *for* $k = 1, \ldots, q$.

3) $\dim(X_k) = \dim(U_k) = d_k$ *for* $k = 0, \ldots, q$.

4) *Write*

$$\alpha_n = \begin{pmatrix} \alpha_n(00) & \ldots & \alpha_n(0q) \\ \vdots & \ddots & \vdots \\ \alpha_n(q0) & \ldots & \alpha_n(qq) \end{pmatrix}, \quad \alpha_n(ij)\colon U_j \to X_i .$$

Then

$$\alpha_n(u_0 + \cdots + u_q) = x_0 + \cdots + x_q, \quad x_i = \sum_{j=0}^{q} \alpha_n(ij) u_j ,$$

and

$$\alpha_n \sim \hat{\alpha}_n = \operatorname{diag}(\alpha_n(00), \ldots, \alpha_n(qq))\colon u_0 + \cdots + u_q \mapsto \alpha_n(00)u_1 + \cdots + \alpha_n(qq)u_q .$$

5) *Moreover* $\alpha_n(kk)^{-1}\alpha_{n+1}(kk) \to \gamma^{(k)}$ *for* $n \to \infty$, $k = 0, \ldots, q$.

6) *We may write* $\hat{\alpha}_n = \alpha_n(00) \otimes \cdots \otimes \alpha_n(qq)$ *if we identify* L *with* $X_0 \times \cdots \times X_q$ *and* $\mathbb{R}^d$ *with* $U_0 \times \cdots \times U_q$. *Then*

$$\hat{\alpha}_n^{-1}\hat{\alpha}_{n+1} \to \gamma = \gamma^{(0)} \otimes \cdots \otimes \gamma^{(q)} .$$

17 Heavy tails – examples

In this section we consider the domain $\mathcal{D}^\infty(\rho)$ for excess measures ρ with diagonal symmetries γ^t. We shall look at three cases: scalar symmetries and scalar normalizations; scalar symmetries and non-scalar normalizations; diagonal symmetries and diagonal normalizations. The rotational component due to the imaginary part of the eigenvalues of the generator, may be incorporated in our description of the domain of attraction of excess measures with scalar symmetries. In view of the Spectral Decomposition Theorem this yields a complete picture of the domains of attraction of excess measures whose generators have a basis of complex eigenvectors. Excess measures with maximal symmetry are treated in Section 17.5.

Section 17.6 is devoted to multivariate stable distributions. Section 17.7 treats the limit laws for *high risk scenarios* $Z^{E_n^c}$ where (E_n) is an increasing sequence of ellipsoids. As in the case of exceedances over horizontal thresholds we give a Representation and an Extension Theorem. For elliptic thresholds there are non-degenerate limit vectors which live on the boundary ∂B. Moreover the symmetry group of the limit measure ρ may split: $\gamma^t = \gamma_0^t \otimes \sigma^t$ on $\mathbb{R}^{d_0} \times \mathbb{R}^{d_1}$, where γ_0^t is an expansion group, and σ^t a one-parameter group of orthogonal transformations.

17.1 Scalar normalization. Scalar normalizations are widely applied. The limit expression for high risk scenarios has a simple form:

$$Z^r/r \Rightarrow W, \quad r \to \infty, \tag{17.1}$$

where Z^r denotes the vector Z conditional on $\|Z\| \geq r$.

Proposition 17.1. *If $Z \in \mathcal{D}^\infty(\rho)$ with scalar normalizations, then (17.1) holds, where W has distribution $d\rho_0 = 1_{B^c}d\rho/\rho(B^c)$.*

Many authors use the term multivariate regular variation to describe the tail behaviour of random vectors which satisfy (17.1). Mikosch [2005] gives a very readable account of the theory of excess measures with scalar normalizations.

The tail asymptotics of the vector Z in (17.1) are determined by the scalar function

$$p(r) = \mathbb{P}\{\|Z\| \geq r\}, \quad r > 0, \tag{17.2}$$

which varies regularly, and the spectral measure ρ^*. The symmetries γ^t, $t > 0$, of the excess measure ρ are scalar expansions. The tail exponent $\lambda > 0$ together with the spectral measure determines the excess measure

$$\rho(rB^c) = a/r^\lambda, \quad a = \rho(B^c).$$

The symmetries γ^t and their generator C have the form

$$\gamma^t(w) = e^{\tau t}w, \quad C = \tau I, \quad \tau = 1/\lambda > 0.$$

Heavy tails correspond to large τ. The tail function p in (17.2) varies regularly with exponent $-\lambda$.

Example 17.2. A vector Z with continuous density of the form

$$f(z) = f_r(\theta)L(r)/r^{d+\lambda}, \quad r = \|z\|, \ \theta = z/r \in \partial B,$$

where $L\colon [0,\infty) \to (0,\infty)$ varies slowly in infinity, and where f_r converges uniformly to a continuous function f_∞ on ∂B, will belong to $\mathcal{D}^\infty(\rho)$ with exponent λ and spectral density $\propto f_\infty$. ◊

Scalar normalization does not imply tails of the same weight in all directions. The vector Z may have non-negative components! If $Z \geq 0$, and if the marginal tails are comparable,

$$1 - F_i(t) \sim c_i L(t)/t^\lambda, \quad t \to \infty, \ c_i > 0, \ i = 1, \dots, d \tag{17.3}$$

for some slowly varying function L, then $Z \in \mathcal{D}^\infty(\rho)$ with scalar normalizations if and only if the *componentwise maxima* converge.

Let us now take a closer look at the *balance* condition (17.3). For scalar normalizations one may use (17.2) to write (16.1) as

$$(r^{-1}I)(\pi)/p(r) \to \rho \ \text{ weakly on } \varepsilon B^c,\ r \to \infty, \quad \varepsilon > 0. \tag{17.4}$$

The df F_d of the vertical component Z_d satisfies

$$\frac{1 - F_d(r)}{p(r)} \to \mathbb{P}\{W_d \geq 1\}, \quad r \to \infty. \tag{17.5}$$

More generally (17.4) implies

$$\frac{\mathbb{P}\{Z \in rE\}}{p(r)} \to \rho(E), \quad r \to \infty \tag{17.6}$$

for any Borel set E whose closure does not contain the origin, and which satisfies $\rho(\partial E) = 0$.

We shall apply this result to *positive-homogeneous* functions $s\colon \mathbb{R}^d \to \mathbb{R}$ of degree one. Such functions are linear on rays:

$$s(rw) = rs(w), \quad w \in \mathbb{R}^d,\ r \geq 0. \tag{17.7}$$

Example 17.3. The positive-homogeneous functions of degree one form a linear space. Examples are $w_1, w_1+\dots+w_d, \xi w, \max_i w_i, |w_1|-3w_2, \sqrt{w_1^2+\dots+w_d^2} = \|w\|_2$, $\|w\|_p$. The maximum of two positive-homogeneous functions is positive-homogeneous. In particular the positive part $s_+ = s \vee 0$ of a positive-homogeneous function is a non-negative positive-homogeneous function. ◊

For a convex open set U containing the origin the *gauge function* of U is the unique function $n_U\colon \mathbb{R}^d \to [0,\infty)$, which satisfies (17.7) and $\{n_U < 1\} = U$. These functions are more general than those defining the rotund-exponential densities in Section 9.2. The functions n_U are continuous and convex. The zero set of n_U is a closed convex cone, $\{n_U = 0\} = U^{++}$, see (5.9).

Theorem 17.4 (Balance). *Let Z have df π that satisfies (17.4). Let s and t be positive-homogeneous Borel function of degree one which are ρ-a.e. continuous. Suppose $\rho\{t \geq 1\}$ is positive and finite. Then*

$$\frac{\mathbb{P}\{s(Z) \geq r\}}{\mathbb{P}\{t(Z) \geq r\}} \to \frac{\rho\{s \geq 1\}}{\rho\{t \geq 1\}} \in [0,\infty], \quad r \to \infty. \tag{17.8}$$

In particular, if ρ charges the open upper halfspace $\{w_d > 0\}$, then $\rho\{w_d \geq 1\} = c \in (0,\infty)$, and

$$\frac{\mathbb{P}\{s(Z) \geq r\}}{1 - F_d(r)} \to \frac{\rho\{s \geq 1\}}{c} \in [0,\infty], \quad r \to \infty.$$

Proof. It suffices to prove the second relation. To this end set $E = \{s \geq 1\}$. Then $\partial E ß \{s = 1\} \cup D$ where D is the discontinuity set of s. (If w is a continuity point of s and $s(w) \neq 1$ there is a neighbourhood U of w and an $\varepsilon \in (0, 1)$ such that either $s < 1 - \varepsilon$ on U, or $s > 1 + \varepsilon$. Hence $w \notin \partial E$.) Now apply (17.6) and (17.5), and remark that $\mathbb{P}\{W_d \geq 1\} > 0$ holds if ρ charges the open halfspace $\{w_d > 0\}$. □

For a nice application to insurance see Wüthrich [2003].

Scalar expansions are geometric and do not depend on the coordinates. If $Z \in \mathcal{D}^\infty(\rho)$ with scalar normalizations then $A(Z) \in \mathcal{D}^\infty(A(\rho))$ for any linear transformation A. If one regards A as a change of coordinates then the excess measure remains the same, but the unit ball and hence the spectral measure will change in general. The norms with respect to the old and new unit ball are positive-homogeneous; the relation between the two limit relations is expressed in the Balance Theorem.

The simplicity of the asymptotic theory for heavy tailed distributions which allow scalar normalizations seems to be due to the fact that one may introduce polar coordinates, and write $Z = R\Theta$ with $R = \|Z\|$. We may and shall assume that Z does not charge the origin. The basic limit relation (17.4) is equivalent to asymptotic independence of the radial part and the angular part conditional on $R \geq r$ for $r \to \infty$.

Proposition 17.5. *Let $Z = R\Theta$. Suppose $p(r) = \mathbb{P}\{R \geq r\}$ varies regularly for $r \to \infty$ with exponent $-\lambda < 0$. Let μ_r be the distribution of the angular part Θ conditional on $R \geq r$. If $\mu_r \to \mu_\infty$ weakly on ∂B then (22) of the Preview holds where W has distribution $\mu_\infty(d\theta) \times \lambda dr/r^{\lambda+1}$ on B^c in terms of polar coordinates, and (17.4) holds with exponent λ and spectral measure $\rho^* = \mu_\infty$.*

Proof. We have to show that weak convergence $\mu_r \to \mu_\infty$ on ∂B implies convergence $Z^r/r \Rightarrow W$. Let μ_t^s denote the conditional distribution of Θ conditional on $s \leq R < t$. Convergence of the distribution of $(R/r, \Theta)$ conditional on $R \geq r$ to a product measure follows from the obvious equality

$$p(t)\mu_t = (p(t) - p(t + s))\mu_t^s + p(t + s)\mu_{t+s}, \quad s > 0$$

which shows that $\mu_t^s \to \mu_\infty$ for $t = cs \to \infty$ for any $c > 1$. The equivalence of asymptotic independence for the polar coordinates and convergence of high risk scenarios for complements of ellipsoids was shown in Theorem 16.15. □

In polar coordinates, *positive-homogeneous* functions have the form $s(r, \theta) = rs_1(\theta)$ for some function $s_1 : \partial B \to \mathbb{R}$. This allows a simple extension of the Balance Theorem:

Proposition 17.6. *If s and t are non-negative positive-homogeneous functions such that s_1 and t_1 are ρ^*-a.e. continuous on ∂B, and if $q > 0$ and $\mathbb{E}t^q(W) \in (0, \infty)$, where W has distribution $1_{B^c} d\rho/\rho(B^c)$, then*

$$\frac{\mathbb{E}s(Z)^q[\|Z\| \geq r]}{\mathbb{E}t(Z)^q[\|Z\| \geq r]} \to \frac{\mathbb{E}s^q(W)}{\mathbb{E}t^q(W)}, \quad r \to \infty.$$

The *Euclidean Pareto* excess measure ρ, with density $1/\|w\|^{d+\lambda}$ on $\mathbb{R}^d \setminus \{0\}$ is symmetric for scalar expansions. All halfspaces J supporting the unit ball have the same mass, $\rho(J) = C$, where $C = C(d, \lambda)$ is given in (12.2). Hence

$$\rho\{\theta \geq r\} = C/r^\lambda, \quad r > 0,\ \theta \in \partial B.$$

If λ is an integer the converse does not hold. Two excess measures which give the same weight to all halfspaces need not be equal. This result is due to *Kesten*. See Hult & Lindskog [2006a] for details. The example below is a variation of an example in Basrak, Davis & Mikosch [2002].

A probability measure on $\mathbb{R}^d$ is determined by its mass on halfspaces since this determines the multivariate characteristic function. One may restrict to halfspaces H which do not contain the origin. For excess measures this is not true. Let ρ_0 be an excess measure on the plane. Suppose $\gamma^t(\rho_0) = e^t \rho_0$ where $\gamma^t(w) = e^{\tau t} w$ for $\tau = 1/2m$ with m a positive integer. There are infinitely many excess measures ρ with these symmetries which give the same mass to all halfspaces:

$$\rho(H) = \rho_0(H), \quad H \in \mathcal{H}.$$

Example 17.7. Restrict attention to halfplanes $H(\theta, t) = \{w \in \mathbb{C} \mid \Re(e^{i\theta} w) \geq t\}$ with $t = 1$. Note that $\rho(H(\theta, t)) = \rho(H(\theta, 1))/t^{2m}$. We may assume that $\rho\{|w| \geq 1\} = 1$. The corresponding random vector is $W = R\omega$, where R and ω are independent, R has Pareto density $2m/r^{1+2m}$ on $[1, \infty)$, and ω lives on the unit circle in $\mathbb{C}$ with distribution ρ^*. This distribution is characterized by its Fourier coefficients $c_k = \mathbb{E}\omega^k$, $k \in \mathbb{Z}$. Obviously $c_0 = 1$ and $c_{-k} = \mathbb{E}\omega^{-k} = \mathbb{E}\bar{\omega}^k = \bar{c}_k$. We shall assume that $-\omega$ is distributed like ω. This implies that $(-1)^k c_k = \mathbb{E}(-\omega)^k = \mathbb{E}\omega^k = c_k$. The odd coefficients vanish. An example of ρ^* is given by the density

$$f(e^{i\varphi}) = 1 + a_2 \cos 2\varphi + b_2 \sin 2\varphi + a_4 \cos 4\varphi + \cdots \tag{17.9}$$

with a_{2n} and b_{2n} real, and $\sum |a_{2n}| + \sum |b_{2n}| \leq 1$ to ensure that f is non-negative. The Fourier coefficients are $c_{2n} = (a_{2n} + i b_{2n})/2$ for $n = 1, 2, \ldots.$

Let $H(\theta) = \{w \in \mathbb{C} \mid \Re(e^{i\theta} w) \geq 1\}$. Set $X = \Re(e^{i\theta}\omega)$. By independence

$$\rho(H(\theta)) = \mathbb{P}\{RX \geq 1\} = \mathbb{P}\{X \geq 1/R\} = 2m \int_1^\infty \mathbb{P}\{X \geq 1/r\} r^{-(1+2m)} dr.$$

Since $-X$ is distributed like X, and $|X| \leq 1$, we find, setting $s = 1/r$,

$$\rho(H(\theta)) = m \int_1^\infty \mathbb{P}\{|X| \geq 1/r\} r^{-(1+2m)} dr = \frac{1}{2} \int_0^1 \mathbb{P}\{X^{2m} \geq s\} ds = \frac{1}{2} \mathbb{E} X^{2m}.$$

Write $X = (e^{i\theta}\omega + e^{-i\theta}\bar{\omega})/2$. We see that $\mathbb{E}X^{2m}$ is a linear combination of terms $\mathbb{E}\omega^{2k}$ with $|k| \leq m$. We conclude that $\rho(H)$ for any halfplane $H = H(\theta, t)$

is determined by t, θ and the Fourier coefficients c_{2k}, $|k| \leq m$, of the spectral measure ρ^*. The coefficients a_n and b_n in (17.9) for $n > 2m$ do not affect $\rho(H)$. In particular the densities

$$g_0(w) = 1/|w|^{2+2m}, \quad g(w) = (1 + \sin(2 + 2m)\varphi)/w^{2+2m}, \quad w = re^{i\varphi}$$

give the same weight to all halfspaces. ◊

17.2 Scalar symmetries. If the excess measure ρ is symmetric for scalar expansions, the normalizations for vectors in the domain of attraction need not be scalar. The normalizations may be chosen to vary like a group of scalar expansions. For max-stable limit laws with identical heavy tailed marginals $G_i(t) = e^{-1/t^\lambda}$, $t > 0$, the normalizations are diagonal matrices $\alpha(t) = \mathrm{diag}(a_1(t), \ldots, a_d(t))$. The functions $a_i(t)$ vary regularly with the same exponent λ, but need not be asymptotically equal. In the geometric theory of excess measures the domains are larger. As a bonus of our analysis of these domains we shall also obtain the structure of the domains when γ^t, $t \in \mathbb{R}$, is a group of rotational expansions: $\gamma^t = e^{\tau t} R^t$ for a one-parameter group R^t in $\mathrm{O}(d)$. Our only condition is that the complex *Jordan form* of the generator is diagonal, and that the diagonal elements all have the same real part.

Excess measures symmetric for *scalar expansions* are determined by an exponent $\lambda = 1/\tau > 0$ which determines the rate of decay for the tails, and a spectral measure which determines the distribution of the excess measure over the directions. Let D be a bounded convex open set containing the origin. Then D is *adapted* and

$$\rho(rD^c) = c_D/r^\lambda, \quad r > 0, \ c_D = \rho(D^c).$$

The exponent does not depend on the set D.

Let α_n be linear transformations. Suppose $\alpha_n^{-1}\alpha_{n+1}(w) \to cw$ for a constant $c > 1$. The ellipsoids $E_n = \alpha_n^{-1}(B)$ allow one to visualize the linear expansions α_n. For simplicity take α_n such that

$$p_n = \mathbb{P}\{Z \in E_n^c\} \sim e^{-n}.$$

Asymptotically the ellipsoids grow by a fixed amount at each step

$$E_{n+1} \sim cE_n, \quad c = e^\tau.$$

The ellipsoids E_n actually only contain half the information of the sequence (α_n) about the asymptotic behaviour of π. Even if the ellipsoids are balls, $E_n = r_n B$, with $r_{n+1}/r_n \to c$, this does not mean that the α_n are scalar expansions.

Example 17.8. Let the excess measure ρ have density $h(r\theta) = h_0(\theta)/r^{d+\lambda}$ for a continuous function $h_0\colon \partial B \to (0, \infty)$. Assume $\rho(B^c) = 1$. The probability distribution π with density $f = h1_{B^c}$ lies in $\mathcal{D}^\infty(\rho)$. So does the distribution with the twisted density $\hat{f}(r\theta) = h_0(R_r\theta)/r^{d+\lambda}$ where $r \mapsto R_r \in \mathrm{O}(d)$ is a continuous curve of rotations which varies slowly:

$$R_{r_n}^{-1} R_{c_n r_n} \to I, \quad r_n \to \infty,\ c_n \to c,\ c \in (0, \infty).$$

A similar *twisting* is possible whenever $Z \in \mathcal{D}^\infty(\rho)$ may be normalized by scalars. Simply replace π_r by $R_r^{-1}(\pi_r)$, where π_r is the conditional distribution of Z given $\|Z\| = r$. ♢

The ellipsoids need not be balls.

Example 17.9. Let $E_n = \alpha_n(B)$ be centered ellipses in the plane. Assume $E_{n+1} \sim cE_n$ for $c > 1$. Write $E_n = c^n F_n$. Then $F_{n+1} \sim F_n$. The ellipses F_n may grow at a polynomial rate, and so may each of the axes, for different polynomials. Assume for concreteness that the major semi-axis has length n, and the minor length $1/\sqrt{n}$. The ellipses become very elongated, and their area grows like $\sqrt{n}$. Let φ_n denote the angle of the major axis of E_n with the horizontal. Since $F_{n+1} \sim F_n$ implies that the overlap of F_n and F_{n+1} is large, the difference $\varphi_{n+1} - \varphi_n$ has to be small. Indeed a simple computation shows that $F_{n+1} \sim F_n$ implies $\varphi_{n+1} - \varphi_n = o(1/n^{3/2})$. So the sequence φ_n has a limit, φ_∞. The ellipses become needles which converge to the line with slope φ_∞.

If the angles φ_n satisfies the stronger condition $\varphi_n - \varphi_\infty = o(1/n^{3/2})$, then one may rotate the ellipsoids E_n so that the major axes all lie along the line with direction φ_∞ without destroying the asymptotics.

Is it possible to describe the asymptotic behaviour of the ellipses F_n in general terms? The condition $F_{n+1} \sim F_n$ implies $|F_n| = r_n|B|$ where $r_{n+1}/r_n \to 1$. So scale the ellipses to have area π. Such an ellipse is determined by two numbers, the length $q \geq 1$ of the major semi-axis, and its angle $\varphi \in [0, \pi)$. Is it possible to represent such ellipses E, F by points $z_E, z_F \in \mathbb{R}^2$ so that the asymptotic relation $F_n \sim E_n$ translates into the simpler relation $d(z_{F_n}, z_{E_n}) \to 0$ in some appropriate metric? The representation $z_E = re^{i\theta} = (q-1)e^{2i\varphi} \in \mathbb{C}$ is continuous and has the desired property in terms of the Riemannian metric

$$ds^2 = \left(\frac{dr}{r+1}\right)^2 + r^2 d\theta^2. \tag{17.10}$$

Then $E_t = c^t r(t) E_t^*$, where $r(t+s)/r(t) \to 1$, and where the E_t^* are ellipses with area π corresponding to a continuous curve $z(t)$ in $\mathbb{C}$ with the property that it bogs down in the metric (17.10):

$$d(z(t_n + s_n), z(t_n)) \to 0, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}.$$

The ellipses E_t give one half of the picture. To see the linear expansions α_t we introduce the open square S of area two inscribed in the unit disk B. The parallelograms $S_n^* = c^{-n}\alpha_n(S)/r(n)$ inscribed in the ellipses E_n^* determine α_n. These parallelograms of area two move around inside E_n^*. We now have an extra degree of freedom for *twisting*. Even when the direction φ_n of the major axis of the ellipse E_n^* converges to a limit direction, the parallelograms inside the ellipses are allowed to keep on revolving, their form oscillating ever slower between a rectangle and a diamond. ◇

Suppose the excess measure ρ has density $h(r\theta) = h^*(\theta)/r^{d+\lambda}$ on $\mathbb{R}^d \setminus \{0\}$ where h^* is a continuous positive function on ∂B. The domain $\mathcal{D}^\infty(\rho)$ falls apart into subdomains. Every subdomain is determined by a sequence α_n such that $\alpha_n^{-1}\alpha_{n+1} \to 2I$, or rather by the equivalence class of such sequences (α_n) under the relation of asymptotic equality. The sequence (α_n) determines a sequence of ellipsoids $E_n = \alpha_n(B)$ such that $E_{n+1} \sim 2E_n$. The ellipsoids E_n may be normalized to have the volume of the unit ball

$$E_n = r_n E_n^*, \quad |E_n^*| = |B|.$$

Then $r_{n+1}/r_n \to 2$ and E_n^* are ellipsoids with constant volume which change shape slowly

$$|E_{n+1}^* \setminus E_n^*| \to 0.$$

In addition to the sequence (r_n) and the sequence (E_n^*) there is a sequence of transformations $S_n \in \mathrm{O}(d)$ which map the ball B onto itself. These transformations tell us how the spectral density h^* fits on ∂E_n, or on ∂E_n^*. The ellipsoids E_n^* may rotate clockwise and at the same time the spectral density on ∂E_n^* may rotate anticlockwise. The linear normalization α_n determines E_n, but it does not determine S_n. To understand the rotations S_n we compare two sequences α_n and $\beta_n = \alpha_n S_n$ which yield the same ellipsoids E_n. Let f be the function associated with α_n as described above: $f(\alpha_n(\theta)) = h^*(\theta)$ for $\theta \in \partial B$. Define g similarly on ∂E_n in terms of β_n. Then on the elliptic surface g may be derived from f by a rotation S_n (in terms of coordinates in which E_n is a ball). Thus for excess measures ρ whose symmetries are scalar expansions, and which have continuous positive densities, the domain $\mathcal{D}^\infty(\rho)$ is determined by:

the exponent $\tau = 1/\lambda > 0$;
the spectral density $h^*\colon \partial B \to [0, \infty)$, the restriction of the density of ρ to ∂B;
a sequence of ellipsoids E_n^* of volume $|B|$, such that $E_{n+1}^* \sim E_n^*$;
a sequence of positive numbers r_n such that $r_{n+1}/r_n \to 2$;
a sequence $S_n \in \mathrm{O}(d)$ such that $S_{n+1} - S_n \to 0$;
a roughening of Lebesgue measure.

The *roughening of Lebesgue measure* depends on the geometry induced by the ellipsoids E_n, as described in Section 16.7; the first five objects in the list may be chosen freely. There is no natural construction of α_n in terms of $E_n = r_n E_n^*$ and S_n.

The measure ρ may have extra *symmetries*. Let $\mathcal{G}$ be the symmetry group of ρ. Every $\tau \in \mathcal{G}$ has the form $\tau = \sigma\gamma^t$ since $\tau(\rho) = e^t\rho$ for some $t \in \mathbb{R}$. Let $\mathcal{G}^*$ be the subgroup of symmetries σ which preserve mass, $\sigma(\rho) = \rho$. This is a normal subgroup, and is compact. (Else there is a sequence of ellipsoids $\sigma_n(B)$ which diverges to a degenerate ellipsoid contained in a hyperplane $\{\xi = 0\}$, and the complement has finite mass $\rho(B^c)$. Contradiction.) By Theorem 18.74 $\mathcal{G}^*$ is a subgroup of $\mathrm{O}(d)$ in suitable coordinates. The unit ball is adapted in any coordinates since the symmetries are scalar. The spectral density h^* is invariant under $\sigma \in \mathcal{G}^*$. So if the rotations S_n above lie in $\mathcal{G}$ they have no effect. We shall not pursue this matter here. The extreme case $\mathcal{G}^* = \mathrm{O}(d)$ was treated in the Preview and in Section 12.1.

Now consider rotational expansions. The complex Jordan form of the generator is diagonal, and the real parts of the diagonal elements are equal. All eigenvalues of γ lie on the same circle in $\mathbb{C}$. We now have a restriction on the sequence of rotations S_n in the list above. Suppose $d = 2m$ and $\gamma^t(w) = e^{t(\tau+i\varphi)}w$ for $w \in \mathbb{C}^m$ where we identify $\mathbb{R}^2$ with $\mathbb{C}$. The condition (16.25) implies that the limit points T of the sequence (S_n) above should lie in $\mathrm{U}(m)$, the group of unitary matrices in $\mathrm{GL}(\mathbb{C}, m)$. The converse also holds by a Lie algebra argument. In general the limit points lie in a subgroup of $\mathrm{O}(d)$ of the form

$$\mathrm{U}(m_1) \times \cdots \times \mathrm{U}(m_q) \times \mathrm{O}(d_0), \quad d_0 + 2m_1 + \cdots + 2m_q = d.$$

So far we have not imposed any regularity conditions on the excess measure ρ except that it does not live on a linear hyperplane. Now assume more. Suppose the convex support of ρ is the whole space. Then for every distribution F in its domain of attraction the marginals $F_1, \dots, F_d$ have upper and lower tails which vary regularly with the same exponent. In general the direction of the halfspace $H_t = \alpha_t\{\eta \ge 1\}$ will wander all over the unit sphere. With these halfspaces H_t we may associate the high risk scenarios Z^{H_t}. It is this family of high risk scenarios, normalized by α_t^{-1} which converges in distribution to the limit law $d\rho^J = 1_J d\rho/\rho(J)$ associated with the horizontal halfspace $J = \{\eta \ge 1\}$.

Theorem 17.10. *Suppose $Z \in \mathcal{D}^\infty(\rho)$ for an excess measure ρ with a symmetry group of scalar expansions with exponent $\tau > 0$. Assume $\rho\{\theta \ge 1\} > 0$ for all unit functionals θ on $\mathbb{R}^d$. Let F be the df of the random variable $Y = \eta(Z)$ where $\eta \ne 0$ is a linear functional. Then $1 - F(r)$ varies regularly with exponent $-1/\tau$ for $r \to \infty$.*

Proof. Assume $\rho(B^c) = 1$, and let $E_t = \beta(t)(B)$. Then (11) in the Preview gives $p(t) = \pi(E_t^c) \sim e^{-t}$ for $t \to \infty$, where π is the distribution of Z, and $E_{t+s} \sim e^{\lambda s}E_t$ by (12). For $r > 1$ let $E(r) = E_{t(r)}$ be the ellipsoid of the continuous increasing family $E_t, t \ge 0$, supporting the halfspace $H_r = \{\eta \ge r\}$. Then $t(e^c r) - t(r) = c/\lambda + o(1)$ for $r \to \infty$. Write $H_r = A_{t(r)}J_r$ where the halfspace J_r supports the unit ball in the point $w(r)$. Then $J_{e^c r}$ supports the unit ball in $w(e^c r)$, and

$\|w(e^{c_n}r_n) - w(r_n)\| \to 0$ for $r_n \to \infty$ and $c_n \to c$ since $e^{-c_n/\lambda} A^{-1}_{t(r_n)} A_{t(e^{c_n}r_n)} \to I$. By weak convergence and the condition on the convex support $\pi(H_r) \sim \rho(J_r)p(t(r))$ for $r \to \infty$. The conditions on ρ ensure that $w \mapsto \log \rho(J_w)$ (with J_w the halfspace supporting B in the point $w \in \partial B$) is continuous on ∂B. Hence

$$\pi(H_{e^c r})/\pi(H_r) \sim p(t(e^c r))/p(t(r)) \to e^{-\lambda c}, \quad r \to \infty.$$

This is what we claimed. □

Similarly one proves that $R(y) = \mathbb{P}\{|Y| \geq y\}$ varies regularly in infinity if ρ is full, see MS Theorem 6.1.31, or for instance that $\mathbb{P}(\{Y \leq -2y\} \cup \{y < Y < 3y\})$ varies regularly. However the extra conditions in the theorem do not ensure that F is *balanced*. The quotient $(1 - F(y))/R(y)$ need not converge for $y \to \infty$.

Example 17.11. Suppose $d = 2$, and identify $\mathbb{R}^2$ with $\mathbb{C}$. Let ρ have density g where $g(re^{is}) = g_0(e^{is})/r^3$ for some continuous positive function g_0 on the unit circle in $\mathbb{C}$, such that $p := \rho\{v \geq 1\}/\rho\{|v| \geq 1\} > 1/2$. Let $Z = (X, Y)$ have density $f(re^{is}) = g_0(e^{i(s+L(r))})/r^3$ on $r > r_0$ for a continuous slowly varying unbounded function L. Then $\mathbb{P}\{Y \geq y\} \sim \rho(H_{L(y)})/y$ for $y \to \infty$, where H_s is the halfplane which supports the unit disk in the point e^{is}. Let Y have df F, and let $R(y) = F(-y) + 1 - F(y)$. Then $(1 - F(y_n))/R(y_n)$ tends to p for $y_n \to \infty$ if $L(y_n) \equiv 0 \mod 2\pi$, and to $1 - p$ if $L(y_n) \equiv \pi \mod 2\pi$. ◊

Proposition 17.12. *Let $Z \in \mathcal{D}^\infty(\rho)$, where ρ has scalar symmetries $\gamma^t(w) = e^{\tau t}w$ for some $\tau > 0$. Suppose $\varphi\colon \mathbb{R} \to [0, \infty)$ is bounded, vanishes on a neighbourhood of infinity, and is continuous outside a closed null set. Assume φ does not vanish a.e. on $(0, \infty)$. If φ vanishes a.e. on $(-\infty, 0)$ assume that ρ does not vanish on the halfspace $\{\theta \geq 1\}$ for any unit functional θ, otherwise assume ρ does not live on a hyperplane through the origin. Then the function*

$$\bar{\varphi}\colon (\theta, r) \mapsto \mathbb{E}\varphi(\theta(Z)/r), \quad \|\theta\| = 1, \ r \geq 1$$

varies regularly for $r \to \infty$ with exponent $-1/\tau$ uniformly in $\theta \in \partial B$.

Proof. Let $\theta_n \in \partial B$, $t_n \to \infty$, $c_n \to c \geq 1$. Then

$$\bar{\varphi}(\theta_n, c_n e^{t_n})/\bar{\varphi}(\theta_n, e^{t_n}) \to 1/c^{1/\tau}$$

since $\theta(Z)/ce^t = \theta'(A(t')Z)/c$ for suitable $\theta'_n \in \partial B$, and $t'_n \geq 0$. One may write the quotient as

$$\int \varphi(\theta'_n(w)/c_n)d\rho_n(w) \Big/ \int \varphi(\theta'_n(w))d\rho_n(w), \quad \rho_n = A(t'_n)^{-1}(\pi) \to \rho.$$

Take a convergent subsequence $\theta'_n \to \theta' \in \partial B$, say. The continuity conditions ensure that $\varphi(\theta'_n(w_n)/c_n) \to \varphi(\theta'(w)/c)$ for $w_n \to w \neq 0$ by the Continuity Theorem, Theorem 4.13, since the univariate marginals of ρ have GPDs of the same type, with continuous density. □

If one restricts the normalizations to be scalar (or diagonal) one obtains stronger projection theorems. For diagonal normalizations coordinate projections on *coordinate* subspaces preserve convergence – as is obvious by ignoring the remaining coordinates. This is the situation for coordinatewise extremes. As a consequence, for scalar normalizations all projections preserve convergence.

Proposition 17.13. *If* $Z \in \mathcal{D}^\infty(\rho)$ *with scalar normalizations then* $A(Z) \in \mathcal{D}^\infty(A\rho)$ *with scalar normalizations for any linear map* A *from* $\mathbb{R}^d$ *onto* $\mathbb{R}^m$*,* $1 \le m \le d$*.*

17.3* Coordinate boxes. A sample cloud of n independent observations $Z_1, \dots, Z_n$ from a distribution on $\mathbb{R}^d$ determines a coordinate box. Define

$$X_n = Z_1 \vee \cdots \vee Z_n, \quad Y_n = (-Z_1) \vee \cdots \vee (-Z_n). \tag{17.11}$$

There is a minimal box B_n that contains the sample:

$$B_n = [-Y_n^{(1)}, X_n^{(1)}] \times \cdots \times [-Y_n^{(d)}, X_n^{(d)}] = [-Y_n, X_n]\text{ß}\mathbb{R}^d \tag{17.12}$$

For dimension $d > 1$ this box is not the convex hull of the sample cloud. The vectors X_n and $-Y_n$ are *coordinatewise extremes*. However, if the boxes B_n, suitably normalized, converge in distribution to a non-degenerate limit box then the convex hulls of the normalized sample clouds converge, and moreover the distribution of the limit box determines the distribution of the limit convex hull.

We first look at the univariate situation.

Proposition 17.14. *Let* $Z_1, Z_2, \dots$ *be independent observations from the distribution* π *on* $\mathbb{R}$ *with df* F*. Define* X_n *and* Y_n *by* (17.11)*. Suppose* $R(t) = F(-t) + 1 - F(t)$ *is positive for* $t > 0$ *and varies regularly in* ∞ *with exponent* $-\lambda < 0$*. If*

$$F(-t)/R(t) \to p \in [0,1], \quad t \to \infty, \tag{17.13}$$

then

$$(r^{-1}I)(\pi)/R(r) \to \rho \ \text{ weakly on } (-\varepsilon, \varepsilon)^c, \quad \varepsilon > 0,\ r \to \infty$$

where ρ *is a Radon measure on* $\mathbb{R} \setminus \{0\}$*, and*

$$\rho(-\infty, -r] = p/r^\lambda, \quad \rho[r, \infty) = (1-p)/r^\lambda, \quad r > 0.$$

Let $R(c_n) \sim 1/n$*. Then*

$$(X_n, Y_n)/c_n \Rightarrow (U, V)$$

where $U = (1-p)^{1/\lambda} U_0$ *and* $V = p^{1/\lambda} V_0$ *are independent and* U_0 *and* V_0 *have df* $G(t) = e^{-1/t^\lambda}$ *on* $(0, \infty)$*. If* (17.13) *does not hold there are no positive affine transformations* α_n *such that* $(\alpha_n^{-1}(X_n), \alpha_n^{-1}(Y_n))$ *converges in distribution to a non-constant limit.*

Proof. The first part follows from regular variation and (17.13) as in Section 6. Let π_r be the distribution of Z conditional on $|Z| \geq r$. Convergence $\pi_r/R(r) \to \rho$ implies convergence of the normalized sample clouds, and their convex hulls $[-Y_n, X_n]/c_n$ to the convex hull $[-V, U]$ of the Poisson point process with mean measure ρ. This gives the distribution of (U, V). Conversely if there exist positive affine α_n yielding a non-constant limit $(\bar{U}, \bar{V})$ then

$$\bar{U} = \beta((1-p)^{1/\lambda} U_0), \quad \bar{V} = \beta(p^{1/\lambda} V_0).$$

This determines p. Hence in (17.13) there is a unique limit point. □

Theorem 17.15. *Let $B_n = [-Y_n, X_n]$ be the coordinate box determined by the first n points of a sequence of independent observations $Z_1, Z_2, \dots$ from the df F with marginals F_i, $i = 1, \dots, d$. Suppose $R_i(t) = F_i(-t) + 1 - F_i(t)$ varies regularly with exponent $-\lambda_i < 0$, and the balance conditions hold:*

$$F_i(-t)/R_i(t) \to p_i \in [0,1], \quad t \to \infty,\ i = 1, \dots, d. \tag{17.14}$$

Let $\gamma_n = \operatorname{diag}(c_{n1}, \dots, c_{nd})$ where $R_i(c_{ni}) \sim 1/n$. Suppose

$$(\gamma_n^{-1}(X_n), \gamma_n^{-1}(Y_n)) \Rightarrow (U, V) \in \mathbb{R}^{2d}.$$

Then the normalized sample clouds converge:

$$N_n = \{\gamma_n(Z_1), \dots, \gamma_n(Z_n)\} \Rightarrow N_0 \ \textit{weakly on } \mathbb{R}^d \setminus \varepsilon B, \quad \varepsilon > 0,$$

where N_0 is a Poisson point process on $\mathbb{R}^d \setminus \{0\}$. Its mean measure ρ is symmetric for the group of expansions Q^t with generator $\operatorname{diag}(\tau_1, \dots, \tau_d)$, $\tau_i = 1/\lambda_i$.

Proof. It helps to introduce the maps $z \mapsto z^+$ and J_δ on $\mathbb{R}^d$:

$$z^+ = (z_1 \vee 0, \dots, z_d \vee 0), \quad J_\delta = \operatorname{diag}(\delta_1, \dots, \delta_d), \quad \delta \in \{-1, 1\}^d.$$

We may assume $0 \in B_n$ replacing X_n by X_n^+ and Y_n by Y_n^+, or by adding the origin to the sample cloud. The normalized vertices of the box B_n may be written as

$$W_n^\delta = J_\delta(\overline{W}_n^\delta), \quad \overline{W}_n^\delta = \max\{J_\delta(\gamma_n^{-1}(Z_1^+)), \dots, J_\delta(\gamma_n^{-1}(Z_n^+))\}.$$

Let W^δ denote the vertices of the limit box $[-V, U]$. Then

$$(W_n^\delta, \delta \in \{-1,1\}^d) \Rightarrow (W^\delta, \delta \in \{-1,1\}^d) \in \mathbb{R}^{d+2^d}.$$

Convergence $W_n^\delta \Rightarrow W^\delta$ makes W^δ max-stable and gives $n\gamma_n^{-1}(\pi_\delta) \to \rho_\delta$ weakly on $[0,\infty)^d \setminus [0, w]$ for $w \in (0,\infty)^d$ by Theorem 7.3 where $\pi_d = (J_\delta \pi)^+$ and the ρ_δ are Radon measures on $[0,\infty)^d \setminus \{0\}$. Finally one has to check that the family ρ_δ, $\delta \in \{-1,1\}^d$, determines a unique Radon measure ρ on $\mathbb{R}^d \setminus \{0\}$ which is finite on

εB^c for $\varepsilon > 0$ such that $(J_\delta \rho)^+ = \rho_\delta$, and that the weak convergence above implies $n\gamma_n^{-1}(\pi) \to \rho$ weakly on εB^c for $\varepsilon > 0$. This is simple algebra. Observe that

$$\tilde{\rho} = \sum_{\delta} \rho_\delta = \sum_{K\,\text{ß}\,D} \rho^K, \quad \rho^K = p_K(\rho)$$

where $D = \{1, \dots, d\}$ and $p_K \colon \mathbb{R}^D \to \mathbb{R}^K$ the natural coordinate projection. Applying p_K to both sides gives $\tilde{\rho}^K = 2^{|D \setminus K|} \sum_{I\,\text{ß}\,K} \rho^I$. Recursively express ρ^K as a linear combination of $\tilde{\rho}^I$ with $I\,\text{ß}\,K$. □

The *balance* condition (17.14) may be dropped if one allows non-linear normalizations of the form $\varphi = \varphi_1 \otimes \cdots \otimes \varphi_d$ where the $\varphi_i \in \mathcal{M}^\uparrow$ are linear and strictly increasing on the two half axes but the slopes may differ.

17.4 Heavy and heavier tails. If the γ^t, $t > 0$, are expansions by diagonal matrices $\gamma^t = \operatorname{diag}(e^{a_1 t}, \dots, e^{a_d t})$ with $0 < a_1 < \cdots < a_d$ the tails in different directions have different weights. The group γ^t imposes natural coordinates in the space where the excess measure ρ lives. This also holds for vectors in the domain $\mathcal{D}^\infty(\rho)$ of an excess measure with these symmetries, but only to a certain extent. By the *Spectral Decomposition Theorem* there exist coordinates in which the normalizations may be chosen to be diagonal.

Example 17.16. Consider a unimodal probability density f on $\mathbb{R}^2$ of the form

$$f(x, y) = f_1(|x|) \wedge f_2(|y|)$$

where f_1 and f_2 are continuous strictly decreasing functions on $[0, \infty)$ and $f_2(t) << f_1(t)$ for $t \to \infty$. The level sets $\{f \geq c\}$ are closed rectangles for $0 < c < c_0$. The vertical height is negligible compared to the horizontal width of the rectangle for $c \to 0$. Hence the behaviour of the density along rays is determined by the lighter tail, except for horizontal rays. ♢

For exceedances over linear thresholds it is the other way round. There the heavier tails dominate the scene.

Proposition 17.17. *Let $Z = (X, Y) \in \mathbb{R}^{h+1}$. Suppose the upper tail of Y varies regularly with exponent $-\lambda < 0$, and is heavier than the tail of $\|X\|$,*

$$\mathbb{P}\{\|X\| \geq t\} / \mathbb{P}\{Y \geq t\} \to 0, \quad t \to \infty.$$

Then for any $c \in \mathbb{R}^h$

$$\mathbb{P}\{Y + c^T X \geq t\} \sim \mathbb{P}\{Y \geq t\}, \quad t \to \infty.$$

Proof. Assume $c \neq 0$. Let Z have distribution π. Let $y_n \to \infty$. There exist $\delta_n \to 0+$ such that

$$\mathbb{P}\{\|X\| \geq \delta_n y_n\}/\mathbb{P}\{Y \geq 2y_n\} \to 0,$$

since this holds for fixed $\delta > 0$. Define the open vertical cylinder $C_n = \{\|x\| < \delta_n y_n\}$. The two sets

$$\{y \geq y_n + \delta_n y_n\} \cap C_n \text{ß} \{y \geq y_n - \delta_n y_n\} \cap C_n$$

enclose the sets $\{y \geq y_n\} \cap C_n$ and eventually $\{y + c^T x \geq y_n\} \cap C_n$. The relations

$$\mathbb{P}\{Y \geq y_n + \delta_n y_n\} - \pi(C_n) \sim \mathbb{P}\{Y \geq y_n + \delta_n y_n\} \sim \mathbb{P}\{Y \geq y_n - \delta_n y_n\}$$

imply that the probability of the extreme sets above are asymptotically equal. Since we may neglect $\pi(C_n)$ this gives the desired asymptotic equality. □

If successive components of Z have heavier tails this does not make the coordinates natural. To each coordinate we may add a linear combination of the preceding coordinates without altering the tail behaviour.

Even if the tails of the components of Z vary regularly with distinct exponents there are no canonical coordinates. There is a canonical sequence of $d-1$ projections onto spaces of decreasing dimension, such that each projection produces lighter tails.

Theorem 17.18. *Suppose $Z = (Z_1, \dots, Z_d)$ and $R_i(t) = \mathbb{P}\{|Z_i| \geq t\}$ varies regularly with exponent $-\lambda_i < 0$. If*

$$R_{i-1}(t)/R_i(t) \to 0, \quad t \to \infty,\ i = 2, \dots, d$$

then $\lambda_1 \geq \cdots \geq \lambda_d > 0$. Let $\xi_1, \dots, \xi_d$ be independent linear functionals. Set $X = (X_1, \dots, X_d)$ with $X_i = \xi_i(Z)$, and $S_i(t) = \mathbb{P}\{|X_i| \geq t\}$. Suppose

$$S_{i-1}(t)/S_i(t) \to 0, \quad t \to \infty,\ i = 2, \dots, d.$$

Then $S_i(t)/R_i(t) \to c_i \in (0, \infty)$ for $t \to \infty$ and

$$X_i = c_{i1} Z_1 + \cdots + c_{ii} Z_i, \quad i = 1, \dots, d$$

with $|c_{ii}|^{\lambda_i} = c_i$ for $i = 1, \dots, d$.

Proof. By induction. □

In the geometric approach nearly all linear functionals have the tail behaviour of the heaviest component by Proposition 17.17; in the coordinate based approach it is special to have d variates with heavy tails which are comparable. Let us say a few words about this disparity. Even if the coordinates measure incomparable quantities like returns and changes in volatility, the assumption in looking at the bivariate distribution, is that there is an organic bond. Both variates form part of a description of some complex underlying model. If one believes that the symmetry visible in the sample cloud is spurious, then it can hardly be used to determine high risk scenarios in halfspaces which contain only a few or no sample points, and statements about risks and losses in terms of unbounded loss functions become dubious. On the other hand if the observed symmetry fits in the model of a limiting excess measure the dependency might be said to reveal some hidden structure of the underlying system which may persist over regions which are of interest to us. Persistence is more likely in the algebraic structure than in the spectral measure; it is more likely for simple *scalar expansion* groups and uniform spectral measures than for groups with complex generators and asymmetric distributions on the sphere. In cases where symmetry is present a geometric approach seems to be appropriate. The ellipsoids associated with the normalization of the sample cloud are geometric objects. These ellipsoids suggest a more gentle change in the rate of decrease of the tail as a function of the direction than what is prescribed by the asymptotics for exceedances and densities, where almost all tails have comparable rates of decrease.

For non-negative vectors in $\mathcal{D}^{\vee}(\rho)$ the ellipsoids are coordinate ellipsoids, and the minimal axes are coordinate axes. If the d coordinate tails are comparable, then so are the tails of all positive linear functionals.

Divergence of the shape of a sequence of ellipsoids does not imply that the direction of the major axes converges, as we know from Example 17.9. It is also possible that the ellipsoids are coordinate ellipsoids, that the axes are not comparable, that the tails of the marginals vary regularly with the same exponent, and that $Z \in \mathcal{D}^{\infty}(\rho)$ for an excess measure ρ whose symmetries are scalar expansions, but that normalization by diagonal matrices is not possible.

Example 17.19. Let f_0 have a continuous unimodal density with elliptic level sets

$$\{f_0 > e^{-3t}\} = E_t = \alpha_t(B) = \{x^2 + y^2/t^2 < e^{2t}\}, \quad t \geq t_0.$$

Then f_0 lies in the domain of the excess measure with density $1/r^3$. The vertical tail of f_0 is heavier than the horizontal tail, and the normalizations α_t are diagonal. Replace the excess measure by an excess measure with density $h(r\theta) = h_0(\theta)/r^3$ where h_0 is continuous and positive, but not constant. Replace α_t by $\beta_t = \alpha_t S_t$ where S_t is a rotation over $\varphi(t) = \sqrt{t}$. The tails of the corresponding density f are comparable to those of f_0 since the geometry of the ellipses E_t is not altered, but the normalizations β_t can not be diagonalized. For any unit vector w_0 the curve $\beta_t(w_0)$ spirals out to infinity. See Section 16.8. ◊

By the *Spectral Decomposition Theorem* such spiralling behaviour is not allowed for distributions in the domain of excess measures with diagonal symmetries, $\gamma^t =$

$\operatorname{diag}(r_1^t, \dots, r_d^t)$ with $r_1 < \cdots < r_d$. At this point the reader is advised to have a look at Example 18.13 in the section on the SDT.

17.5* Maximal symmetry. For *scalar expansion groups* $\gamma^t(w) = e^{\tau t} w$ on $\mathbb{R}^d$, there is a unique excess measure ρ with density g which is 1 on the unit sphere ∂B:

$$g(w) = 1/\|w\|^{d+\lambda}, \quad w \neq 0,\ \lambda = 1/\tau.$$

In this section we study excess measures ρ for linear expansions, which have maximal symmetry. In first instance we look at excess measures with continuous *unimodal* densities on $\mathbb{R}^d \setminus \{0\}$, with elliptic *level sets*.

If the density g is one on the unit sphere, it is continuous, unimodal, and has elliptic level sets, since $g(\gamma^t(w)) = 1/q^t$ for $w \in \partial B$ with $q = e \det \gamma$ by (14) in the Preview. Suppose $g = c_1 < 1$ on ∂E_1 for some ellipsoid E_1. This information does not determine g as we have seen in Section 16.8. However if the generator of the symmetry group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, is symmetric, or if γ^t is symmetric for all $t \in \mathbb{R}$ (or for a sequence $t_n \to 0$, $t_n \neq 0$, or for two values t_1, t_2 which are rationally independent), then one may apply an orthogonal coordinate transformation to bring C and the linear transformations γ^t into diagonal form. The new coordinate planes are symmetry planes of all the ellipsoids E_t. The family of ellipsoids E_t, and hence g, is uniquely determined by E_1, and the condition that the symmetry planes of E_1 are symmetry planes of E_t for every $t \in \mathbb{R}$.

Questions which interest us are: How symmetric is the excess measure ρ with density g? Does there exist a simple analytic formula for the density? Do there exist *typical densities* f, unimodal, with elliptic level sets, which are invariant under the measure preserving symmetries of ρ?

Before looking into these questions let us discuss the link with *linear expansion groups* $\gamma^t = e^{tC}$ whose generators have complex diagonal *Jordan form*. The diagonal entries of the diagonal form of C lie in $\Re > 0$, and every non-real diagonal element $\tau + \lambda i$ is matched by an element $\tau - \lambda i$. The real Jordan form of C is a blocked diagonal matrix, with blocks of size two representing the entries $\tau + \lambda i$ with $\lambda > 0$, as in (18.25), and blocks of size one containing the real entries. Choose a basis $e_1, \dots, e_d$ on which C has this blocked form. This basis is not unique; it may be replaced by $f_k = t_k e_k$, with $t_k \neq 0$, provided that $t_{k+1} = t_k$, if the two vectors e_k, e_{k+1} correspond to a pair of conjugate non-real complex eigenvectors, a block of size two. The basis determines an inner product, and hence a unit ball B. Set $E_t = \gamma^t(B)$. The one-parameter group of matrices $\bar{\gamma}^t = e^{t\bar{C}} = \operatorname{diag}(e^{\tau_1 t}, \dots, e^{\tau_d t})$, where $\tau_k \in \mathbb{R}$ are the diagonal elements of the real Jordan form of C, generates the same ellipsoids:

$$\bar{\gamma}^t(B) = E_t = \gamma^t(B), \quad t \in \mathbb{R},$$

and hence the same density g, if we specify $g = 1$ on ∂B.

There are many ways leading to excess measures whose densities have elliptic level sets.

For diagonal expansions $\gamma^t = e^{tC}$, $C = \mathrm{diag}(\tau_1, \dots, \tau_d)$, with $0 < \tau_1 \le \cdots \le \tau_d$, let g be the density of the corresponding excess measure ρ such that $g \equiv 1$ on ∂B. The level sets are coordinate ellipsoids. This density depends on the Jordan basis. Assume $\tau_d > \tau_1$. Then, by (16.8), the spectral measure ρ^* on ∂B is not a multiple of the *uniform distribution*. Besides, there is no simple formula for this density g. For each $w \neq 0$ there exists $t = t(w)$ such that $\gamma^{-t(w)}(w) \in \partial B$. One has to solve

$$e^{-2\tau_1 t} w_1^2 + \cdots + e^{-2\tau_d t} w_d^2 = 1. \tag{17.15}$$

The right hand side is strictly decreasing in t. So there is a unique solution $t = t(w)$ to this simple equation. Only in very special situations the solution is an explicit analytic expression of w. The density g has the form

$$g(w) = 1/q^{t(w)}, \quad q = e^{1+\tau_1+\cdots+\tau_d}.$$

(If $\tau_d = \tau_1$ then $\|\gamma^{-t}(w)\| = e^{-\tau t}\|w\|$ and hence $t(w) = \log\|w\|/\tau$ and $q^{t(w)} = \|w\|^{d+1/\tau}$.) The density g is unimodal with the elliptic level sets:

$$\{g > 1/q^c\} = \{t(w) < c\} = \left\{\frac{w_1^2}{e^{2c\tau_1}} + \cdots + \frac{w_d^2}{e^{2c\tau_d}} < 1\right\}, \quad c \in \mathbb{R}.$$

The symmetry of the measure ρ is meager. If $\tau_1 < \cdots < \tau_d$, then the group of measure preserving symmetries is discrete: it consists of the 2^d linear maps which change the sign of some components.

Proposition 17.20. *Let $\sigma_1 < \cdots < \sigma_m$ be the distinct entries of $(\tau_1, \dots, \tau_d)$ with multiplicities $d_1, \dots, d_m$ with sum d. Suppose ρ is the excess measure with density g on $\mathbb{R}^d \setminus \{0\}$ as described above. An affine transformation α such that $\alpha(\rho) = \rho$ has the form*

$$\alpha(w) = (\alpha_1(w_1), \dots, \alpha_m(w_m)), \quad w = (w_1, \dots, w_m) \in \mathbb{R}^{d_1+\cdots+d_m}$$

with $\alpha_i \in \mathrm{O}(d_i)$, $i = 1, \dots, m$.

Proof. The affine transformation α preserves the level sets $\{g > c\}$. So $\alpha(B) = B$ for the unit ball $B = \{g > 1\}$. Thus α is orthogonal. Now consider the ellipsoid $E = \{g > 1/2\}$. Then $\alpha(E) = E$ and hence the half-axes of $\alpha(E)$ have the same size τ_k, as those of E. It follows that α preserves the m d_k-dimensional subspaces associated with $\sigma_1, \dots, \sigma_m$, and is orthogonal on each of these subspaces. □

Let us now describe all such maximally symmetric excess measures associated with the generator $C = \mathrm{diag}(\tau_1, \dots, \tau_d)$. The symmetry group $\mathcal{G}$ is $\mathcal{S} \times \mathcal{H}$ where $\mathcal{S} = \mathrm{O}(d_1) \times \cdots \times \mathrm{O}(d_m)$ and $\mathcal{H}$ is the one-dimensional group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$.

Let ρ_0 be a Radon measure on $\mathcal{X} = [0,\infty)^m \setminus \{0\}$ such that $\beta^t(\rho_0) = e^t \rho_0$, $t \in \mathbb{R}$, where $\beta^t = e^{tS}$ for $S = \mathrm{diag}(\sigma_1, \dots, \sigma_m)$. Let $\mu = \mu_1 \times \cdots \times \mu_m$ be the uniform distribution on $\mathcal{T} = \partial B_1 \times \cdots \times \partial B_m$, where B_k is the unit ball in $\mathbb{R}^{d_k}$. The space $\mathcal{T}$ is a $(d-m)$-dimensional compact manifold. The extremely symmetric excess measure ρ in the proposition above has the form $\rho = \Phi(\mu \times \rho_0)$ where $\Phi\colon \mathcal{T} \times \mathcal{X} \to \mathbb{R}^d \setminus \{0\}$ is defined by

$$\Phi(\omega, x) = (x_1\omega_1, \dots, x_m\omega_m) \in \mathbb{R}^{d_1+\cdots+d_m}.$$

Let D_p be the unit ball of $\mathbb{R}^m$ with the l^p-norm for a $p \in [1,\infty]$. One may describe the measure ρ_0 on $\mathcal{X}$ in terms of a spectral measure ρ_p^* on $[0,\infty)^m \cap \partial D_p$ as the image of the product measure $d\rho_p^* \times e^t dt$ on $([0,\infty)^m \cap \partial D_p) \times \mathbb{R}$. If ρ_p^* is concentrated in one point $\xi \in [0,\infty)^m \cap \partial D_p$ then ρ is an elementary excess measure concentrated on the $(d+1-m)$-dimensional orbit $\mathcal{G}\xi$. Only for $p = \infty$ there exists an explicit expression for the density g. The unit ball D_∞ is the cube $(-1,1)^m$. As above define t so that $\gamma^{-t}(x) \in \partial D_\infty$ for $x \in \mathcal{X}$:

$$t = \max_{1 \le k \le m} (\log x_k) - \sigma_k \iff \max\{x_1/e^{\sigma_1 t}, \dots, x_m/e^{\sigma_m t}\} < 1.$$

Define $t = t(w) = \max_{1 \le k \le m}(\log \|w_k\|) - \sigma_k$ for $w \in \mathbb{R}^d \setminus \{0\}$. Then

$$g(w) = g_0(\|w_1\|/e^{\sigma_1 t(w)}, \dots, \|w_m\|/e^{\sigma_m t(w)}), \quad w = (w_1, \dots, w_m) \in \mathbb{R}^{d_1+\cdots+d_m}.$$

Proposition 17.21. *Let ρ be an excess measure on $\mathbb{R}^d \setminus \{0\}$ with symmetries $\gamma^t(\rho) = e^t\rho$, $t \in \mathbb{R}$, where $\gamma^t = \mathrm{diag}(e^{\tau_1 t}, \dots, e^{\tau_d t})$ with $0 < \tau_1 \le \cdots \le \tau_d$. Suppose ρ has maximal symmetry. Let $\tilde{\gamma}^t = e^{t\widetilde{C}}$. Then $\tilde{\gamma}^t(\rho) = e^t\rho$ for $t \in \mathbb{R}$, if and only if $\widetilde{C}$ with respect to the decomposition $\mathbb{R}^d = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_m}$ has block diagonal form with blocks $\widetilde{C}_k = \sigma_k + A_k$ where A_k is skew symmetric for $k = 1, \dots, m$, and the σ_i are the distinct diagonal elements.*

Proof. We may regard the $\mathbb{R}^{d_k}$ as subspaces of $\mathbb{R}^d$. By the proposition above $\mathbb{R}^{d_k}$ is invariant under $e^{-t\sigma_k}\tilde{\gamma}^t$, and the restriction $e^{-t\sigma_k}\tilde{\gamma}^t_k = e^{t(\widetilde{C}_k - \sigma_k)}$ is a one-parameter group in $\mathrm{O}(d_k)$. Hence $\widetilde{C}_k - \sigma_k$ is skew-symmetric. See Example 18.69. □

The density of ρ is determined by its restriction to the $(d-1)$-dimensional surface $S = \partial(B_1 \times \cdots \times B_m)$, where B_k is the unit ball in $\mathbb{R}^{d_k}$, or its restriction on $\gamma^t(S)$. The symmetry imposes the condition that the density is constant on $T = \partial B_1 \times \cdots \times \partial B_m$, and on the sets $\partial r_1 B_1 \times \cdots \times \partial r_m B_m$. The compact manifold T is a subset of S, but its dimension is $d - m$.

What do *typical densities* look like? Let ρ have a continuous density g of maximal symmetry.

If $\tau_1 < \cdots < \tau_d$, then the typical density is determined by d regularly varying functions. A typical density is a continuous function of the form

$$f(z) = h_0(r_1(|z_1|), \dots, r_d(|z_d|)), \quad z = (z_1, \dots, z_d) \in \mathbb{R}^d$$

where h_0 is a continuous function which agrees with cg outside some ball rB – to ensure that f is bounded. (We are only interested in the behaviour of f far out.) The $r_k\colon [0,\infty) \to [0,\infty)$, for $k = 1, \dots, d$, are homeomorphisms of the halfline $[0,\infty)$ onto itself which vary regularly with exponent one in infinity:

$$r_k(c_n t_n)/r_k(t_n) \to c, \quad t_n \to \infty,\ c_n \to c > 0.$$

The inverse functions $z_k = r_k^{\leftarrow}$ have the same properties.

Let $\varphi\colon \mathbb{R}^d \to [0,\infty)$ be a continuous function such that $\{\varphi < r\} = E_r$ is a centered ellipsoid for $r > 0$. Assume $\{\varphi = r\} = \partial E_r$ for $r > 0$, and $\{\varphi = 0\} = \{0\}$. We shall say that φ lies in $\mathcal{R}(d)$ if moreover

$$E_{c_n r_n} \sim c E_{r_n}, \quad r_n \to \infty,\ c_n \to c > 0.$$

For $d = 1$ the ellipsoids E_r are intervals $(-z(r), z(r))$, where $z = \varphi^{\leftarrow}$ as above is a homeomorphism from $[0,\infty)$ onto itself which varies regularly in infinity with exponent one.

If $\tau_1 = \cdots = \tau_d = \tau$, then $g(w) = c/\|w\|^{d+1/\tau}$, and typical densities have the form $f(z) = h_0(\varphi(z))$, with $\varphi \in \mathcal{R}(d)$, and $h_0\colon [0,\infty) \to (0,\infty)$ a continuous function which agrees with the function $c/r^{d+1/\tau}$ for $r \geq r_0$ for some $r_0 > 0$.

In general the sequence $\tau_1, \dots, \tau_d$ contains d_k entries σ_k, with $0 < \sigma_1 < \cdots < \sigma_m$, and typical densities are continuous functions f on $\mathbb{R}^d$ of the form

$$f(z) = h_0(\varphi_1(z_1), \dots, \varphi_m(z_m)), \quad z = (z_1, \dots, z_m) \in \mathbb{R}^{d_1+\cdots+d_m}$$

with $\varphi_k \in \mathcal{R}(d_k)$ for $k = 1, \dots, m$, and h_0 a continuous function which agrees with g_0 on $[0,\infty)^m \setminus rB$ for some $r > 0$.

Proposition 17.22. *Let ρ be an excess measure on $\mathbb{R}^d \setminus \{0\}$ such that $\gamma^t(\rho) = e^t \rho$, $t \in \mathbb{R}$, where $\gamma^t = \operatorname{diag}(e^{t\tau_1}, \dots, e^{t\tau_d})$ with $0 < \tau_1 \leq \cdots \leq \tau_d$. Assume ρ has maximal symmetry and a continuous density g. Let $\beta\colon [0,\infty) \to \mathcal{A}$ vary like γ^t. There exists a probability distribution π on $\mathbb{R}^d$ with typical density f such that $e^t \beta(t)^{-1}(\pi) \to \rho$ weakly on εB^c for all $\varepsilon > 0$.*

Proof. By the Spectral Decomposition Theorem, Theorem 18.18, we may choose orthogonal affine coordinates such that $\beta(t) \sim \beta_1(t) \otimes \cdots \otimes \beta_m(t)$, $t \to \infty$, where $\beta_k\colon [0,\infty) \to \mathrm{GL}(d_k)$ is continuous, varies like $e^{\tau_k t} I$ for $k = 1, \dots, m$, and such that the function $\varphi_k \in \mathcal{R}(d_k)$ has level sets $\{\varphi_k < e^{\tau_k t}\} = \beta_k(t)(B)$. Now check that

$$e^{t_n} f(\beta_1(t_n) w_{n1}, \dots, \beta_m(t_n) w_{nm}) \to g(w),$$
$$t_n \to \infty,\ (w_{n1}, \dots, w_{nm}) \to w \in \mathbb{R}^{d_1+\cdots+d_m}.$$

Weak convergence follows as in Section 16.5. □

For exceedances over horizontal thresholds distributions of maximal symmetry may be handled in the same way. For vertical translations maximal symmetry is cylinder symmetry and typical densities are determined by a positive function g^* on $(0, \infty)$, specified by the circle symmetric spectral measure, which satisfies (15.21), and a homogeneous-elliptic function in $\mathcal{HE}_0$, which reflects the normalization, as described in Section 15.2.

The symmetry of the excess measure may be less than the symmetry allowed by the generator. If the generator $C = \mathrm{diag}(\tau_1, \dots, \tau_d)$ is real diagonal with m distinct diagonal entries $\sigma_1, \dots, \sigma_m$ with multiplicities $d_1, \dots, d_m$ as above, then one may replace the uniform distribution $d\mu$ on the manifold $T = \partial B_1 \times \cdots \times \partial B_m$ by some other distribution, say $q(\theta_1, \dots, \theta_m)d\mu(\theta_1, \dots, \theta_m)$. This determines the excess measure. Assume g_0 on $[0, \infty)^m \setminus \{0\}$ and q on T are continuous and positive. Then so is the density g of the excess measure ρ. Any probability distribution $\pi \in \mathcal{D}^\infty(\rho)$ has the form $d\pi = f d\nu$ where f is a typical density and ν a roughening of Lebesgue measure. The typical density f may be expressed in terms of the symmetric typical density, the function q and a continuous curve

$$R_1 \otimes \cdots \otimes R_m \colon [0, \infty) \to \mathrm{O}(d_1) \times \cdots \times \mathrm{O}(d_m)$$

which varies like the identity, as in the case of scalar symmetries in Section 17.2. This description of typical densities reflects the decomposition of the normalization curve given by the Spectral Decomposition Theorem. Similar techniques apply when the generator of the excess measure has a complex diagonal Jordan form.

17.6* Stable distributions and processes. For very heavy tails sample sums behave like sample extremes. In any direction one term will predominate. Given a sequence of independent observations $Z_1, Z_2, \dots$ from such a heavy tailed distribution one may form the partial sums $S_n = Z_1 + \cdots + Z_n$. If these can be normalized to converge in distribution to a non-degenerate vector W

$$\alpha_n^{-1}(S_n) \Rightarrow W,$$

with $\alpha_n(w) = A_n w + a_n$, then W is said to have a *stable* or *operator stable distribution*. The distribution of W is infinitely divisible and there is a Lévy process $S \colon [0, \infty) \to \mathbb{R}^d$ such that $S(1) = W$. This *Lévy process* is stable in the sense that all vectors $S(t)$ are of the same type. The most famous example of a stable distribution is the Gaussian distribution. The associated stable process is Brownian motion. Here we shall only consider stable distributions with heavy tails. For the general case see MS.

Lemma 17.23. *Let N_n be an n-point sample cloud on $\mathbb{R}^d$ with mean measure $\rho_n = n\,d\pi_n$, and let N_0 be a Poisson point process with mean measure the Radon* *measure ρ_0 on $O = \mathbb{R}^d \setminus \{0\}$. Suppose $\rho_n \to \rho_0$ weakly on εB^c for all $\varepsilon > 0$. Set*

$$W_n = \int_O w(dN_n - \chi d\rho_n), \quad n \geq 1, \tag{17.16}$$

where χ is continuous on $\mathbb{R}^d$ and $1_B \leq \chi \leq 1_{2B}$. If for each $\varepsilon > 0$ there exists $r > 0$ such that

$$\int_{rB} \|w\|^2 d\rho_n < \varepsilon, \quad n \geq n_\varepsilon,$$

then the integral (17.16) defines a random vector W_0 for $n = 0$, and $W_n \Rightarrow W_0$.

Proof. Vague convergence implies $\int_{rB} \|w\|^2 d\rho_0 \leq \varepsilon$. Hence W_0 is well defined, see Section 2.6. Weak convergence on εB^c for all $\varepsilon > 0$ implies

$$\int_{rB^c} w(dN_n - \chi(w)d\rho_n) \Rightarrow \int_{rB^c} w(dN_0 - \chi(w)d\rho_0)$$

whenever $\rho_0(\partial rB) = 0$. (By Skorohod's Representation Theorem we may assume $N_n \to N_0$ weakly on rB^c almost surely. This implies almost-sure convergence of the integrals $\int_{rB^c} w dN_n \to \int_{rB^c} w dN_0$.) Now observe that the vectors

$$W_n(r) = \int_{rB\setminus\{0\}} w(dN_n - d\rho_n), \quad n \geq 0$$

are centered, and that $\operatorname{var} \|W_n(r)\| \leq \int_{rB\setminus\{0\}} \|w\|^2 d\rho_n$. (Equality holds for $n = 0$ and $\operatorname{var}(\int \varphi dN) \leq \int \varphi^2 d\rho$ holds for any sample cloud with mean measure ρ and any bounded function since $\operatorname{var}(\varphi(Z)) \leq \mathbb{E}\varphi(Z)^2$.) Now apply Lemma 4.10. □

Heavy tails imply light poles. If ρ has a spherically symmetric density $h(w) = 1/\|w\|^{d+1/\tau}$ and τ is large then the level sets $\{h > e^{-n}\}$ are balls which grow fast for $n \to \infty$ since the density decreases slowly. For $n \to -\infty$ the size of the ball decreases fast and hence the pole is light. Instead of balls we shall use the ellipsoids $E_n = \alpha_n(B)$ below, but the thrust of the argument is the same. Eigenvalues of the generator in $\Re > 1/2$ ensure that the pole is so light that the integral of r^2 over the unit ball converges.

Proposition 17.24. *Let $Z \in \mathcal{D}^\infty(\rho)$ where ρ is an excess measure with expansion group γ^t. Suppose the eigenvalues of the generator C lie in the open halfplane $\Re > 1/2$. Then $\int \|w\|^2 \wedge 1 d\rho$ is finite and the stochastic integral*

$$W = \int_{\mathbb{R}^d\setminus\{0\}} w(dN - \chi(w)d\rho)$$

is well defined for the Poisson point process N with mean measure ρ and any continuous compensator χ between 1_B and 1_{2B}. Let $Z_1, Z_2, \ldots$ be independent copies of Z. The partial sums $Z_1 + \cdots + Z_n$, properly normalized, converge in distribution to W.

Proof. Let $e^t\alpha_t^{-1}(\pi) \to \rho$ weakly on εB^c for $t \to \infty$, and assume that the ellipsoids $E_t = \alpha_t(B)$ are increasing and cl(E_s)ßE_t for $0 \le s < t$, and that $\alpha_n^{-1}\alpha_{n+1} = \gamma_n \to \gamma$. Choose coordinates such that $\gamma(B) \supset e^{\tau_0}B$ for some $\tau_0 = 1/2+\delta > 1/2$. (Choose a complex Jordan form with $\varepsilon > 0$ close to zero below the diagonal instead of 1.) Then eventually $\gamma_n(B) \supset e^{\tau}B$ with $\tau = (1+\delta)/2$, and more generally there exists t_0 such that

$$e^{\tau}E_t \supset e^{\tau s}E_{t-s}, \quad 0 \le s \le t,\ t \ge t_0.$$

(If f is increasing on $[0,\infty)$ and $f(n+1) > f(n)+a$ for $n \ge n_0$ with $0 < a < c$ then $c + f(t) > sc + f(t-s)$ for $t \ge t_0$.) Since $e^t\pi(E_t^c) \to \rho(B^c)$ we also have a constant A such that

$$\pi(E_t^c) \le Ae^{-t}, \quad t \ge 0.$$

The measure $\rho_t = e^t\alpha_t^{-1}(\pi)$ satisfies

$$\rho_t(e^{-s\tau}B^c) = e^t\pi(e^{-s\tau}E_t^c) \le e^t\pi(E_{t-s-1}^c) \le Ae^{s+1}, \quad s \ge 0.$$

For $s \ge t$ the inequality is trivial since $\rho_t(\mathbb{R}^d) = e^t$. Hence

$$\int_{e^{-m\tau}B^c} r^2 d\rho_t \le \sum_{n>m} e^{-2n\tau}e^{2\tau}\rho_t(e^{-n\tau}B^c) \le e^{2\tau}Ae\sum_{n>m} e^{-n\delta} = O(e^{-m\delta}).$$

Now apply the lemma above to the sample clouds with mean measure ρ_{t_n} for $t_n = \log n$. □

The converse is less simple. Convergence of the sample clouds implies convergence of their sums to the sum of the limiting Poisson point process provided $r^2 \wedge 1 d\rho_n$ converges weakly. It is less obvious that convergence of the sums implies convergence of the summands.

For maxima the converse implication was proved in Section 7. In general maxima behave badly. Even the simple equation $a \vee x = b$ may fail to have a solution. However a *max-stable* df H determines the df of the mean measure by the equation $H = e^{-R}$. Convergence of maxima $F_n^n \to G$ entails convergence of the sample clouds by the asymptotics

$$n(1-F_n) \sim -n\log F_n \to \log G.$$

See Theorem 7.3. For sums one may use characteristic functions to show that the relation between the finite measure $\|w\|^2 \wedge 1 d\rho(w)$ and the distribution of the corresponding id vector is a homeomorphism. Moreover sums of large sample clouds may be approximated well by sums of Poisson point processes by the lemma above. See Kallenberg [2002] or MS for details.

17.7* Elliptic thresholds. Exceedances over *elliptic thresholds* suggests that one should be looking at the asymptotic behaviour of $Z^{E_n^c}$, where Z^{E^c} denotes the vector Z conditioned to lie in the complement of the ellipsoid E, and E_1ßE_2ß$\cdots$ is an increasing sequence of open ellipsoids. Do there exist α_n such that

$$W_n := \alpha_n^{-1}(Z^{E_n^c}) \Rightarrow W, \quad E_n = \alpha_n(B), \; p_n = \mathbb{P}\{Z \in E_n^c\} \to 0+, \qquad (17.17)$$

where W lives on the complement of B, and has a non-degenerate distribution? We allow affine normalizations; the ellipsoids need not be centered.

Using exceedances over horizontal thresholds as our guide, it is clear how one should proceed. Show that the limit distribution has the tail property: There are many ellipsoids $E \supset B$ such that W^{E^c} is of the same type as W. Exhibit a one-parameter group $\gamma^t, t \in \mathbb{R}$, such that the tail property holds for all ellipsoids $\gamma^t(E), t > 0$. This yields a Representation Theorem. For the Extension Theorem we choose a sequence $1 < k_1 < k_2 < \cdots$ such that $p_{k_n}/p_{k_{n+1}} \to 2$ say, and construct a continuous curve $\beta\colon [0, \infty) \to \mathcal{A}$ which varies like γ^t such that $E_{k_n} = \beta(t_n)(B)$ for a sequence t_n with $t_{n+1} - t_n \to \log 2$.

We need some conditions on the ellipsoids E_n and the probabilities p_n. We assume $p_{n+1} \sim p_n$. In our analysis we shall encounter three problems:

1) The distribution π_0 of the limit vector W may live on ∂B. Such a distribution need not be degenerate. It will be shown that in that case the vector Z has light tails.

2) The one-parameter group $\gamma^t, t \in \mathbb{R}$, exists, but γ need not be linear, and if it is linear it need not be an expansion. Even if one assumes $\mathrm{cl}(E_n)$ßE_{n+1} for all $n \geq 1$, in the limit one only obtains the weaker inclusion Bß$\gamma^t(B)$ for $t > 0$, or equivalently $\mathrm{cl}(B)$ß$\gamma^t(\mathrm{cl}(B))$. In order to ensure that the γ^t are expansions, one has to impose a condition on the rate of increase of the ellipsoids E_n.

If γ is not an expansion, the tail property may fail to hold. We shall derive a different representation. The limit distribution may be instable. On replacing the ellipsoids E_n by ellipsoids $F_n \sim E_n$, and α_n by $\beta_n \sim \alpha_n$ such that $\beta_n(B) = F_n$, one may obtain a different limit distribution.

3) If we want to embed α_n in a curve which varies like γ^t, we need the symmetries of the limit measure π_0 to be orthogonal.

Let us now first look at the solutions to the inclusion Bß$\gamma(B)$ for affine transformations γ. An example of such a transformation γ is: blow up the unit ball by a factor two from a point z_0 in B, or a point $z_0 \in \partial B$. A rotation is also possible, or, in $\mathbb{R}^3$, a rotation in the horizontal plane, and a linear expansion in the vertical direction.

Recall that Q is a *linear expansion* if the eigenvalues lie outside the unit circle in $\mathbb{C}$. The map $\gamma\colon w \mapsto Q(w - z_0) + z_0$ then is an *affine expansion* with center z_0.

Proposition 17.25. *Suppose $\gamma(B) \supset B$ for an affine transformation γ. There are two possibilities: γ is an affine expansion with center $z_0 \in \mathrm{cl}(B)$, or γ is linear and has an eigenvalue on the unit circle. In the latter case $\mathbb{R}^d = L_0 + L_1$ where L_0 and*

L_1 are invariant, and $L_0 \perp L_1$. On L_0 γ is an orthogonal transformation; on L_1 a linear expansion.

The proof is given below. We need a lemma.

Lemma 17.26. *Suppose $\mathbb{R}^d = L_0 \oplus L_1$ is the direct sum of two proper linear subspaces. Let $r_0 B_0$ be the intersection of the ball $D = p + B$ with L_0. If $D \text{ß} r B_0 + L_1$ for all $r > r_0$, then $p = 0$ and $L_0 \perp L_1$.*

Proof. In the plane this is obvious. The general situation may be reduced to $d = 2$ by restricting to $\mathbb{R}e_0 + \mathbb{R}e_1$ for unit vectors $e_i \in L_i$, $i = 0, 1$. □

Proof of Proposition 17.25. If the origin is a fix point of γ then γ is linear. The sequence $\gamma^{-n}(z)$ is bounded for any $z \in B$. By looking at the Jordan form of the matrix of γ we see that the translation part is lacking, and hence one may assume that γ is linear and satisfies

$$A \text{ß} \gamma(A), \quad A = q + C, \ C = \mathrm{cl}(B).$$

As in Section 18.9 on orbits, the Jordan form yields a decomposition $\mathbb{R}^d = L_0 \oplus L_1$, in invariant subspaces, and $\gamma = \gamma_0 \otimes \gamma_1$. Here the eigenvalues of γ_1 all lie outside the unit circle in $\mathbb{C}$, and the complex Jordan form of γ_0 is diagonal with the diagonal entries on the unit circle. We may write $L_0 = L_{00} + L_{01}$ where L_{00}, the eigenspace of the eigenvalue one, consists of the fix points of γ, and L_{01} is invariant.

First assume $L_{00} \neq \{0\}$. Let $A_0 \text{ß} L_{00}$ be the projection of A along $L = L_{01} + L_1$. Let $p_0 \in \partial A_0$. Then $(p_0 + L) \cap A = \{p\}$ for some point p, and $\gamma(p) = p$. Thus p is a fix point, and hence $p = p_0$. So $A_0 = A \cap L_{00}$. Choose the origin in the center of A_0. Then $A_0 + L \supset A$ and the lemma above implies that $q = 0$ and $L_{00} \perp L$. It remains to prove the proposition when $L_{00} = \{0\}$.

Assume $L_0 = L_{01}$. Then the origin is the only fix point. We claim that A contains the origin. Indeed the averages $a_n = (a + \gamma^{-1}(a) + \cdots + \gamma^{1-n}(a))/n$ lie in A by convexity, and from the Jordan form of γ it is clear that $a_n \to 0$. If $L_0 = \{0\}$ we are done.

So assume $L_0 \neq \{0\}$. Let $A \cap L_0 = A_0 = p_0 + C_0$, a disk of radius $r_0 \geq 0$ centered in p_0. If $r_0 = 0$ then $\gamma(p_0) = p_0$ by invariance, which implies that $p_0 = 0$ (since $L_{00} = \{0\}$). If $r_0 > 0$ let U be uniformly distributed on C_0. Then $\gamma_0(p_0 + U) = p_0 + U$ in distribution since γ_0 preserves Lebesgue measure. Taking expectations, we find $\gamma(p_0) = p_0$ and hence $p_0 = 0$ as above. Let $T = r\partial B_0 + L_1$ with $r > r_0 \geq 0$, and B_0 the unit ball in L_0. Let $z = (z_0, z_1) \in T \cap A$. Then $\gamma^{-n}(z)$ is bounded. Let $w = (w_0, w_1)$ be a limit point. Then $w_1 = 0$, and $\|w_0\| = r$. This contradicts $\|w\| \leq r_0$ for $w \in A \cap L_0$. We conclude that $T \cap A$ is empty. The lemma shows that A is centered, and $L_0 \perp L_1$. □

Let us briefly consider measures ρ which satisfy $\gamma(\rho) = \rho/q$, where $\gamma(B) \supset B$. Let $\mathcal{R}(\gamma, q) = \mathcal{R}(\gamma, q, \mathbb{R}^d)$ be the set of all such measures on $\mathbb{R}^d$. First assume $q \in (0, 1)$ and γ is an affine expansion with center $z_0 \in \mathrm{cl}(B)$. We only consider non-zero Radon measures on $\mathbb{R}^d \setminus \{z_0\}$. The measure ρ is infinite on any open neighbourhood of z_0, and finite on the complement of such a neighbourhood. Any finite measure ρ^* on $\gamma(B) \setminus B$ has a unique extension to a Radon measure $\rho \in \mathcal{R}(\gamma, q)$ on $\mathbb{R}^d \setminus \{z_0\}$ which lives on $D_\gamma = \bigcup_{n \geq 0} \gamma^n(B)$. It is finite on B^c. If $z_0 \in B$ then $D_\gamma = \mathbb{R}^d$; if $z_0 \in \partial B$ then D_γ is a proper open subset of $\mathbb{R}^d$. The set D_γ and its complement are invariant. Hence $\rho(B^c) \geq \rho(D_\gamma^c) = \infty$ if ρ charges the complement of D_γ.

Example 17.27. Let $z_0 = (0, 1) \in \mathbb{R}^2$, and let $\gamma(z) = 2(z - z_0) + z_0$ be a scalar expansion with center z_0. Then $D_\gamma = \mathbb{R} \times (-\infty, 1)$. Let ρ live on D_γ with density $g(w) = 1/\|w - z_0\|^{5/2}$. Then $\rho \in \mathcal{D}(\gamma, q)$ for some $q < 1$, and $\rho(B^c)$ is finite, even though g is not bounded on ∂B. There exist open disks $E_n \to B$ such that $\rho(E_n^c) = \infty$ for all n. ♢

Now suppose γ has an eigenvalue on the unit circle. Then γ is linear and $\gamma = \gamma_0 \otimes \gamma_1$ on $L_0 + L_1$ where γ_0 is orthogonal on L_0, γ_1 a linear expansion on L_1, and $L_0 \perp L_1$. Let $\rho \in \mathcal{R}(\gamma, q)$. If $\gamma_0 = \mathrm{id}$ then the restriction of ρ to $\{x_0\} \times L_1$ lies in $\mathcal{R}(\gamma_1, q, L_1)$. More precisely for any invariant compact set K_0ßL_0 the measure ρ restricted to $K_0 \times L_1$, projected onto L_1 lies in $\mathcal{R}(\gamma_1, q, L_1)$. This holds for any orthogonal γ_0, provided K_0 is invariant, a closed ball in L_0, or a closed ring, for instance. Again set $D_\gamma = \bigcup \gamma^n(B)$. Then $D_\gamma = B_0 \times L_1$ where B_0 is the unit ball in L_0. Any finite measure on $\gamma(B) \setminus B$ extends to a unique Radon measure $\rho_\infty \in \mathcal{R}(\gamma, q)$ on $D_\gamma \setminus L_0$. This measure is finite on B^c. If $\rho \in \mathcal{R}(\gamma, q)$ charges the complement of D_γ then $\rho(B^c) = \infty$.

For $q = 1$ the situation is different. If γ is not orthogonal then for any compact set K disjoint from L_0 or $\{z_0\}$, the images $K_n = \gamma^n(K)$ will diverge. They also have the same mass as K. One may recursively construct a sequence of positive integers $k_1 < k_2 < \cdots$ such that $K_{k_{n+1}}$ is disjoint from $K \cup K_{k_1} \cup \cdots \cup K_{k_n}$. If K is disjoint from B then $\rho(B^c) < \infty$ implies $\rho(K) = 0$. In short we have:

The measures $\rho \in \mathcal{R}(\gamma, q)$ are Radon measures on $\mathbb{R}^d \setminus M$ and live on D_γ. Here

$$M = \begin{cases} \{z_0\} & \text{if } \gamma \text{ is an affine expansion with center } z_0, \\ L_0 & \text{if } \gamma = \gamma_0 \otimes \gamma_1 \text{ where } \gamma_0 \text{ is orthogonal and } \gamma_1 \text{ a linear expansion;} \end{cases} \tag{17.18}$$

$$D_\gamma = \bigcup_{n \geq 0} \gamma^n(B). \tag{17.19}$$

We now turn to the limit relation (17.17). We begin with an example to show that affine transformations are needed in (17.17), even when the symmetries of the limit measure are linear.

Example 17.28. Let the measure ρ have density

$$g(u)/v^2, \quad u \in \mathbb{R},\ v \in \mathbb{R} \setminus \{0\}$$

where g is a probability density on $\mathbb{R}$. Then ρ is a Radon measure on the complement O of the horizontal axis, and an excess measure for the weak expansions $\gamma^t(u,v) = (u, e^t v)$. The image, $\gamma(B)$, of the open unit disk contains B but does not contain the closure of B. We shall assume that g is continuous on $[-1,1]$ and vanishes outside this interval. Then $\rho(B^c)$ is finite. There is a vector $Z = (X,Y) \in \mathcal{D}^\infty(\rho)$ with a continuous strictly positive density f for which affine normalizations may not be replaced by linear normalizations.

Let M and L be continuous functions on $[0,\infty)$ such that $L_2 = L + M$ and $L_1 = L - M$ are positive and increasing, and vary slowly in infinity. Let Y be standard Cauchy, and for $Y = y \geq 0$ (and for $Y = -y$) let X have density

$$f_y(x) = \begin{cases} c(y)g((x - M(y))/L(y)), & -L_1(y) \leq x \leq L_2(y), \\ e^{-x^2/2}/\sqrt{2\pi}, & |x - M(y)| > L(y). \end{cases}$$

Let $q > 0$. Define the rectangle $R = (-1,1) \times (-q,q)$ and set

$$\begin{aligned} R_t &= (-L_1(e^t), L_2(e^t)) \times (-qe^t, qe^t) = \alpha_t(R), \\ \alpha_t(u,v) &= (M(e^t) + L(e^t)u, e^t v). \end{aligned}$$

Then $\mathbb{P}\{Z \notin R_t\} \sim \mathbb{P}\{|Y| \geq e^t\} \sim (2/q\pi)e^{-t}$ and $(\pi/2)e^t \alpha_t^{-1}(\pi) \to \rho$ vaguely on O. The high risk scenarios $Z^{R_t^c}$ normalized by α_t converge weakly to a vector with distribution $1_{R^c}d\rho/\rho(R^c)$. With some extra effort $\alpha_t^{-1}(Z^{E_t^c})$ may be shown to converge to a vector on B^c for the ellipses $E_t = \alpha_t(B)$, and one may make f continuous and strictly positive. If $M(t)/L(t) \to q \in [-1,1]$ one may choose $(0,q)$ as the new origin and use linear normalizations; if there is no limit, linear normalization is not possible. ◊

Lemma 17.29. *Suppose $\nu_n \to \nu_0$ weakly on $\mathbb{R}^d$. Let D_n be open convex sets, $D_n \to D$ with D open and non-empty. Suppose $d\sigma_n = 1_{D_n^c} d\nu_n \to d\sigma$ weakly. Then σ lives on D^c, and $d\sigma = d\nu_0$ on $U = \mathrm{int}(D^c)$.*

Proof. The measure σ lives on D^c since $\int \varphi d\nu_n \to 0$ for any continuous φ with compact support in D. Vague convergence $\sigma_n \to 1_U d\nu_0$ holds on U since $\int \varphi d\sigma_n \to \int_U \varphi d\nu_0$ for any continuous φ with compact support contained in U. □

We are interested in the asymptotic behaviour of the high risk scenarios $Z^{E_n^c}$ where $E_n = \alpha_n(B)$ are increasing ellipsoids. Let π be the distribution of Z, π_0

the non-degenerate distribution of W, and set $\mu_n = \alpha_n^{-1}(\pi)/p_n$. The basic limit relation (17.17) states that $1_{B^c} d\mu_n/p_n \to \pi_0$ weakly on $\mathbb{R}^d$.

Introduce the set Λ of all points $(\gamma, q) \in \mathcal{A} \times (0, 1]$ for which there exist sequences $k_n < m_n$ such that

$$\gamma_n := \alpha_{k_n}^{-1}\alpha_{m_n} \to \gamma, \quad q_n := p_{m_n}/p_{k_n} \to q, \quad m_n > k_n \to \infty.$$

Suppose $(\gamma, q) \in \Lambda$. Then

$$\gamma_n(\mu_{m_n}) = \mu_{k_n}/q_n \;\text{ on } \mathbb{R}^d. \tag{17.20}$$

Set $D_n = \gamma_n(B)$ and $D = \gamma(B)$. Then $D_n \to D$, and D_n and D contain B. One finds

$$\begin{aligned} 1_{B^c} d\mu_{k_n}/q_n &\to \pi_0/q \quad &&\text{weakly on } \mathbb{R}^d, \\ 1_{D_n^c} d\mu_{k_n}/q_n = \gamma_n(1_{B^c} d\mu_{m_n}) &\to \gamma(d\pi_0) &&\text{weakly on } \mathbb{R}^d. \end{aligned}$$

Lemma 17.29 with $\sigma = \pi_0/q - \gamma(\pi_0)$ gives

$$\begin{aligned} \gamma(d\pi_0) &= d\pi_0/q \quad \text{on } \operatorname{int}(D^c), \\ \gamma(d\pi_0) &\leq d\pi_0/q \quad \text{on } \partial D. \end{aligned}$$

Hence $\pi_0(\gamma(K)) = q\pi_0(K)$ for compact K disjoint from $\mathrm{cl}(B)$ and $\pi_0(\gamma(K)) \geq q\pi_0(K)$ for compact Kß∂B. The measure $\sigma = \sigma(\gamma, q)$ lives on ∂B. For compact sets K disjoint from B:

$$\pi_0(\gamma(K)) = q(\pi_0(K) + \sigma(K)), \quad K \text{ß} B^c,\ \sigma = \sigma(\gamma, q). \tag{17.21}$$

The probability measure π_0 above satisfies the relation $\pi_0(\gamma(K)) = q\pi_0(K)$ apart from the defect σ on ∂B. Outside the closed unit ball the probability measure π_0 behaves like the measure $\rho \in \mathcal{R}(\gamma, q)$. We first show that any γ with $(\gamma, 1) \in \Lambda$ is orthogonal.

Proposition 17.30. *If* $(\gamma, 1) \in \Lambda$ *then* γ *is orthogonal, and* $\gamma(\pi_0) = \pi_0$.

Proof. Choose K disjoint from $B \cup M$ and compact. Then $K_1 = \gamma(K)$ is disjoint from $D \cup M$, and hence K_1 is disjoint from $\mathrm{cl}(B) \cup M$, and $\pi_0(\gamma^n(K)) = \pi_0(K_1)$ for all $n \geq 1$. If γ is not orthogonal then by the argument above the sets $\gamma^n(K)$ diverge, $\pi_0(K_1) = 0$, and by (17.21) $\pi_0(K) = 0$. So π_0 lives on M. Since π_0 is non-degenerate $M = L_0 = \mathbb{R}^d$ and $\gamma = \gamma_0$, and is orthogonal. This proves that γ is orthogonal. Then (17.21) with $K = \partial B = \gamma(K)$ gives $\sigma(\partial B) = 0$, and hence $\pi_0(\gamma(K)) = \pi_0(K)$ for all compact sets KßB^c. □

Without conditions one can get any limit.

Example 17.31. Let ρ be a Radon measure on $\mathbb{R}^d \setminus \{0\}$ such that $\rho(B^c) = 1$. We shall construct a vector Z with probability measure π on $\mathbb{R}^d$ such that

$$\alpha_n^{-1}(\pi)/c_n \to \rho \text{ vaguely on } \mathbb{R}^d \setminus \{0\}.$$

Here c_n are positive constants which satisfy $c_{n+1} \sim c_n \to 0$. Moreover the ellipsoids $E_n = \alpha_n(B)$ satisfy $E_1 ß E_2 ß \cdots$ and $p_n = \pi(E_n^c) \sim p_{n+1} \to 0$.

Let $\rho_n \leq \rho$, $\rho_n \to \rho$ vaguely, and $\rho_n(B^c) \to 1$. We assume that ρ_n lives on $R_n = \{1/\sqrt{n} \leq \|z\| \leq \sqrt{n}\}$ with mass $\rho_n(R_n) = r_n \sim r_{n+1} > 0$. Let $Z/(2n)!$ have distribution ρ_n/r_n conditional on $Z \in S_n := \{(2n-1)! \leq \|z\| < (2n+1)!\}$, and let $\pi(S_n) = a_n \sim a_{n+1}$. Set $E_n = \alpha_n(B)$ and $\alpha_n(w) = (2m)!w$. Set $c_n = a_n/r_n$. Then $\alpha_n^{-1}(\pi)/c_n = \rho_n$ on R_n. ♢

We impose two regularity conditions on $E_n = \alpha_n(B)$ and $p_n = \pi(E_n^c)$:

1) $p_{n+1} \sim p_n > 0$, with $p_n := \pi(E_n^c) \to 0$;

2) There exists a constant $r_0 > 1$ such that

$$E_{n+1} ß E_n^{r_0} = \alpha_n(r_0 B). \tag{17.22}$$

If E is centered then $E^r = rE$, otherwise E^r is the ellipsoid E expanded by a factor r from its center. Write $\alpha_n = \alpha_1\delta_2 \ldots \delta_n$. The second condition is equivalent to

$$B ß \delta_n(B) ß r_0 B, \quad n \geq 1$$

since $E_{n+1} = \alpha_n(\delta_{n+1}(B))$. The condition is satisfied if $E_{n+1} \sim E_n$. It ensures that the sequence (δ_n) is relatively compact in $\mathcal{A}$.

We shall now describe how one may obtain a point $(\gamma, q) \in \Lambda$ with $q \neq 1$. We shall construct a point in Λ such that $q \geq 1/2$ and such that $\gamma(\text{cl}(B))$ intersects $2B^c$ if $q > 1/2$. Let $m_n > n$ be the minimal index such that E_{m_n} does not fit in $\alpha_n(2B)$ or $p_{m_n}/p_n < 1/2$. The corresponding sequence (γ_n, q_n) is relatively compact. Choose a subsequence with limit (γ, q) say. Then $q \geq 1/2$. If $q > 1/2$ then $\gamma(\text{cl } B)$ intersects $2B^c$, and hence $q < 1$ by Proposition 17.30. In the same way for any $r > 1$ and $q_0 \in (0, 1)$ one may construct a sequence $m_n \geq n$ such that the corresponding sequence (γ_n, q_n) is relatively compact and all limit points (γ, q) have the property $q = q_0$, or $q > q_0$ and $\gamma(\text{cl } B)$ intersects rB^c. This proves

Proposition 17.32. *There is a sequence $(\gamma_n, q_n) \in \Lambda$ with $q_n < 1$ and $(\gamma_n, q_n) \to (\gamma, 1) \in \Lambda$.*

Corollary 17.33. *Successive ellipsoids are asymptotically equal.*

Proof. Let $\delta_n := \alpha_n^{-1}\alpha_{n+1} \to \text{id}$. Then $\alpha_n^{-1}(E_{n+1}) = \delta_n(B) \to B$ is equivalent to $E_{n+1} \sim E_n$. The sequence (δ_n) is relatively compact. It suffices to prove that limit points are orthogonal transformations. So suppose $\delta_n \to \gamma$. Then $p_{n+1}/p_n \to 1$. Hence $(\gamma, 1) \in \Lambda$. Now apply Proposition 17.30. □

Proposition 17.34. *Suppose* $(\gamma, q) \in \Lambda$ *with* $q < 1$. *If* $\gamma(B) = B$ *then* $\pi_0(\partial B) = 1$.

Proof. Rings $\{r_0 \leq \|z\| \leq r_1\}$ with $1 < r_0 < r_1$ are invariant under γ. The relation (17.21) gives $\pi_0(R) = q\pi_0(R)$, hence $\pi_0(R) = 0$, and $\pi_0\{\|z\| > 1\} = 0$. □

Let us now first look at the case where π_0 lives on ∂B.

Proposition 17.35. *If* W *lives on* ∂B *then* Z *has thin tails. There is a sequence of ellipsoids* $F_n = E_{k_n}$ *such that* $F_{n+1} \sim F_n$ *and* $\mathbb{P}\{Z \in F_{n+1}^c\}/\mathbb{P}\{Z \in F_n^c\} \to 0$. *In particular* $\mathbb{E}\|Z\|^n < \infty$ *for all* $n \geq 1$.

Proof. If $(\gamma, q) \in \Lambda$ and $\gamma(B) = E \neq B$ then (17.21) gives $\pi_0(\mathbb{R}^d \setminus \text{cl}(B)) \geq q\pi_0(\partial B \setminus \partial E) > 0$ since π_0 is non-degenerate and $\partial B \cap \partial E$ lies in a subspace. So $(\gamma, q) \in \Lambda$ implies $\gamma(B) = B$. Let $q_0 \in (0, 1)$ and let $r > 1$ be small. As above choose $m_n > n$ minimal so that E_{m_n} does not fit in $\alpha_n(rB)$ or $p_{m_n}/p_n < q_0$. Let (γ, q) be a limit point of the associated sequence (γ_n, q_n). Since $\gamma(B) = B$ for $(\gamma, q) \in \Lambda$ we conclude that $q = q_0$ and $q_n < q_0$ eventually. This holds for any $r > 1$ and any $q_0 \in (0, 1)$. □

Example 17.36. Let Z have density $f(z) = ce^{-n_D(z)}$ where n_D is the gauge function of a bounded convex open set Dß$\mathbb{R}^d$ containing the origin. The constant may be computed, $1/c = d!|D|$. Let $\alpha_r(w) = rw$ for $r > 0$. Then $\alpha_r^{-1}(Z^{rD^c}) \Rightarrow W$ for $r \to \infty$ where W lives on ∂D with density $g(r\theta) \propto r^d$ with respect to the uniform distribution on ∂B since $\mathbb{P}\{W \in C\} = |C \cap D|/|D|$ for any cone C. So ellipsoids are not needed here. Now suppose $D = B$ is the open unit disk in $\mathbb{R}^2$. The limit vector $W = (U, V)$ is uniformly distributed on the unit circle if $\alpha_n(u, v) = r_n(u, v)$ with $r_n = \sqrt{n}$. Let $\beta_n(u, v) = (r_n u, (r_n + a)v)$. Then $\beta_n \sim \alpha_n$ since $\alpha_n^{-1}\beta_n(u, v) = (u, (1 + a/r_n)v) \to (u, v)$. For $a > 0$ the ellipses $E_n = \beta_n(B)$ are slightly elongated in the vertical direction. The normalized high risk scenarios converge, $\beta_n^{-1}(Z^{E_n^c}) \Rightarrow W'$, but the limit vector is no longer uniformly distributed over the circle. A simple computation gives the density $g(u, v) = c_a e^{-av^2}$ on $\{u^2 + v^2 = 1\}$. The reason for the *instability* of the limit relation becomes clear if one looks at the mean measure ρ_n of the normalized sample clouds. The density of ρ_n goes to infinity uniformly and monotonically on any compact subset of D. ◊

Proposition 17.37. *Suppose* $(\gamma, q) \in \Lambda$, $q < 1$, *and* $\pi_0(\partial B) < 1$. *Then* π_0 *lives on* $D_\gamma \cup (M \cap \partial B)$.

Proof. Let U be the complement of $\text{cl}(D_\gamma)$. Then U is invariant under γ since D_γ is, and $U \cap \text{cl}(B) = \emptyset$. Hence $\pi_0(U) = q\pi_0(U)$ which gives $\pi_0(U) = 0$. The boundary ∂D_∞ is invariant under γ, and $\pi_0(\partial D_\gamma \setminus \text{cl}(B)) = q(\partial D_\gamma \setminus \text{cl}(B))$ gives $\pi_0(D_\gamma \cup \text{cl}(B)) = 1$. Finally observe that D_γ contains B and $\partial B \setminus M$. □

We shall now show that the support S_0 of the limit distribution π_0 determines the kind of transformations γ for which $(\gamma, q) \in \Lambda$, with $q < 1$. Define the linear subspace L

$$L = \bigcap\{\xi = 0\}, \quad \xi(S_0) \text{ is bounded}, \tag{17.23}$$

the intersection of the hyperplanes $\{\xi = 0\}$ for which $\xi(S_0)$ is bounded. If we project π_0 along L we obtain a probability measure with bounded support. If $L = \mathbb{R}^d$ then γ is an affine expansion; if $L \neq \mathbb{R}^d$ then γ has an eigenvalue on the unit circle, $\gamma = \gamma_0 \otimes \gamma_1$ on $\mathbb{R}^d = L_0 + L_1$ and $L = L_1$. We shall write S_γ for the support of the Radon measure $\rho = \rho_\gamma$ on $\mathbb{R}^d \setminus M_\gamma$ which satisfies $\gamma(\rho) = \rho/q$ and agrees with π_0 outside $\mathrm{cl}(B)$.

Proposition 17.38. *Let $(\gamma, q) \in \Lambda$ with $q < 1$. If γ is an affine expansion, then $\xi(S)$ is unbounded for every non-zero linear functional ξ.*

Proof. Let $K \supset B$ be compact. Then $S_0 \setminus K = S_\gamma \setminus K$. If the non-zero linear functional ξ is bounded on this set, then S_γß$\{\xi = \xi z_0\}$ by Proposition 16.10, and $\pi_0 \leq \rho$ on $\mathbb{R}^d \setminus \{z_0\}$ implies S_0ß$\{\xi = \xi z_0\}$. So π_0 is degenerate. □

Similarly one shows:

Proposition 17.39. *Suppose $(\gamma, q) \in \Lambda$ and $q < 1$. Let $L \neq \mathbb{R}^d$. Then $\gamma = \gamma_0 \otimes \gamma_1$ on $L_0 + L_1$ and $L_1 = L$.*

The support of π_0 determines the structure of the symmetries γ, and hence the subspaces L_0 and L_1 if γ has an eigenvalue on the unit circle. The measure π_0 on the complement of the closed unit ball determines the Radon measure ρ.

Proposition 17.40. *If $\pi_0(\partial B) < 1$ then $\pi_0((\partial B) \setminus M) = 0$.*

Proof. Let $(\gamma, q) \in \Lambda$, and $q < 1$. Let $A = (\partial B) \setminus M$, and suppose $\pi_0(A) = p > 0$. Then the sets $A_n = \gamma^n(A)$ are disjoint and $\pi_0(A_1) = p_1 \geq qp$, and $\pi_0(A_{n+1}) = q^n p$ gives $p \leq 1/q - 1$. Now choose a sequence $(\gamma_n, q_n) \in \Lambda$ such that $q_n < 1$ and $q_n \to 1$. If $L \neq \mathbb{R}^d$ then $M = L^\perp$ does not depend on n, and we are done. If $L = \mathbb{R}^d$ the γ_n are affine expansions with centers $z_n \in \mathrm{cl}(B)$. If $z_n \in B$ infinitely often then $\pi_0(\partial B) = 0$. If $\pi_0(\partial B) > 0$ then there exists a point $z_0 \in \partial B$ such that $\pi_0(\partial B \setminus \{z_0\}) = 0$ and γ_n has center z_0 eventually. □

Our next example shows that $\pi_0(\partial B) \in (0, 1)$ is possible.

Example 17.41. Let Y have a standard Cauchy distribution, and let Z have distribution π, where π is a mixture of the distribution of the vector $(0, 0, Y)$ and the *uniform distribution* on the unit ball B in $\mathbb{R}^3$. Let $E_n = \alpha_n(B)$ where $\alpha_n = \mathrm{diag}(1 - \varepsilon_n, 1 - \varepsilon_n, n)$. If $\varepsilon_n = 1/n$ then $\alpha_n^{-1}(Z^{E_n^c}) \Rightarrow W$ where $W = (0, 0, V)$ and V has density $1/2v^2$ on $\mathbb{R} \setminus (-1, 1)$. If $\varepsilon_n = c/n^{2/3}$ with $c > 0$, then the limit distribution is non-degenerate: it is a mixture of the distribution of $(0, 0, V)$ above and the uniform distribution on the unit circle in the horizontal coordinate plane. ◇

What do the limit laws look like? There are two situations. If ξW is bounded for a non-zero linear functional ξ then there is a linear subspace $M = L_0$ determined by the support of W, as described above, such that $\gamma = \gamma_0 \otimes \gamma_1$ for $(\gamma, q) \in \Lambda$ where γ_0 is an orthogonal transformation on L_0 and γ_1 a linear expansion on $L_1 = L_0^{\perp}$. If ξW is unbounded for all $\xi \neq 0$ then γ is an affine expansion with center $z_0 \in \mathrm{cl}(B)$, and $M = \{z_0\}$. In principle z_0 depends on γ.

In general $\pi_0 = \sigma_0 + \mu_0$, where $d\sigma_0 = 1_{\partial B} d\pi_0$ lives on $M \cap \partial B$, $\gamma(\sigma_0) = \sigma_0$, and μ_0 lives on B^c and satisfies

$$\mu_0(\gamma(K)) = q\mu_0(K), \quad K \subset B^c \text{ compact}, (\gamma, q) \in \Lambda. \tag{17.24}$$

The measure σ_0 may change when one replaces the normalizations α_n by normalizations $\beta_n \sim \alpha_n$.

Theorem 17.42 (Representation). *Let Z be a random vector in $\mathbb{R}^d$ with distribution π. Let E_1ßE_2ß$\cdots$ be open ellipsoids. Assume (17.17) holds, where W has a non-degenerate distribution π_0 on B^c. Let (17.22) hold, and let $p_{n+1} \sim p_n$.*

1) *If $\|W\| = 1$ a.s. then $\mathbb{E}\|Z\|^n$ is finite for all $n \geq 1$;*

2) *If $\mathbb{P}\{\|W\| > 1\}$ is positive, there exists a one-parameter group of affine transformations γ^t, $t \in \mathbb{R}$ such that $\gamma^t(B) \supset B$ for $t > 0$, and a Radon measure ρ on $\mathbb{R}^d \setminus M$, which lives on $D_\infty = \bigcup \gamma^n(B)$, such that $\pi_0(M) = \pi_0(\partial B) = \pi_0(M \cap \partial B)$, and*

$$d\pi_0 = d\rho \quad \text{on } (B \cup M)^c, \quad \gamma^t(\rho) = e^t \rho,\ t \in \mathbb{R},$$

and such that one of the following holds.

2a) *γ^t are expansions with center $z_0 \in \mathrm{cl}(B)$, and $M = \{z_0\}$.*

2b) *There are proper invariant linear subspaces L_0 and L_1, such that $\gamma^t = \gamma_0^t \otimes \gamma_1^t$ where γ_0^t are orthogonal linear transformations on $L_0 = M$, and γ_1^t is a linear expansion group on $L_1 = M^{\perp}$.*

Proof. We have to construct a one-parameter group γ^t, $t \in \mathbb{R}$, such that $B \subset E_t = \gamma^t(B)$ for $t > 0$, and $\gamma^t(d\mu_0) = e^t d\mu_0$ on E_t^c. Let us first construct the ellipsoids E_t for $t > 0$. Let $(\gamma_n, q_n) \in \Lambda$ such that $q_n^{m_n} \to 1/e$ for a sequence $m_n \to \infty$, and set $\tau_n = \gamma_n^{m_n}$, and $E_n = \tau_n(B)$. We claim that the sequence (E_n) is bounded. For any ellipsoid E and any unit functional θ define $S_\theta(E) = \{c_1 \leq \theta \leq c_2\}$ to be the smallest closed slice containing E, and set $\delta_0(E) = \inf_\theta \mu_0(S_\theta(E)^c)$. Note that the inf is achieved in some point θ_0. We concentrate on the situation where the γ_n are affine expansions with center $z_n \in \mathrm{cl}(B)$. If μ_0 is degenerate it lives on a hyperplane through z_0, and π_0 is degenerate. Hence $\delta_0(B)$ is positive, and $\delta_0(E_n) \to \delta_0(B)/e$. First suppose $z_n \to z_0 \in B$. If the diameter of E_n diverges, there exist directions θ_n such that the slice $S_{\theta_n}(E_n)$ contains a centered ball $r_n B$ with $r_n \to \infty$, and hence $\delta_0(E_n) \to 0$. If $z_0 \in \partial B$ we replace B by $2B$ and consider the limit of the high risk scenarios $Z^{\alpha_n (2B)^c}$.

By a diagonal procedure we may extract a subsequence so that $\gamma_n^{[m_n t]}(B) \to E_t$ for all $t > 0$, where $E_t, t \geq 0$, is an increasing sequence of ellipsoids, and $\mu_0(E_t)^c = e^t \mu_0(B^c)$. The sets E_t^c play the role here of the horizontal halfspaces in Section 14.8. There exist β_r, $r > 0$, such that $\beta_r(E_t) = E_{t+r}$ for all $t \geq 0$, and $\beta_r(d\mu_0) = e^r(d\mu_0)$ on E_r^c. Indeed for any $r > 0$ there may exist many such β_r. Quotients $\alpha = \beta_r^{-1}\beta_r'$ map B into itself and leave μ_0 invariant. We may now use Lemma 18.79 to construct the generator of the one-parameter group γ^t as in Theorem 14.7. □

With the sequence of high risk scenarios $Z^{E_n^c}$ we have associated an *excess measure* ρ on $\mathbb{R}^d \setminus M$, extending μ_0, and a one-parameter group of affine transformations $\gamma^t, t \in \mathbb{R}$, which satisfy $\gamma^t(\rho) = e^t\rho$ and $\gamma^t(B) \supset B$ for $t > 0$. Let $\mathcal{S}$ denote the *compact group* of affine transformations σ which satisfy $\sigma(\rho) = \rho$. See Theorem 16.11. We introduce three conditions:

1) $\pi_0(\partial B) = 0$;
2) any excess measure ρ_1 which agrees with ρ on B^c coincides with ρ;
3) $\mathcal{S}$ is a group of orthogonal transformations: $\sigma(B) = B$ if $\sigma(\rho) = \rho$.

The first condition holds if we replace B by a slightly larger ball rB. The second condition is discussed in Sections 18.10 and 18.11. It ensures that for each pair $(\beta, e^{-t}) \in \Lambda$ with $t > 0$ one may write $\beta = \gamma^{-t}\sigma$ for some $\sigma \in \mathcal{S}$. The third condition ensures that the symmetries of π_0 are exactly the measure preserving symmetries of ρ.

Theorem 17.43 (Extension). *Let the assumptions of the previous theorem hold, and also the first two conditions above. Set $M_\epsilon = M + \epsilon B$. Then*

$$\alpha_n(\pi)/p_n \to \rho \ \text{ weakly on } M_\epsilon^c,\ \epsilon > 0. \tag{17.25}$$

Moreover there is a continuous curve $\beta\colon [0,\infty) \to \mathcal{A}$, which varies like γ^t such that

$$e^t\beta(t)^{-1}(\pi) \to \rho \ \text{ weakly on } M_\epsilon^c,\ \epsilon > 0,\ t \to \infty. \tag{17.26}$$

If the third condition also holds one may choose the curve β such that $\beta(-\log p_n) = \alpha_n\sigma_n$ for a sequence (σ_n) in $\mathcal{S}$.

Proof. We first prove that the limit relation (17.25) holds for $\epsilon/2$ if it holds for ϵ. Choose r so large that $\sigma(B)$ßrB for all $\sigma \in \mathcal{S}$. Choose $b > 0$ so large that $\gamma^{-b}(M_{2r}) \subset M_1$. Let $p_{k_n}/p_n \to e^b$. Then $\beta_n = \alpha_n^{-1}\alpha_{k_n}$ is relatively compact. For simplicity assume convergence, $\beta_n \to \beta$. Then

$$\alpha_n^{-1}(\pi)/p_n = \beta_n(\alpha_{k_n}^{-1}(\pi)/p_{k_n})(p_{k_n}/p_n) \to e^b\beta(\rho) = \rho$$

on $\beta(M_\epsilon^c) = \gamma^{-b}(M_{r\epsilon}^c) \supset M_{\epsilon/2}^c$. This proves (17.25). In the general case we use a subsequence argument.

Now choose $1 < k_1 < k_2 < \cdots$ such that $p_{k_n} \sim e^{-n}$. Define $\alpha_{k_n} = \alpha_1\delta_1\dots\delta_n$. There is a sequence $\tau_n \in \mathcal{S}$ such that $\beta(n) := \alpha_{k_n}\tau_n$ satisfies $\beta(n)^{-1}\beta(n+1) \to \gamma$. Set $\gamma_n = \tau_{n-1}^{-1}\delta_n\tau_n$, with $\tau_0 = \text{id}$. Then $\alpha_1\gamma_1\dots\gamma_n = \alpha_{k_n}\tau_n = \beta(n)$, and $\gamma_n = \beta_{n-1}^{-1}\beta_n \to \gamma$. Interpolation gives a continuous curve $\beta\colon [0,\infty) \to \mathcal{A}$ which varies like γ^t. See Section 18.2 for details. Then (17.26) holds since ρ satisfies $\gamma^t(\rho) = e^t\rho$. We shall assume $p_{k_n} = e^{-n}$. Then $t_{k_n} = n$ and $\sigma_{k_n} = \tau_n$. Note that $\beta(t_n)(B) \sim E_n = \alpha_n(B)$. Choose $\beta_n \sim \beta(t_n)$ such that $\beta_n(B) = E_n$. Then $\beta_n^{-1}(\pi)/p_n \to \rho$ weakly on M_ϵ^c for $\epsilon > 0$, and hence $\beta_n^{-1}(Z^{E_n^c}) \Rightarrow W$. Hence $\beta_n \sim \alpha_n\sigma_n$ for a sequence $\sigma_n \in \mathcal{S}$ by the Convergence of Types Theorem. Replace $\beta(t_n)$ by $\alpha_n\sigma_n$ and interpolate to obtain a new continuous curve β, which is asymptotic to the old one since $t_{n+1} - t_n \to 0$. □

18 Regular variation and excess measures

This section is a toolbox for the first four sections of the chapter. It contains a short introduction to multivariate regular variation. There are subsections on the Meerschaert spectral decomposition, on Lie groups, and on the Jordan form for one-parameter groups of linear and affine transformations and their generators. We develop the general theory of excess measures, and discuss symmetry and local symmetry of excess measures.

We characterize the one-parameter groups for which excess measures exist. As an introduction we describe the excess measures on the plane. This general theory extends the theory of excess measures for exceedances over horizontal and elliptical thresholds developed above.

We close with an example. The final subsection exhibits a probability distribution on the plane for which exceedances over linear thresholds, exceedances over elliptic thresholds, and coordinatewise maxima give different pictures of the tail behaviour.

18.1 Regular variation. For the convenience of the reader we insert a brief introduction to multivariate regular variation here.

A function $\varphi\colon [0,\infty) \to \mathbb{R}$, whose slope tends to a constant λ in infinity, satisfies the relation:

$$\varphi(t_n + s_n) - \varphi(t_n) \to \lambda s, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{18.1}$$

In particular the sequence $b_n = \varphi(n)$ satisfies $b_{n+1} - b_n \to \lambda$; and, conversely, for any sequence (b_n) such that $b_{n+1} - b_n \to \lambda$ the continuous piecewise linear function ψ with value b_n in $n = 0, 1, \dots$ satisfies (18.1).

Proposition 18.1. *If φ satisfies (18.1), and agrees with the continuous piecewise linear function ψ above in the points $t = 0, 1, \dots$, then $\varphi - \psi$ vanishes in infinity.*

Proof. Suppose $t_n \to \infty$. Set $s_n = t_n - [t_n]$. Assume $s_n \to s \in [0,1]$. Then

$$\varphi(t_n) - \psi(t_n) = \varphi(t_n) - \varphi([t_n]) - \psi(t_n) + \psi([t_n]) \to \lambda s - \lambda s = 0.$$

Conclusion: for any $\varepsilon > 0$ eventually $|\varphi(t) - \psi(t)| < \varepsilon$. □

If φ satisfies (18.1) then $\varphi_0(t) = \varphi(t) - \lambda t$ locally is close to horizontal for $t \to \infty$. What about the global behaviour of φ_0? That may be wild. For any sequence of reals $c_0, c_1, \ldots$ there exists a function φ_0 whose slope tends to zero, and which satisfies $\varphi_0(t_n) = c_n$ for a sequence of time points $0 = t_0 < t_1 < \cdots$. Indeed, one may (obviously) construct φ_0 to be linear on the intervals $[t_n, t_{n+1}]$ with slope $1/2^n$.

If $\dot\varphi(t) \to \lambda$ then $f(t) = e^{\varphi(t)}$ solves the linear ODE $\dot f(t) = \lambda(t) f(t)$ with $\lambda(t) \to \lambda$ for $t \to \infty$; and f also satisfies

$$f(t_n + s_n)/f(t_n) \to e^{\lambda s}, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{18.2}$$

Conversely, by Proposition 18.1, any $f\colon [0,\infty) \to (0,\infty)$ which satisfies (18.2) is asymptotic to a piecewise C^1 function g which satisfies $\dot g(t) = \lambda(t) g(t)$, with $\lambda(t) \to \lambda$, for $t \to \infty$. A function which satisfies (18.2) will be said to *vary like* $e^{\lambda t}$.

If f is measurable, the weaker condition:

$$f(t+s)/f(t) \to e^{\lambda s}, \quad t \to \infty,\ s \in \mathbb{R}, \tag{18.3}$$

implies (18.2) by the Van Aardenne–De Bruijn–Korevaar Uniform Convergence Theorem. The last limit relation is better known in the form

$$R(xy)/R(x) \to y^{\lambda}, \quad x \to \infty,\ y > 0, \tag{18.4}$$

where we write $x = e^t$, $y = e^s$, and $R(u) = f(r)$ for $u = e^r$. The function $R\colon [1,\infty) \to (0,\infty)$ is said to *vary regularly* in infinity with exponent λ. If $R(xy)/R(x)$ has a finite limit $Q(y)$ when $x \to \infty$, for any $y > 0$, then $Q(xy) = Q(x)Q(y)$. If, moreover, R is measurable, then Q is a power function. Regular variation thus has two aspects: It describes the asymptotic behaviour of solutions of linear ODEs; and (in different coordinates) it describes the asymptotic form of a basic functional relation.

A natural question suggested by the limit relation (18.4) is: What are the minimal conditions? Does a decreasing function R, for instance the tail of a df, vary regularly if the limit relation (18.4) holds for $y = 2, 3$ and for positive integers x? See Feller [1971] Section VIII.8 for a very readable answer to this question. Regular variation is of great importance in asymptotics since it is able to pass the *Laplace transform* barrier as was shown by Karamata in 1930. Laplace transforms will not be used in these lectures. So we shall stick to the simpler expression in (18.3). There is an extensive literature on regular variation; see for instance Bingham, Goldie & Teugels

[1989], or Geluk & de Haan [1987], or Jessen & Mikosch [2006]. Aczél [1969] is a good introduction to functional equations.

We now turn to the multivariate case.

Definition. Let $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, be a one-parameter group of affine transformations. A function $\alpha \colon [0, \infty) \to \mathcal{A}$ *varies like* γ^t if

$$\alpha(t_n)^{-1}\alpha(t_n + s_n) \to \gamma^s, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{18.5}$$

Actually one is not so much interested in the function α as in the class of functions β which are asymptotic to α in infinity in the sense that $\alpha(t)^{-1}\beta(t) \to \text{id}$ for $t \to \infty$. If β varies like γ^t, and α is asymptotic to β, then (18.5) holds for α. This is an easy exercise. It is also easy to see that the limit $\gamma(s)$ in (18.5) satisfies the functional equation $\gamma(s_1 + s_2) = \gamma(s_1)\gamma(s_2)$, if it exists for each $s \in \mathbb{R}$.

Definition. If $A \colon [1, \infty) \to \text{GL}$ is measurable and

$$A(xy)A(x)^{-1} \to y^C, \quad x \to \infty,\ y > 0,$$

then A *varies regularly with exponent* C.

Note the order of the factors on the left. Set $\alpha(t) = A^{-1}(e^t)$ (or $\alpha(t) = A^T(e^t)$). By the multivariate version of the Uniform Convergence Theorem α then varies like $\gamma^t = e^{-tC}$ (or like e^{tC^T}). For a proof of the UCT see Balkema [1973] or Meerschaert & Scheffler [2001]. The latter is the standard work on multivariate regular variation. It may seem confusing to have two names for concepts which are so closely related. In practice it is convenient to have two approaches to the same fundamental theory. In the univariate case it is reasonable to describe the normalizations for heavy tails in terms of regular variation. For distributions in the domain of the exponential limit law the scaling varies slowly, but the full normalization varies like a group of translations; see Section 18.3 below. The basic limit relation (18.5) is equivalent to uniform convergence on bounded s-intervals for $t \to \infty$. So, by the UCT, condition (18.5) is weaker than the conditions imposed in the definition of regular variation above.

Definition. A *one-parameter group* of affine transformations γ^t, $t \in \mathbb{R}$, on $\mathbb{R}^d$ is a group of affine transformations which may be represented by matrices of size $1 + d$ as in (2) in the Preview, as $\gamma^t = e^{tC}$. The *generator of a Lie group* C is any matrix of size $1 + d$ with top row zero.

For a matrix in *Jordan form* one can write down the exponential and logarithm by hand, since the power series break off after a finite number of terms; see (18.22).

Example 18.2. Let $\gamma^t = e^{tC}$ be a one-parameter group of affine transformations, with generator $C \neq 0$, and let $\varphi \colon \mathbb{R} \to \mathbb{R}$ satisfy the *Cauchy equation*: $\varphi(t + s) =$

$\varphi(t) + \varphi(s)$. Then $G(t) = e^{\varphi(t)C}$ satisfies $G(t+s) = G(t)G(s)$. There exist non-measurable solutions of the Cauchy equation. (Just write $\mathbb{R}$ as a vector space over the rationals $\mathbb{Q}$, choose an (uncountable) base, say $1, \sqrt{2}, \sqrt{3}, \dots$ and let φ interchange two base elements, say 1 and $\sqrt{2}$.) Then $G(t), t \in \mathbb{R}$, is *not* a one-parameter group. ◊

The parameter t of a curve $\alpha(t)$ which varies like $\gamma^t = e^{tC}$ may be interpreted as time. The asymptotic behaviour of $\alpha(t)$ is related to stability of the corresponding ODE, $\dot{\xi} = \xi C(t)$, with $C(t) \to C$. See Bellman [1969] and Agrachev & Sachkov [2004]. In the multivariate situation often it is the set $\{\alpha(t) \mid t \geq 0\}$ rather than the function $t \to \alpha(t)$ which one is interested in. We could declare two curves $t \mapsto \alpha(t)$ and $t \mapsto \beta(t)$ equivalent if the sets $\{\alpha(t) \mid t \geq 0\}$ and $\{\beta(t) \mid t \geq 0\}$ are equal, or asymptotically equal. This equivalence relation is rather drastic. We shall assume that there is a well-behaved time change which transforms the one curve into the other:

Definition. Suppose α varies like γ^{at}, and β like γ^{bt}, with $a, b > 0$. We say that α and β are related by a *time change* if there is a continuous strictly increasing function $\varphi\colon [0,\infty) \to [0,\infty)$ such that

$$\beta(t) = \alpha(\varphi(t)), \quad t \to \infty, \tag{18.6}$$

where φ satisfies (18.1) with $\lambda = a/b$, and $\varphi(0) = 0$.

Regular variation concerns the normalization. Our primary interest is in the excess measure and limit law. Concentrating our attention on the normalization is like taking a picture, and instead of showing the photograph, only giving technical details, like shutter speed, focal distance, light intensity. For different photographs of the same object these technical details may vary considerably even if the same camera is used. We can only point out here that statisticians have always been concerned with normalizations. In a Gaussian world the affine normalizations represent the location vector and the *covariance* matrix, the objects of supreme interest. There still is a considerable gap between the sequence of normalizations which one needs to keep the edge of the sample cloud in focus, and a continuous curve in $\mathcal{A}$ which varies like the one-parameter group of symmetries of the excess measure, but the relevance of a limit theory for the normalizations to the study of high risk scenarios should be clear.

In the univariate case regular variation often applies to distribution tails. In the multivariate case the term regular variation has been used to describe the tail behaviour of dfs on $[0,\infty)^d$ in the domain of attraction of a max-stable distribution, and more recently for fat-tailed probability distributions on $\mathbb{R}^d$ in relation to coordinatewise extremes. See Resnick [2004], Rootzén & Tajvidi [2006]. A more algebraic approach was developed in Jurek & Mason [1993]. More recently Meerschaert and Scheffler in MS have developed a truly geometric general theory of regular variation for finite measures in the domain of attraction of excess measures on $\mathbb{R}^d \setminus \{0\}$ with heavy tails.

18.2 Discrete skeletons. If the curve $\alpha\colon [0,\infty) \to \mathbb{A}$ varies like γ^t, then $\gamma_n := \alpha(n)^{-1}\alpha(n+1) \to \gamma$, and the sequence $\alpha(0), \alpha(1), \dots$ determines the asymptotic behaviour of the curve. One recovers the original curve (up to asymptotic equality) by a simple *interpolation* formula. The interpolation formula yields a curve which is continuous and piecewise C^1. Such curves are simple to handle, and will play an important role in future constructions.

Let $\alpha_n = \alpha_0\gamma_1 \dots \gamma_n$ with $\gamma_n \to \gamma$. Set

$$\alpha(t) = \alpha_n \gamma_{n+1}^{t-n}, \quad n \le t \le n+1,\ t \ge 0. \tag{18.7}$$

Then $t \mapsto \alpha(t)$ is continuous and varies like γ^t. The sequence $\beta(n)$ will be called a *discrete skeleton* for the curve β.

There is a slight technical difficulty with (18.7), as may be seen from the two examples below.

Example 18.3. There is a sequence of 2 by 2 matrices $\gamma_n \to \gamma$ such that γ^t, $t \in \mathbb{R}$, is a continuous one-parameter group in GL(2), but the γ_n do not even have a square root. One may take γ_n diagonal with determinant $\det \gamma_n = 1$. ◊

Example 18.4. A matrix γ does not determine the group γ^t, $t \in \mathbb{R}$. If R^t is the rotation over the angle t in the plane, and α maps the disk B onto the ellipse $E = \alpha(B)$, then the matrices $\gamma^t = \alpha \circ R^{t\pi} \circ \alpha^{-1}$ map E onto itself, and $\gamma = -I$ whatever α. ◊

The group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, is determined by its generator C, not by γ. So let $\alpha_n = \alpha_0\gamma_1 \dots \gamma_n$, with $\gamma_n \to \gamma = e^C$. There are two questions. Does there exist a continuous curve $\alpha\colon [0,\infty) \to \mathbb{A}$ which varies like e^{tC}, such that $\alpha(n) = \alpha_n$ eventually? If $\beta\colon [0,\infty) \to \mathbb{A}$ varies like e^{tC}, and $\beta_n \sim \alpha_n$ does it follow that $\alpha(t)^{-1}\beta(t) \to I$ for $t \to \infty$?

Proposition 18.5. *Suppose $\alpha_n = \alpha_0\gamma_1 \dots \gamma_n$ where $\gamma_n \sim e^{r_n C}$ for a sequence $r_n \in [1/M, M]$ for some $M > 1$. Set $s_n = r_1 + \cdots + r_n$. There exists a continuous $\alpha\colon [0,\infty) \to \mathbb{A}$ which varies like $\gamma^t = e^{tC}$ such that $\alpha_n = \alpha(s_n)$ eventually.*

Proof. Choose $\varepsilon \in (0, 1/M)$ such that $\|\varepsilon C\| < 1/2$. Set $D = \varepsilon C$ and $\delta^t = e^{tD}$. Then $\|\delta - I\| < 1$. (This follows from $e^{1/2} < 2$.) Hence $\delta_n \to \delta$ implies that $D_n = \log \delta_n$ is well defined by its power series for $n \ge n_0$ (as soon as $\|\delta_n - I\| < 1$), and $D_n \to D$.

Define $\alpha(s_n + t) = \alpha_n\gamma^t$ for $0 \le t \le r_{n+1} - \varepsilon$, and set $\beta_n = \alpha(s_{n+1} - \varepsilon)$. Then

$$\beta_n^{-1}\alpha_{n+1} = \gamma^{\varepsilon}\gamma^{-r_{n+1}}\alpha_n^{-1}\alpha_{n+1} = \delta\gamma^{-r_{n+1}}\gamma_{n+1} =: \delta_n \to \delta$$

since $\gamma_{n+1} \sim \gamma^{r_{n+1}}$ is given. For $n \ge n_0$ write $\delta_n = e^{D_n}$ and close the gap between β_n and α_{n+1} by defining

$$\alpha(s_{n+1} - \varepsilon + t\varepsilon) = \beta_n e^{tD_n}, \quad 0 \le t \le 1,\ n \ge n_0.$$

Finally redefine $\alpha(0) = \alpha_{n_0}\gamma^{-s_{n_0}}$ and $\alpha(s) = \alpha(0)\gamma^s$. The curve $\alpha\colon [0,\infty) \to \mathcal{A}$ is continuous and piecewise C^1, with derivative $\dot\alpha(t) = \alpha(t)C$ outside the intervals $[s_n - \varepsilon, s_n]$, $n > n_0$. On the interior of these intervals $\dot\alpha(t) = \alpha(t)C_{n-1}$ where $C_n = \varepsilon^{-1}D_n \to C$. □

Proposition 18.6. *Suppose α and β both vary like γ^t, and $\alpha(s_n) \sim \beta(s_n)$ for a sequence $s_n \to \infty$ which has bounded increments $s_{n+1} - s_n$. Then $\alpha(t)^{-1}\beta(t) \to \mathrm{id}$ for $t \to \infty$.*

Proof. It suffices to prove the result for a subsequence, which we may assume increasing with increments which satisfy $1 \le s_{n+1} - s_n \le M$ for some $M > 1$. Now proceed as in Proposition 18.1. □

18.3* Regular variation in $\mathcal{A}^+$. What do excess measures on the real line look like? There are two kinds of halflines: upper and lower halflines. By a symmetry argument it suffices to determine all excess measures which are finite on a halfline $[j_0, \infty)$.

In (6.8) we listed three classes of excess measures in their simplest form. One may translate these measures, scale them, or multiply them by a constant to extend the class. This does not yet yield all excess measures with an upper halfline of finite positive mass.

So let us start with the simpler problem of determining all one-parameter groups of affine transformations $\gamma^t\colon v \mapsto a(t)v + b(t)$ on $\mathbb{R}$. We may restrict attention to *positive* affine transformations, $\gamma \in \mathcal{A}^+$. (Similarly in the multivariate situation the linear part of γ^t has positive determinant.) Each $\gamma \in \mathcal{A}^+$ has the form $\gamma(v) = av + b$ with $a > 0$, and determines a unique one-parameter group. If $a = 1$, then γ is a translation over b, and γ^t is a translation over tb; for $a \ne 1$ one may write $\gamma(v) = a(v - c) + c$, with $(1 - a)c = b$. Then γ is a scaling with center c, and $\gamma^t(v) = a^t(v - c) + c$. Write $a = e^\tau$. Use (2) in the Preview to represent these groups γ^t as matrix groups with matrices of size two. The derivative in $t = 0$ yields the generator C since $e^{tC} = I + tC + O(t^2)$ for $t \to 0$:

$$\frac{1}{t}\left(\begin{pmatrix} 1 & 0 \\ bt & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \to \begin{pmatrix} 0 & 0 \\ b & 0 \end{pmatrix},$$

$$\frac{1}{t}\left(\begin{pmatrix} 1 & 0 \\ c - e^{\tau}c & e^{\tau t} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \to \begin{pmatrix} 0 & 0 \\ -c\tau & \tau \end{pmatrix}.$$

The lower right entry τ in C is the *Pareto parameter* of the excess measure, and of the associated GPD G_τ in (5) in the Preview. The lower left entry has no geometric significance. It depends on the coordinates, like the center c. One may choose coordinates so that C has Jordan form. It then is simple to write down the expression for $\gamma^t = e^{tC}$.

If ρ is an excess measure and $\gamma^t(\rho) = e^t\rho$, and $\rho[j_0, \infty) = 1$, then $\rho[\gamma^s(j_0), \infty) = e^{-s}$, and hence

$$\gamma(j_0) > j_0.$$

So start with a positive affine transformation γ and a real j_0 such that $\gamma(j_0) > j_0$. The orbit

$$(j_*, j^*) = \{\gamma^t(j_0) \mid t \in \mathbb{R}\} \tag{18.8}$$

is an unbounded open interval: $\mathbb{R}$ if γ is a translation $v \mapsto v + b$; (c, ∞) if γ is an expansion with center c; $(-\infty, c)$ if γ is a contraction with center c.

Proposition 18.7. *Given a positive affine transformation γ and a real j_0 such that $\gamma(j_0) > j_0$, there is a unique excess measure ρ on (j_*, j^*) such that $\gamma^t(\rho) = e^t\rho$ and $\rho[j_0, \infty) = 1$.*

Proof. Define $R\colon (j_*, j^*) \to (0, \infty)$ by $R(\gamma^t(j_0)) = e^{-t}$. This defines ρ on (j_*, j^*) by $R(v) = \rho[v, \infty)$ for $v \in (j_*, j^*)$, and $\gamma^t(\rho) = e^t(\rho)$ holds since $R(\gamma^t(v)) = R(\gamma^t\gamma^s(j_0)) = e^{-t}\gamma(v)$ for $v = \gamma^s(j_0)$. □

Do we now have all excess measures ρ which are finite and positive on some halfline $[j_0, \infty)$? No! One easily checks that $1/v^2$ is the density of an excess measure ρ on $\mathbb{R} \setminus \{0\}$. Excess measures may live on more than one orbit! In the *univariate* case we shall restrict attention to excess measures which live on an open interval, and satisfy $\rho[j_0, \infty) = 1$ for some $j_0 \in \mathbb{R}$. The *excess measure* then is determined uniquely by the *symmetry group*, and j_0.

We now want to extend the correspondence between one-parameter groups of transformations γ^t and excess measures ρ to a correspondence between curves which vary like γ^t and probability distributions in the domain of ρ. We shall not distinguish curves which are asymptotic, or dfs which are tail-equivalent. We start by selecting a well-behaved curve.

Suppose $\alpha_n = \alpha_0\gamma_1 \ldots \gamma_n$ where $\gamma_n \to \gamma$. Assume $\gamma(j_0) > j_0$. Then $\gamma_n(j_0) > j_0$ for $n \ge n_0$, and hence $\alpha_{n+1}(j_0) > \alpha_n(j_0)$ for $n \ge n_0$. Define $\alpha(t)$ by interpolation, as in (18.7): $\alpha(t) = \alpha_n\gamma_{n+1}^{t-n}$ for $n \le t \le n+1$, $n \ge n_0$. The inequality $\gamma_n(j_0) > j_0$ implies that $t \mapsto \gamma_n^t(j_0)$ is increasing on the orbit of γ_n generated by j_0. Hence $t \mapsto \alpha(t)(j_0)$ is increasing for $t \ge n_0$. We have the following result:

Proposition 18.8. *Suppose $\alpha_n = \alpha_0\gamma_1 \ldots \gamma_n$ where $\gamma_n \to \gamma$ for a positive affine transformation γ. Let $\gamma(j_0) > j_0$. Then $\alpha_n(j_0)$ is eventually increasing. Set $\lim \alpha_n(j_0) = y_\infty \le \infty$, and let $y_0 < y_\infty$. There is a continuous curve $\alpha\colon [0, \infty) \to \mathcal{A}^+$, which varies like γ^t, such that $\alpha(n) = \alpha_n$ eventually, and such that*

$$t \mapsto y(t) = \alpha(t)(j_0), \quad t \ge 0 \tag{18.9}$$

is a strictly increasing continuous map from $[0, \infty)$ onto $[y_0, y_\infty)$.

The curve $\alpha\colon [0,\infty) \to \mathcal{A}^+$ above varies like γ^t. Let $\gamma(j_0) > j_0$, and let ρ denote the excess measure on the orbit (j_*, j^*) of j_0 associated with γ^t and j_0. We shall construct a continuous df F which satisfies

$$e^t\alpha(t)^{-1}(dF) \to \rho \text{ weakly on } [v,\infty), \quad v > j_*,\ t \to \infty. \tag{18.10}$$

Proposition 18.9. *Given α, γ and j_0 as above, define the df F on the interval $[y_0, y_\infty)$ by $F(y(t)) = 1 - e^{-t}$ with $y(t)$ as in (18.9), and let R be the tail function of the excess measure ρ determined by γ and j_0 in Proposition 18.7. Then $T = 1 - F$ satisfies*

$$T(\alpha(t)(v))/T(\alpha(t)(j_0)) = e^t T(\alpha(t)(v)) \to R(v), \quad v \in (j_*, j^*),\ t \to \infty.$$

Proof. Write $v = \gamma^s(j_0)$. Let $t_n \to \infty$. For $n \geq n_0$ there exists a unique s_n such that $\alpha(t_n)(v) = \alpha(t_n + s_n)(j_0)$. (For $r > s$ the limit relation $\alpha(t_n)^{-1}(\alpha(t_n + r)(j_0)) \to \sigma^r(j_0) > v$ implies $\alpha(t_n)(v) < y(t_n + r)$ eventually. Similarly $\alpha(t_n)(v) > y(t_n + q)$ for $q < s$. By the Intermediate Value Theorem there exists $s_n \in (q, r)$ for which equality holds. Since $t \mapsto y(t)$ is strictly increasing the value of s_n is unique.) Now observe that

$$e^{t_n} T(\alpha(t_n + s_n)(j_0)) = e^{-s_n} \to e^{-s} R(j_0) = R(\gamma^s(j_0)) = R(v)$$

by definition of T, s and R. □

If $\beta\colon [0,\infty) \to \mathcal{A}^+$ varies like γ^t, then $\beta(t) \sim \alpha(t)$ for $t \to \infty$, if we define α by interpolation as above with $\alpha_n = \beta(n)$. Let $t_n \to \infty$, and $v_n \to v \in (j_*, j^*)$. Then $v'_n := \alpha(t_n)^{-1}\beta(t_n)(v_n) \to v$. Hence $v'_n = \alpha(\delta_n)(v_n)$ for a sequence $\delta_n \to 0$, and $\beta(t_n)(v_n) = \alpha(s_n)(v_n)$ if we choose $s_n = t_n - \delta_n$. In particular $\beta(t_n)(j_0) \to y_\infty$. The proposition thus gives the following result:

Theorem 18.10. *Let ρ be the excess measure on (j_*, j^*) associated with the group γ^t in $\mathcal{A}^+$, and the point j_0. Thus $\gamma^t(\rho) = e^t\rho$ for $t \in \mathbb{R}$, (j_*, j^*) is the orbit of j_0 defined in (18.8), $\gamma(j_0) > j_0$, and $\rho[j_0,\infty) = 1$. Let $\beta\colon [0,\infty) \to \mathcal{A}^+$ vary like γ^t. There exists a df F_0 such that $1 - F_0(\beta(t)(j_0)) \sim e^{-t}$ for $t \to \infty$. Any probability distribution π with df F tail asymptotic to F_0 satisfies:*

$$e^{t_n}\beta(t_n)^{-1}(\pi) \to \rho \ \textit{weakly on } [v,\infty), \quad t_n \to \infty,\ v > j_*.$$

We shall now prove the converse. A df $F \in \mathcal{D}^+(\tau)$ determines a curve in $\mathcal{A}^+$. We need to introduce the inverse function to the *von Mises function* ψ introduced in Section 6.

If ψ satisfies (6.4), the inverse function $j = \psi^{\leftarrow}$ exists, is C^2 on a neighbourhood of $\psi(y_\infty) = \infty$, and satisfies

$$\begin{aligned} A(t) &:= j'(t) = \frac{1}{\psi'(j(t))} = a(j(t)) > 0, \\ \frac{j''(t)}{j'(t)} &= -\frac{\psi''}{(\psi')^2}(j(t)) = a'(j(t)) \to 0, \quad t \to \infty. \end{aligned} \tag{18.11}$$

By the Intermediate Value Theorem

$$\frac{A(t_n + s_n)}{A(t_n)} \to 1, \quad \frac{j(t_n + s_n) - j(t_n)}{A(t_n)} \to s, \quad t_n \to \infty,\ s_n \to s \in \mathbb{R}. \tag{18.12}$$

These relations are well-known from Π-variation, the second-order theory of regular variation developed by de Haan [1970]. Set $\beta(t)(v) = j(t) + A(t)v$. Then (18.12) gives

$$\beta(t_n)^{-1}\beta(t_n + s_n)(v) = \frac{j(t_n + s_n) + A(t_n + s_n)v - j(t_n)}{A(t_n)} \to s + v =: \sigma^s(v)$$

for $t_n \to \infty$ and $s_n \to s \in \mathbb{R}$.

Theorem 18.11. *Let $F \in \mathcal{D}^+(\tau)$. Let $\gamma^t = e^{tC}$ where C is a generator with lower right entry τ, and let $\gamma(j_0) > j_0$. Let ρ be the excess measure on (j_*, j^*) associated with C and j_0. There exists a continuous curve $\alpha\colon [0,\infty) \to \mathcal{A}^+$ which varies like γ^t such that* (18.10) *holds.*

Proof. Choose coordinates such that ρ lives on $(-\infty, 0)$, $\mathbb{R}$, or $(0,\infty)$, according as $\tau < 0$, $\tau = 0$, or $\tau > 0$; and $j_0 = \operatorname{sign}(\tau)$, $\gamma^t(v) = e^\tau v$ for $\tau \neq 0$, and $\gamma^t(v) = v + t$ for $\tau = 0$. Write $T(y) = 1 - F(y)$.

First suppose $\tau > 0$. Then T varies regularly with exponent $-\lambda$ where $\lambda = 1/\tau$. We may choose T_0 continuous and strictly decreasing on $[0,\infty)$ with $T_0(0) = 1$, such that $T_0(y) \sim T(y)$ for $y \to \infty$. Define $y(t)$ by $T_0(y(t)) = e^{-t}$, and set $\alpha(t)(v) = y(t)v$. Then

$$y(t_n + s_n)/y(t_n) = T_0^{-1}(e^{-t_n + s_n})/T_0^{-1}(e^{-t_n}) \to e^{s\tau}, \quad t_n \to \infty,\ s_n \to s$$

since T_0^{-1} varies regularly with exponent $-\tau$. Hence α varies like γ^t. A similar argument holds for $\tau < 0$. The upper endpoint of the df F is finite, and one may choose coordinates so that it is the origin. For $\tau = 0$ one may assume that $1 - F \sim e^{-\psi}$ where $e^{-\psi}$ satisfies the von Mises condition, and the proof is given above. □

In the univariate case a description of the domains of attraction in terms of curves of positive affine transformations which vary like a group of affine transformations – translations, or expansions or contractions with a given center – may seem like an exercise in abstraction. In the multivariate case regular variation is a natural approach to the asymptotics of exceedances.

18.4 The Meerschaert spectral decomposition. The Meerschaert Spectral Decomposition Theorem, SDT, treats the case where the diagonal elements of the real *Jordan form* are not all equal. The SDT is the counterpart in regular variation of Oseledec's Theorem in the theory of dynamical systems. We shall discuss the theorem and some of its implications. For the proof we refer to Section 16.9, and to MS. The results in this section may be found in MS, Section 4.3. We shall discuss the application of the theorem to exceedances over horizontal thresholds, and give a simple extension for affine transformations. The SDT is a result on multivariate regular variation in GL or $\mathcal{A}$; its application is not limited to heavy tailed distributions.

It will be convenient to assume that the probability distribution of our random vector lives on a d-dimensional linear space L, on which as yet there are no coordinates. We shall write $\mathcal{L}(\mathbb{R}^d, L)$ for the space of all invertible linear maps from $\mathbb{R}^d$ to L. A function $A\colon [0,\infty) \to \mathcal{L}(\mathbb{R}^d, L)$ varies like $\gamma^t = e^{tC}$ if

$$A(t_n)^{-1}A(t_n + s_n) \to \gamma^s = e^{sC}, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{18.13}$$

For the moment γ^t, $t \in \mathbb{R}$, is a one-parameter group of linear transformations on $\mathbb{R}^d$. Affine transformations are treated in Theorem 18.18.

We first discuss the case where $\gamma^t = e^{tC}$ are diagonal linear transformations, $C = \operatorname{diag}(\tau_1, \dots, \tau_d)$, with $\tau_1 < \cdots < \tau_d$. In that case it is possible to choose the normalizations to be diagonal too, in appropriate coordinates.

Proposition 18.12. *Suppose $A\colon [0,\infty) \to \mathcal{L}(\mathbb{R}^d, L)$ satisfies (18.13), where $C = \operatorname{diag}(\tau_1, \dots, \tau_d)$ with $\tau_1 < \cdots < \tau_d$. There exists $T \in \mathcal{L}(\mathbb{R}^d, L)$, and diagonal matrices*

$$B(t) = \operatorname{diag}(e^{\varphi_1(t)}, \dots, e^{\varphi_d(t)}) \sim T^{-1}A(t), \quad t \to \infty, \tag{18.14}$$

where the $\varphi_k\colon [0,\infty) \to \mathbb{R}$ are continuous, piecewise C^1 functions, which satisfy

$$\dot{\varphi}_k(t) \to \tau_k, \quad t \to \infty,\ k = 1, \dots, d.$$

If L is an inner product space one may choose the coordinate map T to be an isometry.

Proof. This is a special case of the SDT below. □

Example 18.13. Let π be a probability distribution on the d-dimensional linear space L. Assume

$$e^t \alpha(t)^{-1}(\pi) \to \rho \ \text{ weakly on } \varepsilon B^c, \quad t \to \infty,\ \varepsilon > 0,$$

where $\alpha\colon [0,\infty) \to \mathcal{L}(\mathbb{R}^d, L)$ varies like $\gamma^t = \operatorname{diag}(e^{t\tau_1}, \dots, e^{t\tau_d})$. We assume

$$0 < \tau_1 < \cdots < \tau_d.$$

Then ρ is an excess measure on $\mathbb{R}^d \setminus \{0\}$ with symmetries $\gamma^t = e^{tC}$. The linear isomorphism $T\colon \mathbb{R}^d \to L$ from the proposition above defines coordinates on L. Let $Z = (Z_1, \ldots, Z_d)$ be the random vector on $\mathbb{R}^d$ with distribution $T^{-1}(\pi)$. Let K be the open cube $K = (-1, 1)^d$, and for $t \geq 0$ let R_t be the rectangle

$$R_t = (-b_1(t), b_1(t)) \times \cdots \times (-b_d(t), b_d(t)) = B(t)(K), \quad b_k(t) = e^{\varphi_k(t)}.$$

Z^t denotes the vector Z conditioned to lie outside R_t. Normalize the components:

$$W_t = (Z_1^t / b_1(t), \ldots, Z_d^t / b_d(t)) \in K^c.$$

Then $W_t \Rightarrow W$ for $t \to \infty$, where W has distribution $1_K^c d\rho / \rho(K^c)$. $\Diamond$

If the generator C of the linear group γ^t is in real Jordan form, the diagonal elements of C denote the expansion rate in the corresponding directions. Let $\sigma_1 < \cdots < \sigma_m$ denote the distinct diagonal elements of the generator, with positive multiplicities $d_1, \ldots, d_m$. Then one may write

$$\mathbb{R}^d = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_m}, \quad \gamma^t = \gamma_1^t \otimes \cdots \otimes \gamma_m^t,$$

where γ_k^t is a one-parameter group on $\mathbb{R}^{d_k}$. The eigenvalues of γ_k^t all lie on the circle of radius $r_k^t = e^{t\sigma_k}$ in $\mathbb{C}$. If ρ is an excess measure such that $\gamma^t(\rho) = e^t \rho$ then the marginals ρ_k on $\mathbb{R}^{d_k}$ satisfy $\gamma_k^t(\rho_k) = e^t \rho_k$. The tail behaviour of this marginal excess measure is determined by the exponent σ_k. For $\varepsilon > 0$

$$r^{-\varepsilon} \ll r^{\sigma_k} \rho_k(rB^c) \ll r^{\varepsilon}, \quad r \to \infty.$$

The SDT gives similar tail bounds for probability distributions in the domain $\mathscr{D}^\infty(\rho)$.

Let us introduce some notation. Given functions $f_k\colon \mathcal{X}_k \to \mathcal{Y}_k$ for $k = 1, \ldots, m$, one defines the function $f = f_1 \otimes \cdots \otimes f_m$ from the product $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ to the product $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ by

$$(f_1 \otimes \cdots \otimes f_m)(x_1, \ldots, x_m) = (f_1(x_1), \ldots, f_m(x_m)) \in \mathcal{Y}, \quad (x_1, \ldots, x_m) \in \mathcal{X}.$$

If L is a d-dimensional linear space, $L = L_1 + \cdots + L_m$, where L_k is a linear subspace of dimension d_k, and $d_1 + \cdots + d_m = d$, then L is isomorphic to $L_1 \times \cdots \times L_m$, and we shall use the same notation for linear maps. Thus if $A_k\colon \mathbb{R}^{d_k} \to L_k$ is a linear map for $k = 1, \ldots, m$, we write $A = A_1 \otimes \cdots \otimes A_m$ for the linear map which maps $(w_1, \ldots, w_m) \in \mathbb{R}^{d_1 + \cdots + d_m}$ into $A_1 w_1 + \cdots + A_m w_m \in L$.

Theorem 18.14 (Meerschaert Spectral Decomposition). *Let $A\colon [0, \infty) \to \mathscr{L}(\mathbb{R}^d, L)$ vary like γ^t, see (18.13). Assume $\gamma^t = \gamma_1^t \otimes \cdots \otimes \gamma_m^t$ where γ_k^t is a one-parameter group of linear transformations on $\mathbb{R}^{d_k}$, all eigenvalues of γ_k lie on the circle of radius $r_k = e^{\sigma_k}$ in $\mathbb{C}$, $r_1 < \cdots < r_m$ and $d_1 + \cdots + d_m = d$. There exist linear subspaces L_kßL of dimension d_k for $k = 1, \ldots, m$, which span L, and continuous curves $B_k\colon [0, \infty) \to \mathscr{L}(\mathbb{R}^{d_k}, L_k)$ which vary like γ_k^t, such that*

$$B(t) = B_1(t) \otimes \cdots \otimes B_m(t) \sim A(t), \quad t \to \infty. \tag{18.15}$$

Proof. See Meerschaert & Scheffler [2001], Theorem 4.3.10, and Corollary 4.3.12. □

The SDT allows one to transfer the spectral decomposition of the one-parameter group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, to linear normalization curves $B(t)$ which vary like γ^t. The theorem thus allows us to restrict attention to curves $A\colon [0,\infty) \to \mathcal{L}(\mathbb{R}^d, L)$ which vary like γ^t, where the γ^t are linear transformations which grow at the same rate in all directions, in the sense that all eigenvalues of γ^t lie on a fixed circle with radius $r^t = e^{t\tau}$ in $\mathbb{C}$.

The SDT becomes more transparent if one regards $A(t)$ as linear isomorphisms from one linear space L' to another linear space L. In the setting of high risk scenarios L' is the space on which the normalized vectors live, the normalized sample clouds, and the excess measure. On L' we use the coordinate w. The z-space L is the domain of the original random vector Z, the sample clouds $\{Z_1,\dots,Z_n\}$, and the high risk scenarios Z^H and Z^{E^c}. The one-parameter symmetry group γ^t, $t \in \mathbb{R}$, of the excess measure acts on L'; the generator C of this group determines the spectral decomposition

$$L' = L'_1 + \dots + L'_m, \quad \dim(L'_k) = d_k, \quad d_1 + \dots + d_m = d = \dim(L). \tag{18.16}$$

The subspaces L'_k are invariant under γ^t, and the eigenvalues of the restriction γ^t_k lie on a circle of radius $r^t_k = e^{t\sigma_k}$ with $r_1 < \dots < r_m$. The SDT gives a similar decomposition for z-space $L = L_1 + \dots + L_m$, with $\dim(L_k) = d_k = \dim(L'_k)$. Moreover one may replace the original normalization curve $A(t)$, $t \geq 0$, by a curve $B(t) = B_1(t) \otimes \dots \otimes B_m(t)$, $t \geq 0$, where $B_k(t)\colon L'_k \to L_k$ varies like γ^t_k. This decomposition of $B(t)$ reflects the decomposition $\gamma^t = \gamma^t_1 \otimes \dots \otimes \gamma^t_m$.

However there is a crucial difference between the decomposition of L' and of L. The decomposition (18.16) of L' is unique, as is clear by writing C and $\gamma^t = e^{tC}$ in Jordan form. The decomposition of L reflects the rate of decrease of the tails of the distribution π of the random vector Z in the different directions. This decomposition is not unique, as we saw in Section17.4. The convention $r_1 < \dots < r_m$ for $r_k = e^{\sigma_k}$ implies that the tails of π become heavier as the index k increases. The subspace L_m of the heaviest tails is unique, and so are the subspaces

$$M_k = L_{k+1} + \dots + L_m, \quad k = 0, \dots, m-1, \tag{18.17}$$

and the dual subspaces $M_k^+ = \{\xi \mid \xi M_k = 0\}$. The subspaces M_k in this decreasing sequence have a simple intrinsic characterization in terms of the original curve $A\colon [0,\infty) \to \mathcal{L}(L', L)$.

The rvs ξZ for which ξ does not vanish on L_m all have the same tail exponent. These tails need not be comparable. The decomposition here differs from the one in Section 17.4, where tail exponents were allowed to be equal.

Let us first discuss uniqueness of the curve $B(t) = B_1(t) \otimes \cdots \otimes B_m(t)$, $t \geq 0$. Let Q be a linear transformation on L. Set $B'(t) = QB(t)$. Then:

1) The curve B' varies like γ^t since $B'(t)^{-1}B'(t+s) = B(t)^{-1}B(t+s)$. The factors Q^{-1} and Q cancel.

2) $B'(t) = B_1'(t) \otimes \cdots \otimes B_m'(t)$ where $B_k'(t) = Q_k B_k(t)$, and $Q = Q_1 \otimes \cdots \otimes Q_m$ maps $L_1 + \cdots + L_m$ onto $E_1 + \cdots + E_m$ with $E_k = Q(L_k)$ because Q is a linear isomorphism.

Conclusion: One may replace $B(t)$ by $B'(t) = QB(t)$ in the SDT provided $B'(t)$ is asymptotic to $B(t)$ for $t \to \infty$.

Proposition 18.15. *Let $A(t)$ and $B(t)$ be as in the SDT. Let $L = M_0 \supset M_1 \supset \cdots \supset M_m = \{0\}$ be the linear subspaces in (18.17). Let $E_1, \ldots, E_m$ be linear subspaces, such that $M_k = E_{k+1} + \cdots + E_m$ and*

$$\dim(E_k) = d_k = \dim(L_k') = \dim(M_{k-1}) - \dim(M_k), \quad k = 1, \ldots, m.$$

There exists a curve of invertible linear maps $A'(t) = A_1'(t) \otimes \cdots \otimes A_m'(t)$, $t \geq 0$, such that $A'(t) \sim A(t)$ for $t \to \infty$, and such that $A_k'(t) \colon L_k' \to E_k$ varies like γ_k^t for $k = 1, \ldots, m$.

For any such curve $A' \colon [0, \infty) \to \mathcal{L}(L', L)$ there exists a linear transformation Q on L such that $Q(L_k) = E_k$ and $A'(t) \sim B'(t) = QB(t)$ for $t \to \infty$.

Proof. Recall from linear algebra that one may identify the linear space $M = L_0 + M_0$ with $L_0 \times M_0$ if M is the direct sum of L_0 and M_0, i.e. if $\dim(L_0) + \dim(L_0) = \dim(M)$. If E is a subspace of M such that M is the direct sum of E and M_0, then one may identify E with the graph of a linear map $F \colon L_0 \to M_0$. The correspondence between E and F is one-to-one. Hence with E_k we may associate a linear map from L_k to E_kßM_{k-1}:

$$w_k \mapsto w_k + Q_{k+1,k}w_k + \cdots + Q_{mk}w_k, \quad Q_{jk} \colon L_k \to L_j.$$

Any linear transformation Q on L determines a matrix (Q_{ij}), where Q_{ij} is a linear map from L_j to L_i. Assume Q_{ii} is the identity on L_i, $Q_{ij} = 0$ for $i < j$, and Q_{ij} for $i > j$ is determined by the decomposition $L = E_1 + \cdots + E_m$ as above. Then $B'(t) = QB(t) = B_1'(t) \otimes \cdots \otimes B_m'(t)$ where $B_k'(t) \colon L_k' \to E_k$ varies like γ_k^t. We claim that $B'(t) \sim B(t)$ for $t \to \infty$.

Given the spectral radii $r_1 < \cdots < r_m$ of γ, choose constants $q_k < r_k < s_k$ such that

$$q_1 < r_1 < s_1 < q_2 < r_2 < s_2 < \cdots < s_{m-1} < q_m < r_m < s_m.$$

There exists $t_0 \geq 0$ such that

$$\|B_k(t)\| < s_k^t \quad \|B_k^{-1}\| \leq 1/q_k^t, \quad t \geq t_0, \ k = 1, \ldots, m.$$

Then $B^{-1}(t)B'(t) = (S_{kj}(t))$ where $S_{kj}(t) = B_k^{-1}(t)Q_{kj}B_j(t)$. Hence $(S_{kj}(t))$ is lower diagonal, $S_{kk}(t)$ is the identity for $k = 1, \dots, m$, and for $t \geq t_0, k > j$

$$\|S_{kj}(t)\| \leq \|B_k(t)^{-1}\|\|Q_{kj}\|\|B_j(t)\| \leq M(s_j/q_k)^t \to 0, \quad t \to \infty.$$

Hence $B(t)^{-1}B'(t) \to \text{id}$ for $t \to \infty$.

So we may take $A' = B' = QB$ in the proposition. This proves existence. Conversely, if A' has the desired form then A' is asymptotic to A, and hence to B by the SDT, and to $B' = QB$ by the argument above. □

Corollary 18.16. *In a Euclidean space L one may choose the subspaces L_k in the SDT to be orthogonal. Under this extra condition the decomposition $L = L_1 + \dots + L_m$ is unique.*

The spaces M_k and their duals $M_k^+ = \{\xi \mid \xi(M_k) = 0\}$ have a simple characterization in terms of the curve $A(t)$, $t \geq 0$ in the SDT.

Proposition 18.17. *Let $\sigma_k < c < \sigma_{k+1}$, where we set $\sigma_0 = -\infty$ and $\sigma_{m+1} = \infty$. Then*

$$\xi(M_k) = 0 \iff \xi_t := e^{-ct}\xi A(t) \to 0, \quad t \to \infty.$$

If $\xi \notin M_k^+$, then $\|\xi_t\| \to \infty$.

Proof. We may assume $c = 0$ by a simple transformation: replace σ_k by $\sigma_k - c$, and $A(t)$ by $e^{-ct}A(t)$. We shall assume that the subspaces L_j are orthogonal, see above, and define $N_0 = M_k$ and $N_1 = N_0^\perp$ to be the subspace of all vectors orthogonal to N_0. We write $\xi = (\xi_0, \xi_1)$ where ξ_i vanishes on N_{1-i} for $i = 0, 1$. Set $\zeta_t = \xi B(t)$. This is a linear functional on $\mathbb{R}^d$. Asymptotic equality $A(t) \sim B(t)$ for $t \to \infty$ implies

$$\|\zeta_t\| = \|\xi_t A(t)^{-1}B(t)\| \sim \|\xi_t\|.$$

So we may replace ξ_t by ζ_t. The ellipsoid $E_t = B(t)(B)$ has the form $E_t = \{Q_{0t} + Q_{1t} < 1\}$, where Q_{it} is a quadratic form on N_i. Since $\|B_i(t)\| \to 0$ for $L_i ß N_1$ and $\|B_i(t)^{-1}\| \to 0$ for $L_i ß N_0$ the ellipsoid $E_{0t} = \{Q_{0t} < 1\}$ on N_0 contains the ball $rB \cap N_0$ eventually for any $r > 1$, and the ellipsoid $E_{1t} = \{Q_{1t} < 1\}$ on N_0 is contained in the ε-ball $\varepsilon B \cap N_0$ eventually for any $\varepsilon > 0$. If $\xi_0 \neq 0$ then $\zeta_{0t} = \xi_0 B(t)$ satisfies $\zeta_{0t}(B) = \xi_0(E_{0t}) \geq r\xi_0(B)$, and hence $\|\zeta_{0t}\| \to \infty$, which gives $\|\zeta_t\| \to \infty$. If $\xi_0 = 0$, then $\zeta_t(B) = \xi_1(E_{1t}) \leq \varepsilon\xi_1(B)$, and hence $\zeta_1 \to 0$, and $\zeta \to 0$. □

Asymptotic equality $A(t) \sim B(t)$ implies that the ellipsoids $F_t = A(t)(B)$ and $E_t = B(t)(B)$ are asymptotic. This not only means that the volumes are asymptotic, $|F_t| \sim |E_t|$. By (18) in the Preview it also implies

$$\|1_{E_t} - 1_{F_t}\|_1 = o(|E_t|), \quad t \to \infty.$$

In particular for any $\varepsilon > 0$ eventually $e^{-\varepsilon} E_t \text{ß} F_t \text{ß} e^{\varepsilon} E_t$. The ellipsoids F_t also converge to the linear subspace N_0 in the sense that for any compact set $K \text{ß} N_0$ and any $\varepsilon > 0$ eventually $K \text{ß} F_t \text{ß} (N_0 + \varepsilon B)$.

We now turn to affine transformations on $\mathbb{R}^d$. These may be regarded as linear transformations on $\mathbb{R}^{1+d}$ by adding a virtual zeroth coordinate, as in (2) in the Preview. However it is not clear that this coordinate is preserved by the matrices B_k in the decomposition. Hence we write affine transformations as $w \mapsto a + Aw$, where a is a vector in L, and A an invertible linear map from $\mathbb{R}^d$ to L.

Theorem 18.18 (Affine Spectral Decomposition). *Let* $\gamma^t = \gamma_1^t \otimes \cdots \otimes \gamma_m^t$, $t \in \mathbb{R}$, *be a one-parameter group of affine transformations on* $\mathbb{R}^d = \mathbb{R}^{d_1 + \cdots + d_m}$. *Write* $\gamma_k^t(w_k) = q_k(t) + Q_k^t w_k$ *for* $w_k \in \mathbb{R}^{d_k}$, *where* Q_k *is an invertible linear transformation on* $\mathbb{R}^{d_k}$. *We assume that the eigenvalues of* Q_k *all lie on the circle of radius* r_k *in* $\mathbb{C}$, *where* $0 < r_1 < \cdots < r_m$.

Let L be a d-dimensional linear space, and let $\alpha(t)$, $t \geq 0$, *be affine transformations:*

$$\alpha(t)(w) = a(t) + A(t)w, \quad \alpha(t)\colon \mathbb{R}^d \to L,$$

where $A(t)$ *are invertible linear maps, and* $a(t)$ *vectors in* L. *Assume*

$$\alpha(t_n)^{-1}\alpha(t_n + s_n) \to \gamma^s, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}.$$

Then one may write $L = L_1 + \cdots + L_m$ *as the sum of linear subspaces* L_k *of dimension* d_k, *and there exist invertible affine maps* $\beta_k(t)$, $t \geq 0$,

$$\beta_k(t)(w_k) = b_k(t) + B_k(t)w_k, \quad \beta_k(t)\colon \mathbb{R}^{d_k} \to L_k,$$

with $b_k(t) \in L_k$, *and* $B_k(t)$ *invertible linear maps, such that*

$$\beta(t) := \beta_1(t) \otimes \cdots \otimes \beta_m(t) \sim \alpha(t), \quad t \to \infty.$$

If L is an inner product space, one may choose the linear subspaces L_k *orthogonal. Suppose* $q_k = 0$ *for* $r_k \neq 1$. *For* $r_k > 1$ *the maps* $\beta_k(t)$ *may then be chosen linear,* $b_k(t) \equiv 0$, *and for* $r_k < 1$ *one may choose* $b_k(t) \equiv p_k$ *for some point* $p_k \in L_k$.

Proof. The curve $A\colon [0, \infty) \to \mathcal{L}(\mathbb{R}^d, L)$ varies like Q^t, $t \in \mathbb{R}$, where $Q = Q_1 \otimes \cdots \otimes Q_m$ is the linear part of γ. By Theorem 18.14 there exists a decomposition $L = L_1 + \cdots + L_m$ and $B_k\colon [0, \infty) \to \mathcal{L}(\mathbb{R}^{d_k}, L_k)$ such that $A(t) \sim B(t) = B_1(t) \otimes \cdots \otimes B_m(t)$ for $t \to \infty$. By the remark following the theorem we may choose the subspaces L_k to be orthogonal. Set $\bar{\beta}(t)(w) = a(t) + B(t)w$. Then $\bar{\beta}(t) \sim \alpha(t)$ since $\alpha(t)^{-1}\bar{\beta}(t)(w) = A(t)^{-1}B(t)w \to w$ for $t \to \infty$. Hence $\bar{\beta}$ varies like γ^t. Write $a(t) = a_1(t) + \cdots + a_m(t)$ with $a_k(t) \in L_k$, and $\bar{\beta}_k(t)(w_k) = a_k(t) + B_k(t)w_k$. Regular variation of $\bar{\beta}$,

$$\bar{\beta}(t_n)^{-1}\bar{\beta}(t_n + s_n)(w) \to \gamma^s(w), \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}$$

implies that $\bar{\beta}_k\colon [0,\infty) \to \mathcal{L}(\mathbb{R}^{d_k}, L_k)$ varies like γ_k^t for $k = 1,\dots,m$. Now suppose $q_k = 0$ for $r_k \neq 1$. If $r_k > 1$ then γ_k^t, $t \in \mathbb{R}$, is a linear expansion group, and $\bar{\beta}_k(t) \sim B_k(t)$ for $t \to \infty$ by Lemma 16.13; for $r_k < 1$ the eigenvalues of γ_k lie inside the unit circle in $\mathbb{C}$, and there exists a vector $p_k \in L_k$ such that $\bar{\beta}_k(t) \sim \beta_k(t)$ with $\beta_k(t)(w) = p_k + B_k(t)w$ by Lemma 15.15. □

It is possible to choose the origin in L such that the $\beta_k(t)$ are linear maps for $r_k \neq 1$. If the linear part of γ has no eigenvalues on the unit circle, then one may choose a new origin in $\mathbb{R}^d$ and in L such that γ^t is a group of linear transformations, and such that $\alpha(t) \sim A(t)$, where $A(t)$ is the linear part of $\alpha(t)$ in these coordinates. If γ has an eigenvalue on the unit circle in $\mathbb{C}$, it may not be possible to replace the affine transformations $\alpha(t)$ by linear transformations, even if γ is linear, as is shown by Example 17.28.

The SDT is a basic tool in the analysis of $\mathcal{D}^\infty(\rho)$. In MS the domain $\mathcal{D}^\infty(\rho)$ is characterized as the set of probability measures which vary regularly with exponent C, where C is the generator of the one-parameter group of symmetries of ρ.

We now turn to the domain $\mathcal{D}^h(\rho)$. Let us first show that the vertical coordinate η may be incorporated in a coordinate system such that the linear subspaces M_k in (18.17) are coordinate spaces. Let $e_1,\dots,e_d$ be a basis for L such that $e_{d_1+\dots+d_k+1},\dots,e_d$ is a basis of M_k for $k = 0,\dots,m-1$, and, equivalently, the coordinates $\xi_1,\dots,\xi_{d_1+\dots+d_k}$ span M_k^+.

Lemma 18.19. *Let $\eta = c_1\xi_1 + \dots + c_j\xi_j$ be a linear functional with $c_j \neq 0$. Then one may replace ξ_j above by η to obtain a new coordinate basis $\xi'_1,\dots,\xi'_d$. This basis also satisfies*

$$M_k^+ \text{ is spanned by } \xi'_1,\dots,\xi'_{d_1+\dots+d_k}, \quad 0 \le k < m.$$

Proof. For each i the elements $\xi_1,\dots,\xi_i$ and $\xi'_1,\dots,\xi'_i$ span the same space. □

Let $\pi \in \mathcal{D}^\infty(\rho)$, where $\gamma^t(\rho) = e^t\rho$, with $\rho^t = e^{tC}$ a linear expansion group. Let $\sigma_1 < \dots < \sigma_m$ be the distinct diagonal elements of the real Jordan form of C, with multiplicities $d_1,\dots,d_m$. Both z-space and w-space have a decomposition $\mathbb{R}^d = \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_m}$, such that

$$\gamma^t = \gamma_1^t \otimes \dots \otimes \gamma_m^t\colon \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_m} \to \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_m}$$
$$e^t\beta(t)^{-1}(\pi) \to \rho \text{ weakly on } \varepsilon B^c, \quad t \to \infty,\ \varepsilon > 0$$
$$\beta(t) = \beta_1(t) \otimes \dots \otimes \beta_m(t)\colon \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_m} \to \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_m},$$

where $\beta_k(t)$ varies like $\gamma_k^t = e^{tC_k}$ on $\mathbb{R}^{d_k}$. Here C_k is in real Jordan form, with all d_k diagonal entries equal to σ_k. Let π_k and ρ_k denote the corresponding projections of π and ρ on $\mathbb{R}^{d_k}$. Suppose the vector $Z = (X,Y) \in \mathbb{R}^{h+1}$ has distribution π.

Let F be the df of the vertical component Y. We are interested in conditions which ensure

1) $1 - F(y)$ varies regularly in ∞ with exponent $-\lambda < 0$;
2) $R(y) = F(-y) + 1 - F(y)$ varies regularly in ∞ with exponent $-\lambda < 0$;
3) $F(-y)/R(y) \to p \in [0, 1]$ for $y \to \infty$;
4) $Z \in \mathcal{D}^h(\rho)$.

By the lemma above we may choose the vertical coordinate η to be one of the coordinates in the decomposition above, $Y = \eta(Z)$, and $\eta = \xi_j$. With this coordinate is associated a rate of decrease σ_k. By Proposition 18.17

$$\begin{aligned} y^c R(y) &\to 0, \quad y \to \infty, \quad c < 1/\sigma_k, \\ &\to \infty, \quad y \to \infty, \quad c > 1/\sigma_k. \end{aligned}$$

Proposition 18.20. *If $C_k = \sigma_k I$ and ρ is full, then* 2) *holds with $\lambda = 1/\sigma_k$. If in addition $\rho_k\{\theta \geq 1\} > 0$ for every unit functional θ on $\mathbb{R}^{d_k}$, then* 1) *holds with $\lambda = 1/\sigma_k$. If $d_k = 1$ then $\beta_k(t)w_k = b_k(t)w_k$ for a non-zero real $b_k(t)$, and ρ_k is a measure on $\mathbb{R} \setminus \{0\}$. Assume $b_k(t)$ is positive eventually, and $\rho_k[1, \infty) > 0$. Then* 1)*,* 2)*,* 3) *and* 4) *hold.*

Proof. The first two statements follow from Proposition 17.12. The last statement holds since γ^t, $t \in \mathbb{R}$, is a one-parameter group in $\mathcal{A}^h$, $\beta(t) \in \mathcal{A}^h$, for $t \geq 0$, $\beta(t)$ varies like γ^t, and the limit measure ρ does not vanish on $\{v \geq 1\}$ßR^d. □

If the eigenvalues of γ^t all lie on the circle with radius r^t, the Jordan form may have entries below the diagonal, indicating a *shear* component, or blocks of size two, indicating a rotational component. The distribution tail of the vertical component Y, or of $|Y|$, need not vary regularly in these cases. However there exist bounds in terms of R-O variation, see MS, Theorem 6.1.32 and Definition 5.3.12.

The condition $(X, Y) \in \mathcal{D}^\infty$ implies that the convex hulls of the sample clouds may be normalized to converge, and $\Delta = \Gamma = \mathbb{R}^d$. In particular horizontal halfspaces which do not contain the origin are steady.

Example 18.21. Suppose $Z = (X, Y) \in \mathbb{R}^3$ lies in the domain $\mathcal{D}^h(\rho)$, where ρ is an excess measure with symmetries $\gamma^t : (u_1, u_2, v) \mapsto (e^{-t}u_1, e^t u_2, v + t)$. Let J_0 be a horizontal halfspace such that $\rho(J_0)$ is positive and finite. One may choose coordinates (Z_1, Z_2, Y) with

$$Z_1 = a_0 + a_1 X_1 + a_2 X_2 + aY, \quad Z_2 = b_1 X_1 + b_2 X_2 + bY$$

such that $\alpha(t)^{-1}(Z^{H(t)}) \Rightarrow W$ for $t \to \infty$ where W has distribution $1_{J_0} d\rho/\rho(J_0)$, and where the $\alpha(t) \in \mathcal{A}^h$ are diagonal,

$$\begin{aligned} &\alpha(t)(u_1, u_2, v) = (e^{-\varphi_1(t)}u_1, e^{\varphi_2(t)}u_2, b(t) + c(t)v), \\ &\dot{\varphi}_i(t) \to 0, \ \dot{b}(t) = c(t), \ \dot{c}(t)/c(t) \to 0. \end{aligned}$$

Here $H(t)$ are horizontal halfspaces $\{y \geq y(t)\}$ such that $\mathbb{P}\{Y \geq y(t)\} \sim e^{-t}$. $\Diamond$

18.5 Limit theory with regular variation. We now return to the basic limit relation:

$$\rho_t := e^t \alpha(t)^{-1}(\pi) \to \rho \text{ vaguely on } O, \quad t \to \infty. \tag{18.18}$$

Here π is a probability distribution on $\mathbb{R}^d$, and ρ a Radon measure on the open set O. We assume that O is maximal and that $\alpha\colon [0,\infty) \to \mathcal{A}$ varies like $\gamma^t = e^{tC}$.

We shall prove that the limit measure satisfies the basic symmetry relation $\gamma^t(\rho) = e^t\rho$, that O is invariant under γ^t, and that this also holds for the *intrusion cone* Δ and the *convergence cone* Γ, introduced in Section 5, and the corresponding sets of halfspaces. We begin with a simple lemma.

Lemma 18.22. *If $\rho_n \to \rho$ vaguely on O, and $\beta_n \to \beta$ in $\mathcal{A}$, then $\beta_n(\rho_n) \to \beta(\rho)$ on $\beta(O)$.*

Proof. Let ψ be a continuous function on $\beta(O)$ with compact support K. Let U be a neighbourhood of $\beta^{-1}(K)$, whose closure is a compact subset of O. Then $\varphi_n = \psi \circ \beta_n$ is a continuous function with support contained in U for $n \geq n_0$. Hence $\int \psi d\beta_n(\rho) = \int \varphi_n d\rho \to \int \varphi d\rho = \int \psi d\beta(\rho)$. $\square$

Proposition 18.23. *The limit measure is symmetric, and O is invariant.*

Proof. Let $s \in \mathbb{R}$. Let $r_n = t_n + s_n$ with $s_n \to s$ and $t_n \to \infty$. Then (18.18) gives

$$\alpha(t_n)^{-1}\alpha(t_n + s_n)\rho_{t_n+s_n} = e^{r_n}\alpha(t_n)^{-1}(\pi) = e^{s_n}\rho_{t_n}. \tag{18.19}$$

By regular variation $\alpha(t_n)^{-1}\alpha(r_n) \to \gamma^s$. Let $n \to \infty$ in (18.19) to obtain the symmetry:

$$\gamma^s(\rho) = e^s\rho, \quad s \in \mathbb{R}.$$

The left side converges on $\gamma^s(O)$. Hence so does the right side. By maximality $\gamma^s(O)\,\text{ß}\,O$. Similarly $\gamma^{-s}(O)\,\text{ß}\,O$, which gives $O\,\text{ß}\,\gamma^s(O)$. $\square$

A variation of this argument gives the powerful Extension Lemma.

Lemma 18.24 (Extension Lemma). *Let $\alpha\colon [0,\infty) \to \mathcal{A}$ vary like γ^t, and let ρ be a Radon measure on the open set $O\,\text{ß}\,\mathbb{R}^d$ such that $\gamma^t(\rho) = e^t\rho$ for $t \in \mathbb{R}$, and $\gamma^t(O) = O$. Let π be a probability measure on $\mathbb{R}^d$, and $U_0\,\text{ß}\,O$ open. If*

$$\rho_t := e^t\alpha(t)^{-1}(\pi) \to \rho \text{ vaguely on } U_0, \quad t \to \infty,$$

then vague convergence holds on $U = \bigcup_{s\in\mathbb{R}} \gamma^s(U_0)$.

Proof. In view of Proposition 5.16 it suffices to prove that vague convergence holds on $U_s = \gamma^{-s}(U_0)$ for each $s \in \mathbb{R}$. So let $\psi : \mathbb{R}^d \to [0,1]$ be continuous with compact support $K ß U_s$. Let $t_n \to \infty$, $s_n \to s$. First observe that $\gamma^s(K)$ is a compact set in U_0. Hence $\gamma^s(K) ß V ß K_1 ß U_0$ for some open set V and compact K_1. If $\delta_n \to \gamma^s$ then $\delta_n(K) ß V$ eventually, $\varphi_n = \psi \circ \delta_n \to \psi \circ \gamma^s$, and $0 \le \varphi_n \le 1_{K_1}$ for $n \ge n_0$ implies

$$\int \varphi_n d\rho_{t_n+s_n} \to \int \psi \circ \gamma^s d\rho. \tag{18.20}$$

Now observe that (18.19) holds. For $\delta_n = \alpha(t_n)^{-1}\alpha(t_n + s_n)$ this gives

$$\int \varphi_n d\rho_{t_n+s_n} = e^{s_n} \int \psi d\rho_{t_n}.$$

The limit (18.20) then yields convergence to $\int \psi \circ \gamma^s(d\rho) = e^s \int \psi d\rho$. □

We do not need the full force of (18.18). Suppose this limit relation only holds for a sequence which diverges to infinity. If the sequence contains the integers $n!$ and $n! + 1$ then (18.19) with $t_n = n!$ and $r_n = n! + 1$ implies that $\gamma(\rho) = e\rho$. If the sequence also contains the numbers $n! + \log 2$ then $\gamma^{\log 2}(\rho) = 2\rho$. It follows by Theorem 18.28 that $\gamma^s(\rho) = e^s\rho$ for all s in the closed additive group generated by 1 and $\log 2$. Since $\log 2$ is irrational this equation then holds for all $s \in \mathbb{R}$, and hence ρ is an excess measure.

Proposition 18.25. *If (18.18) holds for a sequence $t_n \to \infty$, and $t_{n+1} - t_n$ is bounded, and if $\gamma^s(\rho) = e^s\rho$ for all $s \in \mathbb{R}$, then (18.18) holds for $t \to \infty$.*

Proof. Let $r_n \to \infty$. We shall show that (18.18) holds for a subsequence of (r_n). So assume $r_n = t_{k_n} + s_n$ where $s_n \to s \ge 0$. Use (18.19) to write $\rho_{r_n} = e^{s_n}\alpha(r_n)^{-1}\alpha(t_{k_n})\rho_{t_{k_n}}$. Lemma 18.22, applied to the sequence $\alpha(r_n)^{-1}\alpha(t_{k_n}) \to \gamma^{-s}$, allows us to take the limit on the right. Hence $\rho_{r_n} \to e^s\gamma^{-s}(\rho)$. By the symmetry assumption the right side equals ρ. □

If the halfspace $J_0 ß O$ has finite measure, then so has $\gamma^s(J_0)$ for all $s \in \mathbb{R}$. Thus the set $\mathcal{H}_0$ of all halfspaces $J ß O$ of finite mass is invariant under the affine transformations γ^t, $t \in \mathbb{R}$. Recall that a halfspace $J_0 ß O$ is sturdy if J_0 is interior point of $\mathcal{H}_0$; equivalently for any sequence $J_n \to J_0$, eventually $J_n ß O$ and $\rho(J_n) < \infty$. The directions of sturdy halfspaces form an open cone Δ in the dual space, the *intrusion cone*. Let $\gamma^s(w) = a(s) + A(s)w$. Then γ^{-s} maps the halfspace $H = \{\xi \ge c\}$ into $\gamma^{-s}(H) = \{\xi A(s) \ge c - \xi a(s)\}$ since

$$z \in \gamma^{-s}(H) \iff \gamma^s(z) \in H \iff \xi A(s)z + \xi a(s) \ge c.$$

Both the intrusion cone, Δ, and $\mathcal{H}^{\Delta}$, the set of sturdy halfspaces, are invariant under γ^t, $t \in \mathbb{R}$.

Example 18.26. Suppose Δ contains the functional $\xi_0 = (1, 1, 1)$. If the γ^t are scalar expansions all one can say is that Δ contains all functionals ξ sufficiently close to ξ_0 since the intrusion cone is open. If $\gamma^t = (e^t, e^{2t}, e^{3t})$ then Δ contains the orbit $\{(r, r^2, r^3) \mid r > 0\}$ and hence by convexity all $\xi = (a, b, c) \in (0, \infty)^3$ with $b^2 < ac$. The same holds for the convergence cone. ♢

Proposition 18.27. *Suppose* (18.18) *holds. The convergence cone* Γ*, and the set* $\mathcal{H}^\Gamma$ *are invariant under the transformations* γ^t.

Proof. Let $s_n \to s$, $t_n \to \infty$, and set $r_n = t_n + s_n$, and $\beta_n = \alpha(t_n)^{-1}\alpha(r_n)$. Then $\beta_n \to \gamma^s$, and $\beta_n(\rho_{r_n}) = e^{s_n}\rho_{t_n}$ by (18.19). Now suppose $J \in \mathcal{H}^\Gamma$ and $\rho(\partial J) = 0$. Then $\rho_{t_n}(J_n) \to \rho(J)$ for any sequence $J_n \to J$. Choose J_n such that $\beta_n^{-1}(J_n) = \gamma^{-s}(J)$. Then

$$\rho_{r_n}(\gamma^{-s}(J)) = \rho_{r_n}(\beta_n^{-1}(J_n)) = e^{s_n}\rho_{t_n}(J_n) \to e^s\rho(J) = \rho(\gamma^{-s}(J)).$$

Hence $\rho_t(\gamma^{-s}(J)) \to \rho(\gamma^{-s}(J))$ for $t \to \infty$ holds for any $J \in \mathcal{H}^\Gamma$ for which $\rho(\partial J) = 0$. Since $\gamma^{-s}(\mathcal{H}^\Gamma)$ is open in $\mathcal{H}$ it follows that $\gamma^{-s}(\mathcal{H}^\Gamma)\subset\mathcal{H}^\Gamma$ by maximality of $\mathcal{H}^\Gamma$. Equality follows as above. □

18.6 Symmetries. We shall now embark on an investigation of excess measures.

For a measure ρ on $\mathbb{R}^d$ there is a maximal open set O on which ρ is a Radon measure. The set O consists of all points $z \in \mathbb{R}^d$ which have a neighbourhood on which ρ is finite. If α is a symmetry of ρ then $\alpha(O) = O$. Recall that α is a symmetry of ρ if there is a constant $c(\alpha) > 0$ such that $\alpha(\rho) = c(\alpha)\rho$.

Theorem 18.28. *Let* ρ *be a Radon measure on an open set* $O\subset\mathbb{R}^d$*. The set* $\mathcal{G}$ *of symmetries of* ρ *is a closed subgroup of* $\mathcal{A}$*. For* $\alpha \in \mathcal{G}$ *let* $c(\alpha) > 0$ *be the constant such that* $\alpha(\rho) = c(\alpha)\rho$*. The map* $c\colon \mathcal{G} \to (0, \infty)$ *is a continuous homomorphism into the multiplicative group of positive reals.*

Proof. It is clear that α^{-1} is a symmetry if α is, and that $\alpha\beta$ is a symmetry if α and β are symmetries. So $\mathcal{G}$ is a group. It is also clear that $c(\alpha^{-1}) = 1/c(\alpha)$ and $c(\alpha\beta) = c(\alpha)c(\beta)$. It remains to show that $\mathcal{G}$ is closed and c continuous. So let $\alpha_n \in \mathcal{G}$ converge to $\alpha_0 \in \mathcal{A}$. We shall assume that the open set O is maximal. We claim that $\alpha_0(O) = O$. Above we saw that $\alpha_n(O) = O$ by maximality of O. Let $U = w + rB$ be a small open ball in O. The images $E_n = \alpha_n(U)$ are open ellipsoids, and E_0 is contained in the union $E_1 \cup E_2 \cup \dots$. Hence $E_0\subset O$. This proves $\alpha_0(O)\subset O$. Similarly $\alpha_0^{-1}(O)\subset O$. Together this gives $\alpha_0(O) = O$. Now let $\varphi \geq 0$ be continuous with compact support K contained in O, and $\int \varphi d\rho = q > 0$. Let $K_1\subset O$ be a compact neighbourhood of K. Then

$$c(\alpha_n)\int \varphi d\rho = \int \varphi d\alpha_n(\rho) = \int \varphi \circ \alpha_n^{-1} d\rho \to \int \varphi \circ \alpha_0^{-1} d\rho = \int \varphi d\alpha_0(\rho).$$

The sequence $c(\alpha_n)$ is bounded eventually (by $\rho(\alpha_0(K_1)) \max |\varphi|/q$), and hence has a finite limit. This also holds for the sequence $c(\alpha_n^{-1}) = 1/c(\alpha_n)$. So the limit $c(\alpha_0)$ is positive, and does not depend on φ. This shows that α_0 is a symmetry. Hence $\mathscr{G}$ is closed. The relation $c(\alpha_n) \to c(\alpha_0)$ shows that c is continuous. □

The set $\mathscr{S}$ of measure preserving symmetries is a closed normal subgroup of $\mathscr{G}$ since it is the kernel of the continuous homomorphism $\alpha \mapsto c(\alpha)$. If ρ is an excess measure, and $\gamma^t(\rho) = e^t\rho$ for $t \in \mathbb{R}$, then any symmetry α has the form

$$\alpha = \gamma^t\sigma = \sigma'\gamma^t, \quad t = \log c(\alpha),\ \sigma, \sigma' \in \mathscr{S}.$$

The one-parameter group γ^t, $t \in \mathbb{R}$, is not unique, unless the group $\mathscr{S}$ of measure preserving symmetries is discrete. We need some elementary *Lie group* theory to understand the situation.

A one-parameter group $\alpha^t = e^{tA}$ is determined by its generator,

$$A = \lim_{t \to 0} \frac{\alpha^t - \mathrm{id}}{t},$$

where we use the representation (2) in the Preview if the transformations α^t are non-linear. Let $\mathfrak{g}$ be the set of generators A of all one-parameter groups $\alpha^t = e^{tA}$ for which there exists a constant $c = c(\alpha) > 0$ such that $\alpha^t(\rho) = c^t\rho$ for $t \in \mathbb{R}$. The set $\mathfrak{g}$ is a linear space. It is called the *Lie algebra* of the group $\mathscr{G}$. Actually $\mathfrak{g}$ is determined by the connected component of the group $\mathscr{G}$ which contains the identity. The component of the identity is a normal subgroup, $\mathscr{G}_0$, which is both open and closed in $\mathscr{G}$. Any sequence α_n in $\mathscr{G}$ which converges to the identity will lie in $\mathscr{G}_0$ eventually, and has the form $\alpha_n = e^{A_n}$, $n \geq n_0$, for a null-sequence A_n in $\mathfrak{g}$. Section 18.13 gives more details.

The homomorphism $\alpha \mapsto c(\alpha)$ on $\mathscr{G}$ corresponds to a linear functional θ on $\mathfrak{g}$:

$$c(e^A) = e^{\theta(A)}, \quad A \in \mathfrak{g}.$$

The kernel $\mathfrak{s} = \{\theta = 0\}$ß$\mathfrak{g}$ of θ is the Lie algebra of the group $\mathscr{S}$ of measure preserving transformations. The $\gamma \in \mathscr{G}$ which satisfy (9) in the Preview are determined by a hyperplane in the Lie algebra $\mathfrak{g}$:

$$\gamma^t = e^{tC}, \quad C \in \{\theta = 1\}ß\mathfrak{g}.$$

Example 18.29. If ρ is *Lebesgue measure* on $(0, \infty)^3$ then $\mathscr{G}_0$ is the group of positive diagonal matrices, and $\mathscr{G}$ is the group of matrices of size three with one positive entry in each row and column, and six zeros. There is a discrete group of six measure preserving transformations, which permute the axes. The group $\mathscr{S}$ is not compact. The Lie algebra $\mathfrak{g}$ of $\mathscr{G}$ consists of all diagonal matrices; $-\theta$ is the trace. So C is a generator for ρ if C is diagonal, and the diagonal sum is -1. ◊

Let C generate the group γ^t which satisfies $\gamma^t(\rho) = e^t\rho$, and let $\sigma = e^S$ be a measure preserving transformation. The three transformations

$$e^{S+C},\ \sigma\gamma,\ \gamma\sigma; \quad \sigma = e^S,\ \gamma = e^C,\ S \in \{\theta = 0\},\ C \in \{\theta = 1\}$$

map ρ into $e\rho$. In non-commutative groups the three may be distinct. Meerschaert & Veeh [1993] give a thorough analysis of symmetry groups for excess measures associated with linear expansion groups.

Definition. An excess measure ρ is an *XS-measure* if the group of symmetries $\mathcal{G}$ is so large that $\mathcal{G}(J_0)$ is a neighbourhood of J_0, and if ρ lives on an orbit of the component of the identity of $\mathcal{G}$.

Examples are the multivariate GPDs, but also *Lebesgue measure* on $(0,\infty)^d$, or on a cone $\{v > \|u\|\}$. Balkema [2006] contains further examples. Such XS-measures ρ have the tail property of the univariate exponential distribution in an excessive degree. For all halfspaces J close to J_0 the high risk scenarios $d\rho^J = 1_J d\rho/\rho(J)$ have the same shape.

18.7* Invariant sets and hyperplanes. With an excess measure are associated an open set Oß$\mathbb{R}^d$, a one-parameter group of affine transformations $\gamma^t = e^{tC}$, and a halfspace J_0 of finite positive mass. Here we shall prove that $\rho(\partial J_0) = 0$. We first show that invariant sets have mass zero or infinite.

Definition. A set E is *invariant* under γ^t, $t \in \mathbb{R}$, if

$$\gamma^t(E) = E, \quad t \in \mathbb{R}. \tag{18.21}$$

Any invariant set is a union of *orbits* $\Gamma_z = \{\gamma^t(z) \mid t \in \mathbb{R}\}$.

Proposition 18.30. *If E is invariant, then so is its closure* cl(E)*, its interior* int(E)*, the convex hull, the cone, and the affine subspace generated by E.*

Proof. Suppose $z_n \to z \in \mathrm{cl}(E)$. Then $\gamma^t(z_n) \to \gamma^t(z)$. Hence $\gamma^t(z) \in \mathrm{cl}(E)$. If UßE is an open neighbourhood of z then $\gamma^t(U)$ßE is an open neighbourhood of $\gamma^t(z)$. If $z = p_0z_0 + \cdots + p_mz_m$ with $p_i \geq 0$, $p_0 + \cdots + p_m = 1$, and $z_i \in E$, then $\gamma^t(z) = p_0\gamma^t(z_0) + \cdots + p_m\gamma^t(z_m)$. Similar arguments show that the cone and the affine hull generated by E are invariant. □

Invariant Borel sets have mass zero or infinite since

$$\rho(E) = \rho(\gamma^t(E)) = e^{-t}\rho(E).$$

We shall now derive a similar result for affine subspaces. In particular this ensures that $\rho(\partial J_0) = 0$ for any halfspace J_0 of finite mass.

First we prove that any affine subspace contains a maximal invariant affine subspace.

Lemma 18.31. *For a given affine subspace A let A_0 be the union of all orbits $\Gamma \text{ß} A$. Then A_0 is an affine subspace. It may be empty.*

Proof. If z_1 and z_2 are distinct points in A_0, and z is a point on the line L through z_1 and z_2 then $\gamma^t(z)$ is a point on the line $\gamma^t(L)$ through $\gamma^t(z_1)$ and $\gamma^t(z_2)$. Hence A_0 is closed for lines. □

Proposition 18.32. *Let ρ be a σ-finite measure on $\mathbb{R}^d$, and γ^t, $t \in \mathbb{R}$, a one-parameter group of affine transformations such that $\gamma^t(\rho) = e^t \rho$ for $t \in \mathbb{R}$. Then any affine subspace A contains an invariant affine subspace A_0 such that $\rho(A \setminus A_0) = 0$.*

Proof. Let A_0 be maximal. For $z \in A \setminus A_0$ there exists $t > 0$ such that $\gamma^s(z) \notin A$ for $0 < |s| < t$. Choose $t = t(z)$ maximal. We shall see below that $z \mapsto t(z)$ is measurable. Hence $A_n = \{t > 1/n\}$ is a Borel set. Suppose $\rho(A_m) = c > 0$ for some $m \geq 1$. The sets $\gamma^t(A_m)$, $0 \leq t < 1/m$, are disjoint Borel sets, each with mass $\rho(\gamma^t(A_m)) > c/e$. This is not possible for a finite or σ-finite measure. □

For any $E \text{ß} \mathbb{R}^d$ define

$$t_E(z) = \inf\{s > 0 \mid \gamma^s(z) - z \in E\}, \quad z \in \mathbb{R}^d.$$

The function $z \mapsto t_E(z)$ is a first entrance time and is universally measurable if E is a Borel set. We only need a special case.

Lemma 18.33. *If E is open or closed then $z \mapsto t_E(z)$ is measurable.*

Proof. If E is open one may restrict s in the definition of t_E to the positive rationals and t_E is the infimum of functions $q 1_{E_q}$ with $q > 0$ rational. If $E \text{ß} \mathbb{R}^d \setminus \{0\}$ is compact then $t_E = \sup t_{E_n}$ where E_n is the open $1/n$-neighbourhood of E. If $E \text{ß} \mathbb{R}^d \setminus \{0\}$ is closed then $t_E = \inf t_{E_n}$ where $E_n = E \cap \{1/n \leq \|z\| \leq n\}$ is compact. Finally observe that t_E is measurable for $E = \{0\}$ since by the Jordan representation of the generator C for $s > 0$

$$\gamma^s(z) = z \iff z \in E_{\lambda_1} + \cdots + E_{\lambda_m}$$

where E_λ is the eigenspace of C for the eigenvalue $\lambda \in \mathbb{C}$, and $\lambda_1, \ldots, \lambda_m$ are the eigenvalues on the imaginary axis for which $\lambda s / 2\pi i$ is an integer. □

Corollary 18.34. *Excess measures do not charge the boundary of halfspaces of finite mass.*

18.8 Excess measures on the plane. One can explicitly write down all excess measures on $\mathbb{R}^2$, and with some extra effort, on $\mathbb{R}^d$, for $d > 2$. First choose a symmetry group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, of affine transformations on the plane. The generator C may be represented by a matrix C of size 3 with top row zero. If the group contains no translation component, one may choose the origin so that the transformations γ^t are linear, and C is a matrix of size 2. We give the matrices in their *Jordan form*. There are two classes of translations (of order one and two), and four classes of linear transformations: diagonal, scalar, *shear*, and rotation. The generators are

$$\begin{pmatrix}0&0&0\\1&0&0\\0&0&\lambda\end{pmatrix}, \quad \begin{pmatrix}0&0&0\\1&0&0\\0&1&0\end{pmatrix}, \quad \begin{pmatrix}\lambda&0\\0&\mu\end{pmatrix}, \quad \begin{pmatrix}\lambda&0\\0&\lambda\end{pmatrix}, \quad \begin{pmatrix}\lambda&0\\1&\lambda\end{pmatrix}, \quad \begin{pmatrix}\lambda&\tau\\-\tau&\lambda\end{pmatrix}.$$

The parameters λ, μ, τ are real, $\mu < \lambda$, and $\tau \neq 0$. Depending on the sign of λ and μ there are $3 + 1 + 5 + 3 + 3 + 3 = 18$ qualitatively different groups.

Invariant sets have mass zero or infinite, as we saw above. Compact invariant sets have measure zero since ρ is a Radon measure. It follows that there is no excess measure for pure rotations ($\lambda = 0$), or for the trivial group: scalar transformations with $\lambda = 0$. (All disks have measure zero.)

The simplest invariant sets are orbits:

$$\Gamma_z = \{\gamma^t(z) \mid t \in \mathbb{R}\}.$$

Many orbits are graphs. They are of the form

$$x = y^2/2, \quad x = e^{\lambda y}, \quad y = x^{\mu/\lambda}, \quad y = cx, \quad y = x \log x.$$

Only the spiral orbits associated with rotations do not have this simple structure. The image of the measure $e^{-t}dt$ on $\mathbb{R}$ under the map $t \mapsto \gamma^t(z)$ is an elementary excess measure ρ_z on the orbit Γ_z which satisfies $\gamma^t(\rho_z) = e^t\rho_z$, at least if the orbit is not compact. If the orbit is curved then ρ_z is full. Only for pure translations ($\lambda = 0$), pure shears ($\lambda = 0$), and scalar transformations, there are no curved orbits, and the support of a full excess measure has to contain at least two orbits.

For rotations with $\lambda < 0$, the orbits are spirals, and the complement of any disk rB has infinite measure for ρ_z. All halfplanes have infinite measure. This yields a third group for which there is no excess measure. For the remaining 15 groups excess measures exist.

In these lectures we have concentrated on two kinds of excess measures. Those for which there exists a halfspace J_0 of finite measure such that $\gamma^t(J_0)ßJ_0$ for $t > 0$; and those which are symmetric for expansions (linear groups whose generator have only eigenvalues with positive real part). Of the fifteen one-parameter groups which admit an excess measure only the group of pure shears, $\lambda = 0$, is not of one of these two kinds.

Example 18.35 (Pure *shears*: $\gamma^t(u,v) = (u, v + tu)$). Orbits are the vertical lines $\{u = c\}$ with $c \neq 0$, and the points on the vertical axis. Any excess measure is a mixture of excess measures ρ_c on the orbit through $(c, 0)$, $c \neq 0$. For $c > 0$ the excess measure ρ_c is normalized by $\rho_c\{v \geq 0\} = 1$. Then

$$\rho_c(J_1) = e^{-1/c}, \quad J_1 = \{v \geq 1\},$$

and in general $\rho_c\{v \geq b\} = e^{-b/c}$, so that we have $\rho_c(H) = \infty$ for any halfplane $H = \{v \leq b + au\}$. Similarly excess measures with mass to the left of the vertical axis give infinite mass to $v \geq b + au$. To have a halfplane with finite positive mass the excess measure has to live on one side of the vertical axis. We assume that it lives on the right. Then one may write

$$\rho = \int \rho_c d\mu(c)$$

for some measure μ on $(0, \infty)$. Note that ρ has a density if μ has:

$$d\mu(u) = m(u)du \iff d\rho(u,v) = e^{-v/u} u m(u) du dv.$$

(For $m = 1_{(a,b)}$ with $0 < a < b$ both sides give mass $(b-a)e^{-c}$ to $\{v \geq cu\}$.) In general

$$M(s) := \rho(J_s) = \int e^{-s/c} d\mu(c), \quad J_s = \{v \geq s\}.$$

We see that M is a *Laplace transform* (of the image $\tilde{\mu}$ of μ under the map $c \mapsto 1/c$). In particular there exists $s_0 \in [-\infty, \infty]$, the abscissa of convergence, such that the integral is finite for $s > s_0$ and infinite for $s < s_0$. If s_0 is finite one may choose the origin in $(0, s_0)$. The function M then is analytic on $\Re > 0$, and 0 is a singular point. The measure ρ is a Radon measure on $O = \mathbb{R}^2 \setminus \{(0,v) \mid v \leq 0\}$. All halfplanes $H ß O$ have finite mass. If $s_0 = \infty$ there are no halfplanes of finite mass; if $s_0 = -\infty$ then ρ is a Radon measure on $\mathbb{R}^2$, all halfplanes $v \geq au + b$ with $a, b \in \mathbb{R}$ have finite mass, and are sturdy. ◊

Given the one-parameter group one would like to couple mixtures of elementary excess measures to halfspaces of finite mass. If $\rho(H_0)$ is finite then so is $\rho(H_t)$ for $H_t = \gamma^t(H)$. Then ρ is finite for all halfspaces contained in a finite union $H_{t_1} \cup \cdots \cup H_{t_n}$, and also for all halfspaces $H ß O$ with direction $\theta \in \Delta$. For pure shears, and excess measures on the right halfplane, there is only one such family of halfspaces, $H = \{y \geq c + bx\}$, $b \in \mathbb{R}$, $c > c_0$, where c_0 is the abscissa of convergence. In general the situation is less simple.

Example 18.36 (Hyperbolic Orbits). Let ρ be an excess measure which lives on $(0,\infty)^2$ with symmetries $\gamma^t : (u,v) \mapsto (e^{-t}u, e^t v)$. The generator is $C = \mathrm{diag}(-1, 1)$. The orbits in $(0, \infty)^2$ are hyperbolas. Let ρ_c be the measure on $\Gamma_c = \{uv = c\}$

which projects onto Lebesgue measure on the positive horizontal axis. If $\rho = \rho_c$ then $O = \mathbb{R}^2$ and the *intrusion cone* Δ is the open left halfplane.

In general the excess measure ρ on $(0, \infty)^2$ is a mixture of elementary measures: $\rho = \int \rho_c d\mu(c)$ for a measure μ on $(0, \infty)$. There are four classes of halfplanes, represented by

$$\{u \le 1\}, \quad \{v \ge 1\}, \quad \{v \ge u + 1\}, \quad \{u + v \le 1\}.$$

A simple computation gives $\rho\{u \le 1\} = \mu(0, \infty)$, $\rho\{v \ge 1\} = \int c d\mu(c)$, and $\rho\{v \ge u + 1\}$ is finite if and only if $\int (c \wedge \sqrt{c}) d\mu(c)$ is finite. The fourth class of halfplanes, $H_{ab} = \{au + bv \le 1\}$ with $a, b > 0$, is different. If μ is a Radon measure on $(0, \infty)$ and $\mu(0, 1)$ is finite, then ρ is a Radon measure on $\mathbb{R}^2$ and $\rho(H_{ab})$ is finite. However μ need not be a Radon measure on $(0, \infty)$. Suppose μ has density $1/|c - 1|$. Then $\rho(U) = \infty$ for every open set U which intersects the orbit Γ_1. Now ρ is a Radon measure on $O = \mathbb{R}^2 \setminus \Gamma_1$. Halfplanes H_{ab}ßO satisfy $ab > 1/4$. Halfplanes of finite mass cannot see the measure ρ inside the convex region bounded by the hyperbola Γ_1. So halfplanes cannot distinguish measures which agree with ρ outside this convex region. Moreover the halfplane HßO cuts off compact sets from the orbits Γ_c, $0 < c < 1$. As t varies, the halfplane $\gamma^t(H)$ will see both more and less of the orbits it intersects. The basic assumption of the recipe still applies in this situation. The mean measure $\rho(\gamma^t(H)) = e^{-t}\rho(H)$ increases as t decreases. ◊

18.9 Orbits. We now turn to excess measures on open subsets of $\mathbb{R}^d$. Understanding the structure of the orbits of the one-parameter group γ^t in $\mathcal{A}(d)$ makes it possible to describe excess measures as mixtures of elementary excess measures on orbits, thus revealing the product structure. One factor is the exponential measure on $\mathbb{R}$, the other a spectral measure on a disjoint union of affine and quadratic subsets. We shall characterize the one-parameter groups for which excess measures exist.

There are two kinds of affine one-parameter groups, those which contain a translation component, and linear groups. The generator C is a matrix of size $1 + d$ with top row zero. We assume that it has *Jordan form*. If $C_{10} = 1$ then γ^t maps the horizontal coordinate plane $\{\xi_1 = 0\}$ into the horizontal hyperplane $\{\xi_1 = t\}$. If $C_{10} = 0$ then $\gamma^t = e^{tC}$ are linear transformations, and one may take C to have size d.

If there is a translation component the map

$$\Phi\colon (u, t) \mapsto \gamma^t(u, 0), \quad \mathbb{R}^h \times \mathbb{R} \to L = \mathbb{R}^d$$

is a homeomorphism. It maps vertical lines into orbits. A Radon measure ρ on an open invariant set OßL which satisfies $\gamma^t(\rho) = e^t\rho$ is the image under Φ of a product measure $\rho^*(du)e^{-t}dt$ on $\mathbb{R}^{h+1}$. If $\rho(H_+) = 1$ then the spectral measure ρ^* is a probability measure.

Henceforth we assume that $\gamma^t = e^{tC}$ are linear transformations on the linear space $L = \mathbb{R}^d$, and that C is a matrix of size d in complex Jordan form. We shall identify complex numbers with submatrices of size two in the usual way, see (18.25). For unbounded orbits $\Gamma_p = \{\gamma^t(p) \mid t \in \mathbb{R}\}$ there is a measure ρ_p on Γ_p which satisfies $\gamma^t(\rho_p) = e^t \rho_p$, and $\rho_p\{\gamma^t(p) \mid t \geq 0\} = 1$. This measure is the image of the measure $e^{-t}dt$ on $\mathbb{R}$ under the map $t \mapsto \gamma^t(p)$. We claim that it is a Radon measure on the open set $L \setminus L_{00}$, where L_{00} is the linear space containing the bounded orbits. We shall prove that any excess measure is a mixture of such elementary excess measures ρ_p.

One might try to write excess measures as mixtures of such elementary measures with regard to a mixing measure on the space U of unbounded orbits, the quotient space

$$U = (L \setminus L_{00})/\sim.$$

Here the points p and q are equivalent, $p \sim q$, if both lie on the same orbit. Open sets in U correspond to invariant open sets in $L \setminus L_{00}$. Unfortunately the quotient space U need not be Hausdorff.

Example 18.37. Let $C = \mathrm{diag}(-1, 1)$. Orbits are hyperbolas or half-axes. The two positive half-axes do not have disjoint invariant neighbourhoods. ◊

Let us now first see what orbits look like. Orbits, or rather the functions $\gamma^t(z)$, $t \in \mathbb{R}$, are solutions of linear ODEs, $\dot{x} = Cx$, see Hirsch & Smale [1974]. We are interested in the sets Γ_z, rather than the dynamics. The generators C and rC with $r > 0$ give the same orbits, traversed at different speeds; the orbits of $-C$ are the orbits of C traversed in the inverse direction. The matrix C in complex Jordan form falls apart into blocks $C_\lambda^m = J + \lambda I$ of size m with λ's on the diagonal and ones below. The matrices e^{tC} have the same decomposition with blocks $e^{tC_\lambda^m} = e^{\lambda t} e^{tJ}$. For $m = 4$ the matrices are

$$e^{tJ} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ t & 1 & 0 & 0 \\ t^2/2 & t & 1 & 0 \\ t^3/3! & t^2/2 & t & 1 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{18.22}$$

The behaviour of the orbit is determined by its projections on the corresponding m-dimensional invariant subspaces. If λ is not real we identify the m-dimensional complex subspace with a $2m$-dimensional real subspace in the usual way.

In order to get insight in the behaviour of orbits, we write

$$\begin{aligned} L = \mathbb{R}^d &= \mathbb{R}^{d_r} \oplus \mathbb{C}^{d_c} = L_+ \oplus L_0 \oplus L_- \\ p &= (x, z) = (p_+, p_0, p_-) \\ d &= d_r + 2d_c = d_+ + d_0 + d_-, \end{aligned}$$

where d_+, d_0 and d_- are the real dimensions of the invariant subspaces L_+, L_0 and L_- corresponding to eigenvalues in $\Re > 0$, $\Re = 0$ and $\Re < 0$.

Suppose $C = C_\lambda^m = J + \lambda I$ as above. First assume λ lies on the imaginary axis. The orbit Γ_p diverges in both directions, $\|\gamma^t(p)\| \to \infty$ for $t \to \infty$ and for $t \to -\infty$, unless p lies on the vertical axis, in which case Γ_p is a point or a circle. If $\Re\lambda > 0$ then $\gamma^t(p)$ runs from the origin to infinity for $p \neq 0$ as t increases; if $\Re\lambda < 0$, it runs from infinity to the origin.

The decomposition of C into blocks C_λ^m makes it possible to describe the orbits. The orbit Γ_p is bounded if and only if $p \in L_{00}$, where L_{00}ßL_0 is the subspace spanned by the eigenvectors with eigenvalues on the imaginary axis. If the orbit is unbounded then $\|\gamma^t(p)\| \to \infty$ for $t \to \infty$ or for $t \to -\infty$. The orbit diverges in both directions if and only if p_+ and p_- both are non-zero, or if $p_0 \in L_0 \setminus L_{00}$. Now suppose $\{\gamma^t(p) \mid t \to \infty\}$ is bounded, and $\|\gamma^t(p)\| \to \infty$ for $t \to -\infty$. Then $p_+ = 0$, $p_- \neq 0$, and $p_0 \in L_{00}$. If moreover $p_0 \in \ker(C)$ then $\gamma^t(p)$ converges for $t \to \infty$; otherwise the limit points form a torus in L_{00}, the closure of Γ_{p_0}.

If $\gamma^t(z) = z$ for some $t \neq 0$, but not for all, then the orbit is (homeomorphic to) a circle, if $\gamma^t(z) \neq z$ for all $t \neq 0$ the orbit is (homeomorphic to) a line. An orbit may be bounded but not compact. Think of a line wound around a torus. The set of fix points is a linear space. It is the kernel of the generator C. From the description above we see:

Proposition 18.38. *The union of the bounded orbits is a linear subspace L_{00}. This subspace L_{00} is spanned by the eigenvectors whose eigenvalues lie on the imaginary axis.*

Proposition 18.39. *Unbounded orbits are closed in $L \setminus L_{00}$.*

Proposition 18.40. *The following statements are equivalent:*

1) *All orbits are bounded;*

2) *C is diagonal with entries on the imaginary axis;*

3) *if ρ is a Radon measure on an open set and $\gamma^t(\rho) = e^t\rho$, $t \in \mathbb{R}$, then $\rho = 0$.*

Proof. In view of the discussion above it suffices to show that symmetric Radon measures ρ on an open set OßL vanish. Let K be a compact subset of O. Choose an open neighbourhood of K whose closure is a compact subset K_1 of O. Then $\rho(K_1) = c_1$ is finite. By 2) the γ^t lie in the compact group O of orthogonal transformations. Hence there is a sequence $t_n \to \infty$ such that $\gamma^{t_n} \to I$. It follows that $\gamma^{-t_n}(K)$ßK_1 eventually. Hence

$$e^{t_n}\rho(K) = \rho(\gamma^{-t_n}(K)) \leq \rho(K_1) = c_1, \quad n \geq n_0.$$

We conclude that $\rho(K) = 0$. □

If ρ is a Radon measure on L such that $\gamma^t(\rho) = e^t\rho$ for $t \in \mathbb{R}$, then the restriction of ρ to any invariant linear subspace will have the same properties. In particular the restriction to L_{00} is the zero measure. We shall now give a decomposition of L (and ρ) by an increasing sequence of linear subspaces which allows us to write *excess measures* ρ as mixtures of elementary measures ρ_z somewhat in the same way this was done when γ^t had a translation component.

Proposition 18.41. *There is an increasing sequence* $L_{00} = M_0 \text{ß} \ldots \text{ß} M_m = L$ *of invariant linear subspaces, and a set* $S \text{ß} L \setminus M_0$ *with the following properties for* $k = 1, \ldots, m$*:*

1) *the set* $S_k := S \cap (M_k \setminus M_{k-1})$ *is closed in* $M_k \setminus M_{k-1}$*;*

2) *the map* $\Phi_k : (z, t) \mapsto \gamma^t(z)$ *is a homeomorphism from* $S_k \times \mathbb{R}$ *onto* $M_k \setminus M_{k-1}$*;*

3) *if* ρ *is a Radon measure on the open set* $O \text{ß} L$ *such that* $\gamma^t(\rho) = e^t \rho$ *for* $t \in \mathbb{R}$*, then* $\rho(M_0) = 0$*, and there exist Radon measures* ρ_k^* *on* $O \cap S_k$ *such that*

$$\Phi_k(d\rho_k^* e^{-t} dt) = 1_{M_k \setminus M_{k-1}} d\rho, \quad k = 1, \ldots, m.$$

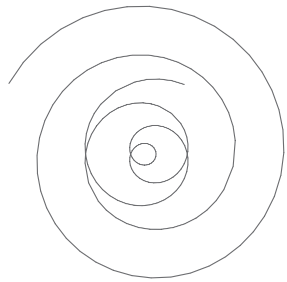
Proof. The construction of the subspaces proceeds by induction, starting with M_{m-1}. If there are unbounded orbits, there is a proper invariant linear subspace $M \text{ß} L = \mathbb{R}^d$, and a closed subset S of $L \setminus M$ such that all orbits in $L \setminus M$ are unbounded and intersect S in a unique point. Only the first coordinate $x_1 = \xi_1(x)$ or the first two coordinates will play a role in the construction. One may write $S = S_0 \times M$, where S_0 is a quadratic surface of real dimension 0, 1 or 3 in the one- or two-dimensional space $\mathbb{R}$ or $\mathbb{R}^2$, or $\mathbb{C}$ or $\mathbb{C}^2$, depending on whether the corresponding eigenvalue is real or complex.

Suppose C has a real non-zero eigenvalue τ. Assume $C_{11} = \tau$. Remember that C has Jordan form. It may have ones below the diagonal. Since $\xi_1 \gamma^t = e^{\tau t} \xi_1$, the map γ^t changes the value of the first coordinate x_1 by a factor $e^{\tau t}$. So take $M = \{\xi_1 = 0\}$ and $S = \{|\xi_1| = 1\}$. Then S_0 is the point pair $\{-1, 1\} \text{ß} \mathbb{R}$. Each orbit in $L \setminus M$ intersects S in a unique point.

Suppose C has an eigenvalue ω which does not lie on the imaginary or the real axis. Assume $C_{11} = \omega$. Then $\zeta_1 \gamma^t = e^{\omega t} \zeta_1$, where ζ_1 is the first complex coordinate. The map γ^t changes the value of the first (complex) coordinate z_1 by a factor $e^{\omega t}$. So take $M = \{\zeta_1 = 0\}$, and $S = \{|\zeta_1| = 1\}$. Then S_0 is the unit circle in $\mathbb{C}$. Since $\Re\omega \neq 0$ the map $t \mapsto |e^{\omega t}|$ from $\mathbb{R}$ to $(0, \infty)$ is a homeomorphism and for any vector $z \in L \setminus M$ there is one value of $t \in \mathbb{R}$ such that $|e^{\omega t} z| = 1$.

Now assume all eigenvalues of C lie on the imaginary axis. If C is complex diagonal then $L = L_{00}$ and we are done. So assume $C_{21} = 1$. First suppose $C_{11} = C_{22} = 0$. Then $\xi_1 \gamma^t = \xi_1$ and $\xi_2 \gamma^t = \xi_2 + t\xi_1$. The linear map γ^t transforms the first two coordinates (x_1, x_2) into $(x_1, x_2 + t x_1)$. So take $M = \{\xi_1 = 0\}$ and $S = \{\xi_2 = 0\} \setminus M$. Then S_0 is the horizontal axis in $\mathbb{R}^2$, with the origin removed. For any vector $x \in L \setminus M$ there is a unique $t \in \mathbb{R}$ such that $x_2 + t x_1 = 0$.

Finally assume $C_{11} = C_{22} = ic$ with $c \neq 0$. Then $\zeta_1 \gamma^t = e^{ict}\zeta_1$ and $\zeta_2 \gamma^t = e^{ict}(\zeta_2 + t\zeta_1)$. So γ^t transforms the first two complex coordinates (z_1, z_2) into $e^{ict}(z_1, z_2 + tz_1)$. As t varies, the second coordinate of the image point $e^{ict}(z_2 + tz_1)$ moves along the line $z_2 + \mathbb{R}z_1$ with constant speed, and at the same time is rotated around the origin (see figure). The first coordinate traverses a circle, the second the spiral above: $t \mapsto e^{it}(1+it)$. The first coordinate just moves at constant speed around the circle with radius $|z_1|$. On the orbit in $\mathbb{C}^2$ there is a unique point (w_1, w_2) with minimal norm. This is at the value of t which minimizes $z_2 + tz_1$. We shall not compute the coordinates w_1 and w_2, but only remark that from plane geometry it is clear that w_1 is perpendicular to w_2. So (w_1, w_2) lies on the complex-homogeneous three-dimensional real quadratic surface $u_1u_2 + v_1v_2 = 0$ in $\mathbb{R}^4$ where we identify $w_k = u_k + iv_k \in \mathbb{C}$ with the vector $(u_k, v_k) \in \mathbb{R}^2$. We set $M = \{\zeta_1 = 0\}$ and $S_0 = \{u_1u_2 + v_1v_2 = 0\} \setminus \{(0,0,0,0)\}$.



The first coordinate traverses a circle, the second the spiral above:

$$t \mapsto e^{it}(1+it).$$

In the construction first use up the non-zero real eigenvalues, and then the complex eigenvalues which do not lie on the imaginary axis. If these have been depleted, we are left with a linear subspace on which all eigenvalues of C lie on the imaginary axis. If C is diagonal, we are done; if not we may use one of the last two procedures to decrease the dimension. □

We can now determine the *one-parameter groups* for which excess measures exist.

Theorem 18.42. *Excess measures exist for the one-parameter group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, if the group has a translation component, $C_{10} = 1$ in the Jordan form, or if the group is linear and C has a real eigenvalue $\lambda < 0$, or an eigenvalue $\lambda \in \{\Re > 0\}$, or if the Jordan form has a diagonal submatrix $\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$, and only in these cases.*

Proof. In the cases mentioned in the theorem one may construct a measure ρ_z living on the orbit Γ_z which is finite on a halfspace J_0 defined in terms of the first or first two coordinates. To obtain a measure that is full one may take a finite mixture of such elementary measures. If the condition is not fulfilled then C is linear. We may assume that all eigenvalues have the form $\zeta = \tau + i\varphi$ with $\varphi \neq 0$ and $\tau \leq 0$, and we may assume that $\mathbb{R}^d = \mathbb{C}^{d_c}$ with $d_c = d/2$. It suffices to show that for any elementary measure ρ_z and any halfspace J the mass $\rho_z(J)$ is zero or infinite. Since $\rho_z(\partial J_0) = 0$ it suffices to show that $\rho_z\{\xi > j_0\} = \infty$ whenever Γ_z intersects the open halfspace $\{\xi > j_0\}$. We may take $j_0 \in \{-1, 0, 1\}$.

Let η be the measure on $\mathbb{R}$ with density e^{-t}. We have to show that $T \cap (-\infty, 0]$

has infinite η-mass whenever

$$T = \{t \in \mathbb{R} \mid \Re\varphi(t) > j_0\}$$

is non-empty, where

$$\varphi(t) = e^{i\varphi_1 t} e^{\tau_1 t} P_1(t) + \cdots + e^{i\varphi_m t} e^{\tau_m t} P_m(t).$$

Here the P_k are complex polynomials, and $\tau_k \le 0$, $\varphi_k \ne 0$, and the complex numbers $\tau_k + i\varphi_k$ are distinct. First assume all τ_k vanish, and the polynomials are constants, $P_k = c_k \ne 0$. The function φ is periodic or almost periodic. It is bounded and so is its derivative. For any interval $[s_1, s_2]$, and $\varepsilon > 0$, $t_0 > 1$, there exists $t > t_0$ such that $|\varphi(s-t) - \varphi(s)| < \varepsilon$ for $s_1 \le s \le s_2$. For T non-empty there exists $\delta > 0$ such that $T \cap (-\infty, 0]$ contains infinitely many intervals of length $> \delta$. If there are $\tau_k < 0$, let τ be the minimal value. We may restrict attention to the terms with $\tau_k = \tau$. Set

$$T = \{t \in \mathbb{R} \mid \Re(e^{i\varphi_1 t} c_1 + \cdots + e^{i\varphi_m t} c_m) > j_0 e^{-\tau t}\}.$$

Observe that the right side vanishes for $t \to -\infty$. Hence the argument above applies here too. Similarly if the polynomials are non-constant with degrees $g_1, \dots, g_m$, we may neglect all powers t^q with $q < g = \max\{g_1, \dots, g_m\}$. □

For exceedances over horizontal thresholds or ellipsoids there is a simple description of the orbits by means of a homeomorphism Φ from the product $\mathcal{X} \times \mathbb{R}$ onto the natural domain of the excess measure:

$$\Phi : (x, t) = \gamma^t(x), \quad x \in \mathcal{X},$$

where $\mathcal{X}$ is the horizontal coordinate plane or the unit sphere in appropriate coordinates. The excess measure ρ is the image of the product measure $\rho^*(dx) \times e^{-t} dt$. In general such a representation holds locally.

Lemma 18.43 (Local Sections). *Let $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, be a one-parameter group of affine transformations on $\mathbb{R}^d$, p a point in $\mathbb{R}^d$, and ξ a non-zero linear functional. Let M be the hyperplane $\{\xi = c\}$ containing p. Define*

$$\Phi(z, t) = \gamma^t(z), \quad z \in M,\ t \in \mathbb{R}.$$

If the derivative $\xi C p$ of the function $t \mapsto \xi\gamma^t(p)$ in $t = 0$ does not vanish there exists $\delta > 0$ such that Φ restricted to $D \times (-\delta, \delta)$ is a diffeomorphism. Here the set $D = M \cap (p + \delta B)$ is the open disk in M of radius δ, centered in p.

Proof. Introduce coordinates such that $p = 0$ and ξ the vertical coordinate. Then M is the horizontal coordinate plane, and we may write $\Phi(x, t) = \gamma^t(x, 0)$, $x \in \mathbb{R}^h$. In these coordinates

$$\Phi'(0,0) = \begin{pmatrix} I & q \\ 0 & s \end{pmatrix}, \quad s = \xi C p \ne 0.$$

Now apply the Inverse Function Theorem. □

In case the orbit Γ_p is unbounded it is natural to wonder whether one can take the disk D in the lemma so small that Φ maps the full open cylinder $D \times \mathbb{R}$ onto an open tube around the orbit Γ_p, and that this map is a diffeomorphism.

If γ is not linear (in any coordinates) then one may choose D to be a horizontal coordinate plane in the Jordan coordinates. Proposition 18.41 above exhibits such diffeomorphisms Φ with $D = \{|\zeta| = 1\}$ a pair of hyperplanes or a cylinder for a real or complex eigenfunctional ζ whose eigenvalue does not lie on the imaginary axis. In general such a cylinder $D \times \mathbb{R}$ need not exist. The example below presents an orbit Γ_0 of a one-parameter group of affine transformations which preserve Lebesgue measure. The orbit is unbounded, $\|\gamma^t(0)\| \to \infty$ both for $t \to -\infty$ and for $t \to \infty$. The images $\gamma^n(\delta B)$ intersect the unit ball infinitely often, no matter how small one takes $\delta > 0$.

Example 18.44. It is simpler to use a group of linear transformations and a point $p \neq 0$. Let $d = 4$, and $C = J$, $\gamma^t = e^{tJ}$ as in (18.22). Points on the vertical axis are fix points; for other points the orbits diverge. Take $p = (0, 0, 1, 1)$, and $z_n = (12/n^2, -6/n, 1, 1)$. Then $\gamma^n(z_n) = (12/n^2, 6/n, 1, 1)$. ♢

18.10* Uniqueness of extensions. If we know the one-parameter group of symmetries of the excess measure ρ, then ρ is determined by the spectral measure. Suppose that ρ has a density g. If ρ is an excess measure for exceedances over horizontal thresholds, it suffices to know g on a horizontal hyperplane; if ρ is an excess measure for exceedances over elliptic thresholds, it suffices to know g on the unit sphere. The extension follows by (14) in the Preview.

If the symmetry group is not given, the situation is less simple. In this section we consider the following situation: We are given a Radon measure μ on an open set U in $\mathbb{R}^d$, an affine transformation γ, and a constant $C > 1$ such that

$$\gamma(\mu) = C\mu \ \text{ on } U_1 = \gamma(U) ß U.$$

The sequence $U_k = \gamma^k(U)$, $k \in \mathbb{Z}$, is decreasing. For $n \geq 1$ the measures $\gamma^n(\mu)$ and $C^n\mu$ agree on U_n. For $-m \leq 0$ one may define $\mu_m = C^m\gamma^{-m}(\mu)$ on U_{-m}. This measure agrees with μ_j on U_{-j} for $j = 0, \ldots, m$. As in Section 14.4 there is a unique measure ρ on $O = \bigcup U_k$, the γ-*extension* of μ which satisfies

$$\gamma(\rho) = C\rho, \quad 1_U d\rho = d\mu.$$

So the Radon measure μ on U above has a unique extension to a Radon measure ρ on an open set O, which satisfies $\gamma(\rho) = \rho$. A problem arises if there is a second affine transformation α with the same properties as γ. Do the two extensions agree?

In the simplest case ρ is an excess measure on $\mathbb{R}^d \setminus \{0\}$ with a continuous density and symmetries $\gamma^t(\rho) = e^t\rho$. Let μ be the restriction of ρ to the complement U of a compact set K, and let $\alpha(K) \supset K$, and $\alpha(\mu) = C\mu$ on $\alpha(U)$. If the α-extension of μ agrees with ρ then $\alpha = \gamma^t\sigma$ for $t = \log C$ and an appropriate measure preserving transformation σ of ρ. In particular α will be linear. It is an open problem whether the α-extension agrees with ρ. A similar problem was encountered in Section 14.4 and left unresolved. In this section and the next we present some partial results.

Suppose α and γ are affine transformation, and $\gamma(B) = \alpha(B)$. This does not imply $\gamma^2(B) = \alpha^2(B)$, nor $\gamma^{-1}(B) = \alpha^{-1}(B)$. If π is the uniform distribution on B, then $\gamma(\pi) = \alpha(\pi)$ does not imply $\gamma^2(\pi) = \alpha^2(\pi)$ or $\gamma^{-1}(\pi) = \alpha^{-1}(\pi)$.

Example 18.45. Take $\gamma = \mathrm{diag}(1, 2)$ and $\alpha = \gamma R$ where R is a rotation over $\pi/2$ counterclockwise. Then $\gamma^2(B)$ is a vertically elongated ellipse, $\alpha^2(B) = 2B$; $\gamma^{-1}(B)$ is a horizontal ellipse, $\alpha^{-1}(B)$ is this ellipse rotated over $\pi/2$. ♢

We begin by showing that for a compact set K with non-empty interior the equalities $\alpha^k(K) = \gamma^k(K)$, $k \geq 1$, imply the same equality for all $k \in \mathbb{Z}$. We need an algebraic result.

Proposition 18.46. *Let $\gamma \in \mathcal{A}$, and let $\mathcal{G}$ be a subgroup of $\mathcal{A}$ such that $\gamma\mathcal{G}\gamma^{-1} = \mathcal{G}$. If $\sigma_n := \alpha^{-n}\gamma^n \in \mathcal{G}$ for $n \geq 1$, then this holds for all $n \in \mathbb{Z}$.*

Proof. First note that

$$\sigma_{-1} = \alpha\gamma^{-1} = \gamma(\gamma^{-1}\alpha)\gamma^{-1} = \gamma\sigma_1^{-1}\gamma^{-1} \in \mathcal{G}.$$

Now suppose $\sigma_{-m} \in \mathcal{G}$. Then

$$\sigma_{-m-1} = \alpha^{m+1}\gamma^{-m-1} = \gamma(\gamma^{-1}\alpha)(\alpha^m\gamma^{-m})\gamma^{-1} = \gamma\sigma_1^{-1}\sigma_{-m}\gamma^{-1} \in \mathcal{G}.$$

By induction the result holds for all $m \geq 0$. □

Corollary 18.47. *Let $\mathcal{G}$ be a compact subgroup of $\mathcal{A}$. If $\alpha^{-n}\gamma^n \in \mathcal{G}$ for $n \geq 1$, then also for $n \leq 0$. Moreover $\mathcal{G}$ contains a compact subgroup $\mathcal{K}$ such that $\gamma\mathcal{K}\gamma^{-1} = \mathcal{K}$, and $\alpha^{-k}\gamma^k \in \mathcal{K}$, $k \in \mathbb{Z}$.*

Proof. Let $\mathcal{H}$ be the algebraic group generated by the elements $\sigma_n = \alpha^{-n}\gamma^n$, $n \geq 1$. The elements

$$\gamma^{-1}\sigma_n\gamma = (\gamma^{-1}\alpha)(\alpha^{-n-1}\gamma^{n+1}) = \sigma_1^{-1}\sigma_{n+1}$$

lie in $\mathcal{H}$ and generate $\gamma^{-1}\mathcal{H}\gamma$. So $\gamma^{-1}\mathcal{H}\gamma\text{ß}\mathcal{H}$. This relation then also holds if we replace $\mathcal{H}$ by its closure, the compact group $\mathcal{K}$, say. Proposition 18.83 gives $\gamma^{-1}\mathcal{K}\gamma = \mathcal{K}$. Hence also $\gamma\mathcal{K}\gamma^{-1} = \mathcal{K}$. The proposition above, applied to $\mathcal{K}$, gives $\alpha^{-k}\gamma^k \in \mathcal{K}$ for $k \in \mathbb{Z}$. □

Theorem 18.48. *Let μ be a finite non-degenerate measure on H_+. Let $\alpha, \gamma \in \mathcal{A}^h$, and suppose $\alpha(H_+) = \gamma(H_+) = H_1\text{ß}H_+$, and $\alpha(\mu) = \gamma(\mu) = C\mu$ on H_1 for some $C > 1$. Then $\alpha^k(H_+) = \gamma^k(H_+) := H_k\text{ß}H_{k-1}$ for $k \in \mathbb{Z}$, and there exists a Radon measure ρ on the open set $O = \bigcup_k H_k$ such that $\alpha(\rho) = \gamma(\rho) = C\rho$ and $1_{H_+}d\rho = d\mu$.*

Proof. Write $H_k = \mathbb{R}^h \times [c_k, \infty)$. Then $C > 1$ implies $c_1 > c_0 = 0$. Write $\tilde{\alpha}(v) = av + c_1, \tilde{\gamma}(v) = cv + c_1$ with $a, c > 0$, and set $M(t) = \tilde{\mu}[t, \infty)$. Here $\tilde{\mu}$ is the vertical component of μ, and $\tilde{\alpha}$ and $\tilde{\gamma}$ describe the effect of α and γ on the vertical coordinate. Then $M(\tilde{\alpha}^n(0)) = M(\tilde{\gamma}^n(0)) = 1/C^n$. The asymptotic behaviour of M determines the value of a and c. In particular $a = c$, as some reflection will show. Hence $\tilde{\alpha} = \tilde{\gamma}$ and $\tilde{\alpha}^k(0) = \tilde{\gamma}^t(0) = c_k$ implies $\alpha^k(H_+) = \gamma^k(H_+)$ for $k \in \mathbb{Z}$. Let $\mathcal{G}$ be the compact group of those $\sigma \in \mathcal{A}$ which preserve μ and map H_+ onto itself. See Theorem 18.75. Then $\alpha^n(H_+) = \gamma^n(H_+)$ for $n \geq 1$ together with $\alpha(\mu) = \gamma(\mu) = C\mu$ implies $\alpha^n(\mu) = \gamma^n(\mu)$ for $n \geq 1$ by induction. Hence $\alpha^{-n}\gamma^n \in \mathcal{G}$ for $n \geq 1$. Corollary 18.47 gives this relation for $n \in \mathbb{Z}$. Hence $\mu_m = C^m\alpha^{-m}(\mu) = C^m\gamma^{-m}(\mu)$ on H_{-m} for $m \geq 1$, and the γ- and α-extensions of μ agree. □

Suppose μ is a finite non-degenerate measure on $U = K^c$, where K is a compact set with non-empty interior. Let $\alpha, \gamma \in \mathcal{A}$ map K onto the same compact set $K_1 \supset K$.

Proposition 18.49. *Let μ, α, γ and U, K, K_1 be as above. Let $C > 1$. Suppose $\alpha(\mu) = \gamma(\mu) = C\mu$ on K_1^c. If $\alpha^n(K) = \gamma^n(K)$ for $n \geq 1$, then $\alpha^k(K) = \gamma^k(K) = K_k$ß$K_{k+1}$ for $k \in \mathbb{Z}$. There exists an affine subspace M such that $\bigcap K_k = M \cap K$. There exists a unique Radon measure ρ on M^c such that*

$$\alpha(\rho) = \gamma(\rho) = C\rho, \quad 1_U d\rho = d\mu.$$

We may choose the origin in a point $p \in K$ such that α and γ are linear, and M is a linear subspace.

Proof. First observe that $\sigma_n := \alpha^{-n}\gamma^n$ lies in the compact group of affine transformations mapping K onto itself for $n = 1, 2, \ldots$, and that $\sigma_n(\mu) = \mu$. The compact group $\mathcal{K}$ generated by $\sigma_1, \sigma_2, \ldots$ contains the elements $\sigma_k = \alpha^{-k}\gamma^k$ for $k \in \mathbb{Z}$, and $\gamma\mathcal{K}\gamma^{-1} = \mathcal{K}$ by Corollary 18.47). As above the extension ρ is unique. As in Section 17.7 the intersection $\bigcap K_k$ has the form $K \cap M$ for an affine subspace M. □

Theorem 18.50 (Uniqueness Theorem). *Let ρ be a Radon measure on $\mathbb{R}^d \setminus \{0\}$ and γ a linear expansion such that $\gamma(\rho) = C\rho$ for some constant $C > 1$. Let F be a convex compact set such that $\rho(F^c)$ is finite. Let $\alpha \in \mathcal{A}$ such that Fß$\alpha(F)$ and $\alpha(1_{F^c} d\rho) = C d\rho$ holds on the complement of $\alpha(F)$. If α is linear then $\alpha(\rho) = C\rho$. If the convex hull of the support of ρ is the whole space $\mathbb{R}^d$ then α is linear.*

Proof. Let $U = U(\varepsilon)$ be the union of all open halfspaces of mass ε. We choose $\varepsilon > 0$ so small that F lies in the complement K of U. Let K_n be the complement of $U(\varepsilon/C^n)$ for $n \geq 0$. Then KßK_1 and $K_n = \alpha^n(K) = \gamma^n(K)$ and $\alpha(\mu) = \gamma(\mu) = c\mu$ on K_1^c. Now apply the previous proposition. If α is linear the same argument applies with $U(\varepsilon)$ the union of all sets $\{|\xi| > 1\}$ of mass ε. □

18.11* Local symmetries. An excess measure may be more symmetric in some points than others. The local symmetry enables us to show that excess measures for expansions agree if they agree on large enough open sets.

If μ is a non-zero finite measure on the open set U, and ρ_1 and ρ_2 are multivariate GPD measures which agree with μ on U then $\rho_1 = \rho_2$, unless $\tau_1 = \tau_2 = -1$. This is obvious. The densities are analytic functions. If they agree on the set U they agree everywhere. If the Pareto parameter is -1 then the GPD measure has a density which is constant on some open paraboloid P. If μ is uniformly distributed on U, all one can say is that P contains U. In this section we derive two uniqueness results for excess measures.

Example 18.51. Let ρ on $O = \mathbb{C} \times \mathbb{R} \setminus \{0\}$ be the sum of a measure on the open cone $C = \{|z| < v\}$ with density $1/(|z|^2 + v^2)^2$, and a measure ρ_0 on the spiral (e^{t+it}, e^t) on the boundary of C, which satisfies $\rho_0\{v \geq c\} = 1/c$ for $c > 0$. Then ρ is an excess measure for the group of linear expansions $\gamma^t(z, v) = e^t(e^{it}z, v)$. On C the local symmetry group consists of the affine transformations with determinant ± 1; off $\mathrm{cl}(C)$ the local symmetry group consists of all affine transformations; on the spiral it consists of the linear transformations γ^t, and elsewhere on $\partial C \setminus \{0\}$ it is the linear group generated by the rotations in the horizontal plane and scalar expansions. $\diamondsuit$

From the example we see that for each $w \in O$ in the support of ρ there exists a neighbourhood U and a collection of symmetries α such that $\alpha(\rho) = \Theta(\alpha)\rho$ on $U \cap \alpha(U)$. Let $\mathcal{G}_w$ denote this group of symmetries. Observe that $\Theta \colon \mathcal{G}_w \to (0, \infty)$ is a homomorphism. If ρ vanishes on a neighbourhood of w, then $\mathcal{G}_w = \mathcal{A}$, but Θ is not defined.

We shall restrict attention to symmetries α close to id for which α^t also is a symmetry for $|t| \leq 1$. Hence it is more convenient to work with the Lie algebra $\mathfrak{g}_w$ and the linear map $\theta \colon \mathfrak{g}_w \to \mathbb{R}$ corresponding to the homomorphism Θ. See Section 18.13 for details.

Definition. Let ρ be a Radon measure on the open set Oß$\mathbb{R}^d$. For each $w \in O$ in the support of ρ define $(\mathfrak{g}_w, \theta_w)$. Here $\mathfrak{g}_w$ is a Lie algebra, a subalgebra of the Lie algebra $\mathfrak{a}$ of the group of all affine transformations, and $\theta_w \colon \mathfrak{g}_w \to \mathbb{R}$ is a linear functional on $\mathfrak{g}_w$ which vanishes on the Lie bracket, $\theta_w([A, B]) = 0$ for $A, B \in \mathfrak{g}_w$. If there exists an open neighbourhood $U = w + \varepsilon B$ of w and an open ball $W = \delta B$ in $\mathfrak{g}_w$ such that

$$\alpha(\rho) = e^{\theta(A)}\rho \ \text{ on } U \cap \alpha(U), \quad \alpha = e^A, \ A \in W,$$

then $(\mathfrak{g}_w, \theta_w)$ is called a *local symmetry* of ρ in w; if $\mathfrak{g}_w$ is maximal, it is called the local symmetry.

It is not obvious that local symmetries may be described in terms of Lie algebras. Let us say a few words to clarify the matter.

For simplicity assume ρ has a smooth positive density g, and $\alpha^t = e^{tA}$ and $\beta^t = e^{tB}$ preserve the density locally, say on the open set Uß$\mathbb{R}^d$. One may regard A and B as vector fields on U. Then $g(\alpha^t(z)) = g(z)$ for $z \in U$, if $t > 0$ and $\alpha^s(z) \in U$, $0 \le s \le t$. In terms of the vector field A this means that $Ag = 0$ on U. Similarly $Bg = 0$. It follows that $Xg = 0$ for all $X \in \mathbb{R}A + \mathbb{R}B$. So the vector fields which preserve g around a point w form a linear space. Now define $\gamma(t) = \alpha^t\beta^t\alpha^{-t}\beta^{-t}$. Let $z \in U$, and assume U contains the elements $\alpha^{t_1}\beta^{t_2}\alpha^{s_1}\beta^{s_2}z$ for $-\delta \le s_i, t_i \le \delta$, $i = 1, 2$. Then g is constant on the curve $t \mapsto \gamma(t)(z)$. By continuity this implies that $([A, B]g)(z) = 0$. We conclude that the generators of the local one-parameter groups which preserve the density form a *Lie algebra*. If $g(\alpha_i^t(z)) = e^{\theta_i t}g(z)$ locally, then

$$(A_i g)(z) = \lim_{t\to 0}\frac{g(\alpha_i^t(z)) - g(z)}{t} = \lim_{t\to 0}\frac{e^{\theta_i t} - 1}{t}g(z) = \theta_i g(z),$$

and $Xg = \theta(X)g$ where θ is the linear function with value θ_i in A_i. Note that $[A, B]g$ vanishes since $g(\gamma(t)(z)) = g(z)$ – the positive factors cancel. It remains to show that these identities for the vector fields allow one to choose $\varepsilon > 0$ and $\delta > 0$ so small that

$$g(\alpha) = e^{\theta(A)}g \quad \text{on } U_0 \cap \alpha(U_0), \quad \alpha = e^A,\ A \in \delta B,\ U_0 = w + \varepsilon B. \tag{18.23}$$

Choose $\varepsilon > 0$ so small that U contains the compact closure K of $U_0 = w + \varepsilon B$. Let $r > \varepsilon$, and $\alpha_n \to$ id in $\mathcal{A}$. Then $\alpha_n(K)$ß$w + rB$ eventually. So there exists $\delta > 0$ such that $e^A(K)$ßU for all $A \in \delta B$ß$\mathfrak{g}_w$. (Even for all $A \in \delta B$ß$\mathfrak{a}$.) For $z \in U_0$ and $\|A\| < \delta$ the curve $t \mapsto z(t) = e^{tA}(z)$, $0 \le t \le 1$, lies in U, and hence $(Ag)(z(t)) = dg(z(t))/dt = cg(z(t))$ with $c = e^{\theta(A)}$ gives $g(e^{tA}(z)) = e^{t\theta(A)}g(z)$, and $g(\alpha(z)) = e^{\theta(A)}g(z)$. We have established (18.23). The group structure of local symmetries given here is the Frobenius Theorem. See Varadarajan [1974] for details.

The flow $z \mapsto z_t = e^{tC}z$, $t \in \mathbb{R}$, on $\mathbb{R}^d$ induces a flow of measures $\mu \mapsto \mu_t$. This is clear if μ is the point mass in a point z_0, or a mixture of such point masses. In general for a flow on O one may define for any function $\varphi\colon O \to \mathbb{R}$ the function φ_t by $\varphi_t(z) = \varphi(z_t)$. For functions with compact support we define $\mu_t\varphi = \mu\varphi_t$, in the same way in which one sets $T(\pi)\varphi = \int \varphi(T)d\pi = \mathbb{E}\varphi(T(Z))$ if π is the distribution of Z and T any measurable map. The flow is measure preserving if $\mu_t = \mu$ for all $t \in \mathbb{R}$. For smooth functions φ with compact support we may interchange integral and derivative, and find $\mu C\varphi = 0$. The flow is symmetric if $\mu_t = e^{\theta t}\mu$, or equivalently $\mu C = \theta\mu$ for the vector field C. As above the generators of the local symmetries form a Lie algebra in each point $w \in O$ in the support of μ.

Example 18.52. It may be impossible to extend local symmetries to global symmetries. Let ρ on $O = \mathbb{C} \setminus \{0\}$ have density $g(re^{i\varphi}) = e^{\varphi}/r^3$ for $0 \le \varphi < 2\pi$, $r > 0$.

Then ρ is an excess measure: $\gamma^t(\rho) = e^t \rho$ for $\gamma^t : w \mapsto e^t w$. The local symmetry $(\mathfrak{g}_w, \theta_w)$ is the two-dimensional *Lie algebra* $\mathfrak{g}_w$ with generators

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and the linear map $\theta(I) = \theta(J) = 1$ for almost all $w \in \mathbb{C}$. The associated Lie group $\mathcal{G}$ consists of the maps $w \mapsto zw$, $z \in \mathbb{C} \setminus \{0\}$. The set O is open, connected, and invariant under $\mathcal{G}$; but $\mathcal{G}$ is not the symmetry group of ρ. ♢

We shall now give a simple, more positive extension result.

Lemma 18.53. *Let ρ be a Radon measure on the open set $O \text{ß} \mathbb{R}^d$. Let $a > 0$, and $\alpha \in \mathcal{A}$. Suppose for each $w \in O$ there exists an open neighbourhood U_w such that*

$$\alpha(U_w) \text{ß} O, \quad \alpha(1_{U_w} d\rho) = a 1_{\alpha(U_w)} d\rho. \tag{18.24}$$

Then $\alpha(O) \text{ß} O$, and $\alpha(\rho) = a\rho$ on $\alpha(O)$.

Proof. If two measures are defined and equal on the sets $\alpha(U_w)$ they coincide on the union $\alpha(O)$. □

Corollary 18.54. *Let $\delta > 0$ and $\theta \in \mathbb{R}$. Suppose (18.24) holds on U_w for $\alpha = \gamma^t$, $|t| < \delta$, with $a(t) = e^{\theta t}$ for each $w \in O$. Then $\gamma^t(O) = O$ for $|t| < \delta$, and $\gamma^t(\rho) = e^{\theta t}\rho$ on O. These relations then also hold for $t \in \mathbb{R}$.*

Proposition 18.55. *Let ρ be a Radon measure on the open set $O \text{ß} \mathbb{R}^d$. Let O be invariant under the one-parameter group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$. Suppose $C \in \mathfrak{g}_w$ for each $w \in O$ in the support of ρ, and $\theta_w(C) = \theta_0$ for some constant $\theta_0 \in \mathbb{R}$. Then $\gamma^t(\rho) = e^{\theta_0 t}\rho$, $t \in \mathbb{R}$.*

Proof. The invariant points of the one-parameter group form an affine subspace M_0 in $\mathbb{R}^d$, which may be empty. For $w \in O \setminus M_0$, by Lemma 18.43 there exists $\delta > 0$, a centered open ball D in $\mathbb{R}^h$, a smooth map $\Phi : D \times \mathbb{R} \to O$, and an affine map $\varphi_0 : \mathbb{R}^h \to \mathbb{R}^d$, such that $\varphi_0(\mathbb{R}^h)$ is a hyperplane, $\varphi_0(0) = w$, and such that

$$\Phi(x, t) = \gamma^t(\varphi_0(x)), \quad x \in \mathbb{R}^h, t \in \mathbb{R}$$

restricted to $D \times (-\delta, \delta)$ is a diffeomorphism onto an open neighbourhood U of w. There is a measure $d\mu = d\mu_0 \times e^{-\theta_0 t} dt$ on $D \times \mathbb{R}$ such that the image of μ on $D \times (-\delta, \delta)$ under Φ is ρ on U. The restriction of Φ to $D \times (t - \delta, t + \delta)$ is a diffeomorphism onto an open subset U^t of O for each $t \in \mathbb{R}$, and the image of μ on $D \times (t - \delta, t + \delta)$ is the measure ρ on U^t. (Let C be a compact disk in D, and suppose the relations hold for μ restricted to $C \times [0, s)$. The local symmetry in the points $w \in \Phi(C \times \{s\})$ allows one to increase s slightly.) The restriction ρ_0 of ρ to the open set $\Phi(D \times \mathbb{R})$ need not be the image of μ since Φ need not be injective, even if the orbit through w is unbounded, as we saw in Example 18.44. However $\gamma^t(\rho_0) = e^{\theta_0 t}\rho_0$ holds for $t \in (-\delta, \delta)$ since the equality holds on slices U^t. Hence $\gamma^t(\rho_0) = e^{\theta_0 t}\rho_0$ holds for all t. This relation then also holds for ρ on O. If $\theta_0 \neq 0$ then ρ vanishes on M_0; if $\theta_0 = 0$ then ρ may be an arbitrary Radon measure on $M_0 \cap O$. □

Example 18.56. Let ρ have density g on $O = \mathbb{R}^2 \setminus \{0\}$, where $g(u, v) = u^2 + v^2$ for $u, v > 0$, $g = 1$ on $(-\infty, 0)^2$, and g vanishes elsewhere. Here the exponent θ depends on w. There is local symmetry, but no global symmetry, even though O is connected. ◊

If ρ is an excess measure, the local symmetry is constant along orbits of the global symmetry. If $\mathfrak{g}$ß$\mathfrak{a}$ is a Lie algebra, and $\alpha \in \mathcal{A}$, then $\mathfrak{g}' = \alpha\mathfrak{g}\alpha^{-1}$ is a Lie algebra. If θ is a linear functional on $\mathfrak{g}$ which vanishes on $[\mathfrak{g}, \mathfrak{g}]$, then $\theta' \colon \alpha C\alpha^{-1} \mapsto \theta(C)$ is a linear functional on $\mathfrak{g}'$ which vanishes on $[\mathfrak{g}', \mathfrak{g}']$ since $\alpha AB\alpha^{-1} = \alpha A\alpha^{-1}\alpha B\alpha^{-1}$.

Proposition 18.57. *If $\alpha = e^A$ for some $A \in \mathfrak{g}$ then $\mathfrak{g}' = \mathfrak{g}$ and $\theta' = \theta$.*

Proof. The Lie groups $\mathcal{G}$ and $\mathcal{G}' = \alpha\mathcal{G}\alpha^{-1}$ coincide for $\alpha \in \mathcal{G}$. Locally one may define $\Theta(e^C) = e^{\theta C}$ for $C \in \mathfrak{g}$. Then $\Theta(\beta\gamma) = \Theta(\beta)\Theta(\gamma)$ for β and γ close to id. In particular $\Theta(\alpha^t\gamma^s\alpha^t) = \Theta(\gamma^s)$ for $\gamma^s = e^{sC}$, $|t| \leq \delta$, $|s| < \delta_C$; and hence $\theta(\alpha^t C\alpha^{-t}) = \theta(C)$ for $|t| \leq \delta$. By iteration

$$\theta(\alpha^t C\alpha^{-t}) = \theta(\alpha^s C_n\alpha^{-s}) = \theta(C_n) = \theta(C), \quad t \geq 0, t = n\delta + s, 0 \leq s < \delta$$

where $C_0 = C$ and $C_n = \alpha^\delta C_{n-1}\alpha^{-\delta}$ for $n \geq 1$. □

Corollary 18.58. *Let ρ be a Radon measure on the open set Oß$\mathbb{R}^d$, and γ^t, $t \in \mathbb{R}$, a one-parameter group of affine transformations such that $\gamma^t(O) = O$, and $\gamma^t(\rho) = e^{qt}\rho$ for all $t \in \mathbb{R}$. Then*

$$(\mathfrak{g}_{\gamma^t(w)}, \theta_{\gamma^t(w)}) = (\mathfrak{g}_w, \theta_w), \quad w \in O,\ t \in \mathbb{R}.$$

Theorem 18.59. *Suppose ρ on $\mathbb{R}^d \setminus \{0\}$ is an excess measure for the linear expansions γ^t, and ρ_0 on $\mathbb{R}^d \setminus \{w_0\}$ is an excess measure for the affine expansions γ_0^t with center w_0. Let U be a bounded open set containing the origin. Suppose each $w \in \partial U$ has a neighbourhood U_w on which ρ_0 agrees with ρ. Then $w_0 = 0$ and $\rho_0 = \rho$.*

Proof. Let $\gamma^t = e^{tC}$ and $\gamma_0^t = e^{tC_0}$. Then C and C_0 lie in $\mathfrak{g}_w$ for all $w \neq 0$ by the proposition above. Observe that $w_0 \notin \partial U$ since $\rho_0(w_0 + \varepsilon B) = \infty$ for all $\varepsilon > 0$ and $\rho(w_0 + \varepsilon B)$ is finite for $\varepsilon < \|w_0\|$. From Lemma 18.43 it follows that $w_0 \neq 0$ implies that ρ is finite on a neighbourhood of the origin. Contradiction. So $w_0 = 0$, and $O = \mathbb{R}^d \setminus \{0\}$ is invariant under γ_0^t. Apply Proposition 18.55 to ρ and γ_0^t, and conclude that ρ and ρ_0 both are excess measures for γ_0^t. Since for each $w \neq 0$ the orbit $\gamma_0^t(w)$ intersects ∂U, the measures ρ and ρ_0 agree on a neighbourhood of w. □

We now turn to exceedances over horizontal thresholds. We restrict attention to excess measures ρ which vanish off their natural domain $\mathbb{R}^h \times (j_*, j^*)$, where the interval (j_*, j^*) is the domain of the vertical component $\tilde{\rho}$ of ρ. The symmetry group of $\tilde{\rho}$ is a one-parameter group of positive affine transformations on $\mathbb{R}$, and (j_*, j^*) is an orbit of this group. There exists a point $j_0 \in (j_*, j^*)$ such that the restriction of $\tilde{\rho}$ to $[j_0, j^*)$ is a probability measure. This probability measure is a *GPD*, and the shape parameter τ of the GPD is the Pareto parameter of ρ.

Theorem 18.60. *Let μ be a finite non-zero measure on the horizontal slice $U = \{a < v < b\}$. Let ρ_1 and ρ_2 be excess measures for exceedances over horizontal thresholds on their natural domain. Suppose ρ_1 and ρ_2 agree with μ on U. If $\rho_1 \neq \rho_2$ then $\tau = -1$, and one may choose the vertical axis so that μ is a product measure on $\mathbb{R}^h \times (a, b)$.*

Proof. If ρ_1 and ρ_2 agree on U then the vertical components $\tilde{\rho}_1$ and $\tilde{\rho}_2$ have the same density on (a, b). The logarithm of the density is an analytic function on the subinterval where the density is positive. Hence $\tilde{\rho}_1 = \tilde{\rho}_2$ unless $\tau_1 = \tau_2 = -1$. If $\tilde{\rho}_1 = \tilde{\rho}_2$, the domains agree, and the argument of the previous theorem implies that ρ_1 and ρ_2 agree on a neighbourhood of every point w in their common domain $\mathbb{R}^h \times (j_*, j^*)$. If $\tau = -1$ then the vertical component of μ is uniformly distributed over an interval (a, c) with $c \leq b$. If $c < b$ then $c = j_1^* = j_2^*, \tilde{\rho}_1 = \tilde{\rho}_2$, and $\rho_1 = \rho_2$ by the argument above. If $\tilde{\mu}$ is uniformly distributed over (a, b) then $j_1^* \neq j_2^*$ is possible. However if μ is not a product measure then one may compute j^* from the conditional distribution of μ given v, for $v \in (a, b)$. Bring the generator C of ρ_1 into Jordan form, and reduce dimension to $d = 2$ or 3 by a suitable projection. If there are eigenvalues outside the imaginary axis, or if the complex Jordan form is not diagonal then the corresponding projection of μ on $\mathbb{R}^q \times (a, b)$ will reveal j_1^*. So too if there is a non-zero eigenvalue and μ is not symmetric. □

The criteria above have been developed for excess measures with extra *symmetry*. If μ is a finite measure on an open set U, and the local symmetry in $w \in U$ is $(\mathfrak{g}_w, \theta_w)$ where $\mathfrak{g}_w = \mathbb{R}C$ is a one-dimensional Lie algebra, then two excess measures which agree with μ on U will have the same symmetry group $\gamma^t = e^{tC}$ and agree on the invariant open set

$$O = \bigcup_{t \in \mathbb{R}} \gamma^t(U), \quad \gamma^t = e^{tC}.$$

18.12 Jordan form and spectral decompositions. Any square matrix may be written in a simple form, either as a diagonal matrix, or with some ones added just below the diagonal.

Let $\gamma^t = e^{tC}$ be a group of linear transformations on $\mathbb{R}^d$. Assume C is in Jordan form. If all eigenvalues of the generator C are real then on a suitable basis C

has diagonal block form with blocks $C_\zeta^m = \zeta I + J$ of size $m \geq 1$ where ζ is an eigenvalue, I the identity matrix and J the matrix with ones just below the diagonal and zeros elsewhere: $J_{ij} = 1$ if $i = j + 1$ and zero else. The matrix J maps the base vector e_i into e_{i+1} for $i = 1, \dots, m-1$ and e_m into the zero vector. Hence J^q maps e_i into e_{i+q} for $i = 1, \dots, m-q$, and maps the remaining base vectors into the zero vector. In particular $J^m = 0$. So the power series expansion gives

$$e^{t(\zeta I+J)} = e^{\zeta t}(I + tJ + t^2J^2/2 + \cdots + t^hJ^h/h!), \quad h = m-1.$$

There are as many blocks for the eigenvalue ζ as there are independent eigenvectors for ζ. If there are complex eigenvalues then one has a complex matrix of this form, and since the non-real eigenvalues come in pairs one may replace the complex block $\zeta I + J$ of size q with $\zeta = \xi + i\eta$, $\eta > 0$, by a real block of size $2q$ by replacing each complex entry $z = x + iy$ by a 2 by 2 real matrix

$$z = x + iy \longleftrightarrow \begin{pmatrix} x & y \\ -y & x \end{pmatrix}. \tag{18.25}$$

It is often simpler to regard C as a map from $\mathbb{R}^{d_r} \oplus \mathbb{C}^{d_c}$ into itself, where $d_r + 2d_c = d$. The eigenvalues of γ^t are the complex numbers $e^{\zeta_k t}$ where $\zeta_k = \xi_k + i\eta_k$ are the eigenvalues of the generator C. Hence the absolute value of the eigenvalues of $\gamma^t = e^{tC}$ are determined by the diagonal entries ξ_k in the real Jordan form of the generator C. Let $\xi_1 < \cdots < \xi_q$ be the distinct diagonal elements in the real Jordan form of the generator, with multiplicities $d_1, \dots, d_q$. Then all eigenvalues of γ lie on one of the q circles of radius $r_k = e^{\xi_k}$ in $\mathbb{C}$. The *Jordan spectral decomposition* writes

$$\begin{aligned} &\gamma^t(w) = (\gamma_1 \otimes \cdots \otimes \gamma_q)^t(w) = (\gamma_1^t(w_1), \dots, \gamma_q^t(w_q)), \\ &w = (w_1, \dots, w_q) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_q} \end{aligned} \tag{18.26}$$

where γ_k has all its eigenvalues on the circle of radius r_k in $\mathbb{C}$. See MS Sections 2.1 and 2.2 for details.

The reader will not miss much if he envisages the Jordan matrix to be diagonal with all eigenvalues real. Jordan blocks with ones below the diagonal are rare. If one picks a matrix according to a probability measure with a density on $\mathbb{R}^{d^2}$ Jordan matrices with ones below the diagonal form a null event; see Hirsch & Smale [1974]. Matrices with non-real eigenvalues do occur with positive probability. Non-real eigenvalues translate into rotations and spirals in the distribution of the excess measure.

For the result below we need to know how Jordan bases are constructed. We shall assume that C is nilpotent: $C^d = 0$. The general case may be reduced to this situation.

A Jordan basis for a nilpotent matrix C is a basis e_{ij}, $1 \leq j \leq j_i$, $i \in I$, where for each $i \in I$ the vectors $e_{i1}, \dots, e_{ij_i}$ form a chain. Here a *chain* is a finite sequence of non-zero vectors $a_1, \dots, a_k$ such that $Ca_1 = 0$ and $Ca_i = a_{i-1}$ for $i = 2, \dots, k$.

For a nilpotent matrix C a Jordan base may be constructed easily. First observe that $\ker(C)$ contains a decreasing sequence of linear subspaces $K_m = \ker(C) \cap \operatorname{im}(C^m)$, $m = 0, \dots, m_0$. Choose a base for K_{m_0}, extend to a base for K_{m_0-1}, and proceed until we have a base g_i, $i \in I$, for $\ker(C)$. This base contains subsets which form a base for K_m, $m = 0, \dots, m_0$. For each vector g_i now choose a maximal chain $g_{i1}, \dots, g_{im_i}$ with $g_{i1} = g_i$ and $C g_{ij} = g_{i,j-1}$ for $j = 2, \dots, m_i$. The collection of all these vectors g_{ij} is a Jordan base.

Proof. It suffices to prove independence. So assume

$$\sum_i \lambda_i g_{ij_0} = \sum_{j<j_0} \sum_i \lambda_{ij} g_{ij}$$

where one of the λ_is is non-zero. Apply C^{j_0-1} to both sides. This gives $\sum_i \lambda_i g_i = 0$. Contradiction. □

Proposition 18.61. *Let C be nilpotent and let $a_1, \dots, a_k$ be a chain. There is a Jordan basis for C which contains this chain.*

Proof. Since $C a_1 = 0$ one may incorporate a_1 as element $g_k \in K_1$ of the base (g_i) of $\ker(C)$. Extend the chain (a_j) to a maximal chain $g_{k1}, \dots, g_{km_k}$. Incorporate this chain in the Jordan base (g_{ij}). □

For an invariant subspace one may choose a Jordan basis which extends to a Jordan basis for the whole space.

Proposition 18.62. *Let $C : \mathbb{R}^d \to \mathbb{R}^d$ be linear. Let L be a linear subspace of $\mathbb{C}^d$ such that $C(L)$ßL. There is a Jordan basis for C such that a subset is a Jordan basis for the restriction C_L of C to L.*

Proof. If F_0 is the kernel of C^d then $F_0 \cap L$ is the kernel of C_L^d. The same argument applies to $C - \lambda$ for each eigenvalue λ. So the decomposition of $\mathbb{C}^d$ into linear subspaces F_λ corresponding to the eigenvalues λ of C induces a decomposition of L into linear subspaces $F_\lambda \cap L$ corresponding to the eigenvalues of C_L. Hence we may assume that C is nilpotent. Let K be the kernel of C. Then $K \cap L$ is the kernel of C_L. For the kernel K there is a decreasing sequence of subspaces $K = K_0 \supset \cdots \supset K_m = \{0\}$. Similarly $K \cap L = E_0 \supset \cdots \supset E_q = \{0\}$. We claim that there is a basis of K such that for each K_i and each E_j a subset is basis of this subspace. This follows from the lemma below. □

Lemma 18.63. *Let $a_1, \dots, a_d$ and $b_1, \dots, b_d$ be bases of the linear space M. Let A_m be the linear subspace spanned by $a_1, \dots, a_m$, and define B_m similarly for $m = 0, \dots, d$. There is a basis $e_1, \dots, e_d$ which for each m contains a subset which is basis for A_m, and a subset which is basis for B_m.*

Proof. By induction on the dimension d. The result is trivial for $d = 1, 2$. Suppose the desired property $\mathcal{P}$ holds for dimension less than d. Let M have dimension d. If a_1 and b_1 are linearly dependent, choose $e_1 = a_1$ and let M_1 be the quotient space M/E with $E = \mathbb{C}e_1$. Then $a_2, \dots, a_d$ and $b_2, \dots, b_d$ yield bases (a_i') and (b_i') for M_1. By hypothesis there is a basis $e_2', \dots, e_d'$ of M_1 with the property $\mathcal{P}$. Choose $e_i \in M$ which map into e_i'. The basis $e_1, \dots, e_d$ of M has the desired property $\mathcal{P}$. If a_1 and b_1 are independent we have to be more careful in our construction. Let $e_1 = a_1$ and $e_2 = b_1$, and let M_2 be the quotient space M/E where $E = \mathbb{C}e_1 + \mathbb{C}e_2$. Suppose $b_1 \in A_k$ with $k > 1$ minimal, and $a_1 \in B_q$ with $q > 1$ minimal. The vectors $a_2, \dots, a_{k-1}, a_{k+1}, \dots, a_d$ are independent modulo E, since together with e_1 and e_2 they span $\mathbb{C}^d$. Similarly the vectors $b_2, \dots, b_{q-1}, b_{q+1}, \dots, b_d$ yields a basis with vectors b_i' for M_2. Let B_m' be the subspace of M_2 spanned by $b_1', \dots, b_m'$. Then $B_1' = \{0\}$ and $B_q' = B_{q-1}'$. Since $\mathcal{P}$ holds in M_2 there is a basis $e_3', \dots, e_d'$ for M_2 with subsets spanning B_m', and A_m'. Let $e_i \in M$ map into $e_i' \in M_2$ for $i = 3, \dots, d$. Suppose $m \geq k$ and A_m' is spanned by $e_3', \dots, e_m'$. Then A_m is spanned by $e_1, \dots, e_m$. Similarly for $m < k$, let A_m' be spanned by $e_3', \dots, e_{m+1}'$. Then A_m is contained in the span of $e_1, \dots, e_{m+1}$. Successively replace $e_3, \dots, e_k$ by $\bar{e}_i = e_i + \beta_i e_2$ so that a_m is a linear combination of $e_1, \bar{e}_3, \dots, \bar{e}_{m+1}$. Then each subspace A_m is spanned by a subset of $\bar{e}_1, \dots, \bar{e}_d$, with $\bar{e}_i$ as above for $3 \leq i \leq k$, and $\bar{e}_i = e_i$ else. This remains true if we replace $\bar{e}_i$ by $\bar{e}_i + \alpha_i e_1$ for $i \geq 3$, and arbitrary α_i. We may choose the α_i so that each of the subspaces B_m is spanned by a subset of these new vectors. □

18.13 Lie groups and Lie algebras. This section is an overview of Lie groups for excess measures rather than an introduction to Lie groups.

With certain linear subspaces of the d^2-dimensional linear space of all real matrices of size d one may associate a group of linear transformations on $\mathbb{R}^d$. Loosely speaking the exponential function maps the linear subspace into the group. The linear subspace is called a Lie algebra, $\mathfrak{g}$; the associated group a Lie group, $\mathcal{G}$. For one-dimensional linear subspaces, $\mathfrak{g} = \mathbb{R}C$, the situation is simple. The associated Lie group $\mathcal{G}$ is the *one-parameter group* generated by C. Actually a one-parameter group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, is a homomorphism from $\mathbb{R}$ into GL, rather than a group. The map $t \mapsto e^{tC}$ is a homomorphism from $\mathfrak{g}$ to $\mathcal{G}$ but need not be an isomorphism. The group $\mathcal{G}$ may be compact. Think of the group of rotations in the plane. The image $\mathcal{G}$ need not be a closed subset in the space GL of all invertible linear transformations on $\mathbb{R}^d$. The image $\mathcal{G} = e^{\mathfrak{g}}$ may spiral around on a torus in GL.

Example 18.64. The one-parameter group $\gamma^t = e^{tC}$ in $\mathbb{C}^2 = \mathbb{R}^4$ with generator $C = \mathrm{diag}(ai, bi)$, with $ab \neq 0$ and $i = \sqrt{-1}$, is dense in the two-dimensional torus

$$\mathcal{T} = \{\mathrm{diag}(e^{\varphi i}, e^{\psi i}) \mid 0 \leq \varphi, \psi < 2\pi\},$$

unless there exists $t \neq 0$ such that $\gamma^t = I$. Then at and bt both are zero modulo 2π. This happens only if a/b is rational. ◊

Theorem 18.65. *If* $\mathcal{G} = \{\gamma^t = e^{tC} \mid t \in \mathbb{R}\}$ *is a one-parameter group of affine transformations, and* ρ *an excess measure, such that* (9) *in the Preview holds, then* $\mathcal{G}$ *is a closed non-compact subgroup of the group* $\mathcal{A}$ *and the map* $\exp\colon \mathfrak{g} = \mathbb{R}C \to \mathcal{G}$ *is an isomorphism.*

Proof. If $\mathcal{G}$ is compact C has a basis of complex eigenvectors with eigenvalues on the imaginary axis, and there is no excess measure. So too if $\mathcal{G}$ is not closed in $\mathcal{A}$. Section 18.9 gives details. □

By (2) in the Preview the group of affine transformations on $\mathbb{R}^d$ is a closed group of linear transformations on $\mathbb{R}^{1+d}$. So we may restrict attention to linear groups. The operator norm on the space of all matrices of size d is defined by $\|A\| = \max\{\|Ax\| \mid \|x\| = 1\}$, where $\|x\|$ is a norm on $\mathbb{R}^d$. The inequality $\|AB\| \le \|A\|\|B\|$ holds, and ensures that power series $f : X \mapsto \sum c_n X^n$ converge uniformly on $\|X\| \le r$ with $\|f(X)\| \le \sum |c_n| r^n$ provided the sum on the right is finite. In particular $X \mapsto \exp X$ converges uniformly on bounded subsets of the space of all matrices of size d, and $X \mapsto \log X$ converges uniformly on compact subsets of the open unit ball $I + B$ around the identity. The exponential function is non-linear, so the proper setting for studying these maps is differential geometry. Indeed the relation between differential geometry and Lie groups is very close. Lie groups give rise to many fundamental concepts in differential geometry.

In the general theory a Lie group is a manifold associated with a Lie algebra. This makes it possible to construct universal covering groups associated with any Lie algebra. In our approach differential geometry does not play a role; our Lie groups are subgroups of $\mathcal{A}(d)$ßGL$(1 + d)$, not necessarily closed.

Definition. A *Lie algebra* $\mathfrak{g}$ is a linear space of matrices of fixed size which is closed for the *Lie bracket*:

$$A, B \in \mathfrak{g} \Rightarrow [A, B] := AB - BA \in \mathfrak{g}.$$

The *Lie group* $\mathcal{G}$ associated with $\mathfrak{g}$ is the smallest group of matrices containing $\exp(\mathfrak{g})$. It is a subgroup of GL but need not be closed.

Let us start with some concrete examples.

Example 18.66. If $\mathfrak{h}$ is a linear space of matrices, with basis $X_1, \dots, X_m$, and if these matrices commute, $X_i X_j = X_j X_i$ for $1 \le i < j \le m$, then $\mathfrak{h}$ is a Lie algebra since $[A, B] = 0$ for $A, B \in \mathfrak{h}$. The map exp is a homomorphism from $\mathfrak{h}$ to $\mathcal{H} = \exp(\mathfrak{h})$ since $e^{A+B} = e^A e^B$ if A and B commute, as one sees by writing out the power series. The set $\mathcal{H}$ is a commutative connected group. Such a group is known to be isomorphic to the product of a vector space and a torus. ♢

In general a Lie algebra is not commutative and $e^{A+B} \ne e^A e^B$. Let us consider the Lie group associated with the vector space $\mathfrak{m}$ of all matrices of size d, which obviously is a Lie algebra. The map sign det from the group GL to the group $\{\pm 1\}$ is a continuous homomorphism. The kernel is the set GL^+ of linear transformations of positive determinant. It is both open and closed as subset of GL. The group GL^+ also is the Lie group associated with the Lie algebra $\mathfrak{m}$ of all matrices of size d. In general it is not the image of $\mathfrak{m}$ under the exponential map. Details are given below.

Example 18.67. Let $\mathfrak{m}$ be the linear space of all matrices of size d, and let $\mathcal{M}$ be the group of matrices of size d with positive determinant. Then exp maps $\mathfrak{m}$ into $\mathcal{M}$. The power series $\log(I + X) = X - X^2/2 + \cdots$ maps the open ball $I + B$ onto a set Uß$\mathfrak{m}$, and has inverse exp, as may be seen by writing out the power series for exp(log(I+X)). See Curtis [1979] for details. The map log yields an analytic homeomorphism between the open neighbourhood $I + B$ of the identity I in the group GL of all linear transformations on $\mathbb{R}^d$ and an open neighbourhood U of the origin in the d^2-dimensional linear space $\mathfrak{m}$ by Brouwer's Theorem on the invariance of domains, see Dugundji [1966]. (The Inverse Mapping Theorem will give a diffeomorphism from some open ball $I + \varepsilon B$ onto an open neighbourhood of the origin in $\mathfrak{m}$.) We may identify GL with the set $\{\det \neq 0\}$ in $\mathfrak{m}$. We claim that for each matrix X with $\det X > 0$ there is a continuous curve in GL^+ linking X to the identity. (Assume X has Jordan form. If there are ones below the diagonal in the complex Jordan form we may continuously alter these into zeros; the diagonal elements may be continuously altered to lie on the unit circle by changing their absolute value; the elements on the unit circle may be altered into ones. We end up with a real diagonal matrix with ones and minus ones. Any pair of minus ones may be altered into a pair of ones by a rotation. So we end up with the identity if the determinant of the original matrix was positive.) Conclusion: A group which contains the ball $I + \varepsilon B$ for some $\varepsilon > 0$ contains all matrices with positive determinant. ♢

Example 18.68. If $d = 2$ then the image $e^{\mathfrak{m}}$ contains the matrices $-I$ and $\mathrm{diag}(1, 2)$, but not their product $\mathrm{diag}(-1, -2)$. There is no one-parameter group $\gamma^t = e^{tC}$ such that $\gamma = \mathrm{diag}(-1, -2)$. A straightforward calculation shows that the equation $X^2 = \mathrm{diag}(-1, -2)$ has no solution. ♢

In order to see what is happening here, one may reduce the dimension and look at matrices with zero trace: the sum of the two diagonal elements vanishes. This yields a three-dimensional group, SL(2). For a readable analysis of this group, see Duistermaat & Kolk [2000]. Even in the simple situation of matrices of size two a full description is quite complex.

Matrix multiplication is notorious for non-commutativity. A geometric example is given by the group SO(3) of all rotations in $\mathbb{R}^3$.

Example 18.69. Take a match box in mind. Let α_1 denote a rotation of 90 degrees around the x-axis, and α_2 a rotation of 90 degrees around the y-axis. Then $\alpha_1^{-1}\alpha_2\alpha_1$ results in a 90 degrees rotation around the z-axis. Since rotations only involve the coordinates perpendicular to the axis, one may write down the matrices for rotations around the vertical axis, and the generator

$$\gamma^t = e^{tC} = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad C = \dot{\gamma}(0) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In any dimension a one-parameter group of rotations $R(t) = e^{tE}$ satisfies

$$R^T(t)R(t) = I \Rightarrow E^T + E = 0$$

by taking derivatives in $t = 0$. (Check that taking the transpose commutes with powers: $(A^T)^n = (A^n)^T$, and hence $(e^A)^T = e^{(A^T)}$.) We conclude that the generator of a one-parameter subgroup of $\mathrm{SO}(d)$ is anti-symmetric. The converse also holds since the implication above also goes in the other direction. In principle we can now determine in any dimension d the subgroup of $\mathrm{SO}(d)$ generated by two given one-parameter rotation groups with linearly independent generators A and B. Extend the set $\{A, B\}$ to a basis for the Lie algebra by adding Lie brackets until one obtains a set of independent matrices $X_1, \dots, X_m$ with the property that all Lie brackets $[X_i, X_j]$, $1 \le i < j \le m$, are linear combinations of $X_1, \dots, X_m$. By Example 18.64 the Lie group associated with such a Lie algebra need not be a closed subgroup of $\mathrm{SO}(d)$; the map exp may curl up the m-dimensional Lie algebra to fit into the compact set $\mathrm{SO}(d)$. $\Diamond$

A good way to measure non-commutativity is by looking at the product

$$\varphi(t) = \alpha^t \beta^t \alpha^{-t} \beta^{-t}, \quad \alpha^t = e^{tA},\ \beta^t = e^{tB},\ t \in \mathbb{R}.$$

A second order Taylor expansion gives

$$\begin{aligned}(I + At + \cdots)&(I + Bt + \cdots)(I - At + \cdots)(I - Bt + \cdots) \\ &= I + ABt^2 - BAt^2 + O(t^3), \quad t \to 0\end{aligned}$$

and hence $(\varphi(t) - I)/t^2 \to AB - BA = [A, B]$. It is clear now why Lie algebras should be closed for the Lie bracket. If e^{tA} and e^{tB} lie in $\mathcal{G}$ then $e^{t[A,B]} \in \mathcal{G}$ since

$$\gamma_n^{[tn]} \to e^{t[A,B]}, \quad \gamma_n = \varphi(1/\sqrt{n}) \in \mathcal{G}. \tag{18.27}$$

Example 18.70. Let α^t and β^t be rotations in $\mathbb{R}^3$. Choose coordinates so that the x-axis is the rotation axis of α^t, and the horizontal plane contains the rotation axis of β^t. The generators are:

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & -a & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & c \\ 0 & 0 & b \\ -c & -b & 0 \end{pmatrix} \quad \Rightarrow \quad [A, B] = \begin{pmatrix} 0 & ac & 0 \\ -ac & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

If A and B are linearly independent then $ac \neq 0$. The three-dimensional group SO(3) of rotations in $\mathbb{R}^3$ with determinant $+1$ contains no two-dimensional subgroups. ◊

Example 18.71. $\mathcal{A}^+$ is the group of positive affine transformations on the real line. Generators are matrices of size two with top row zero. The group $\mathcal{A}^+$ is non-commutative:

$$\left[\begin{pmatrix}0 & 0\\0 & 1\end{pmatrix}, \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix}\right] = \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix} - \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix} = \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix}.$$

Write this as $[A, B] = B$. Then $e^{At}y = \alpha^t y = e^t y$ and $e^{Bt}y = \beta^t y = y + t$. Check that

$$\varphi(t)y = \alpha^t\beta^t\alpha^{-t}\beta^{-t}y = e^t(e^{-t}(y-t)+t) = y + (e^t - 1)t,$$

and hence $(\varphi(t) - \mathrm{id})/t^2 \to B$, as in (18.27). ◊

A two-dimensional non-commutative Lie algebra has a basis X_1, X_2. Set $Y := [X_1, X_2] = aX_1 + bX_2$. We may assume $b \neq 0$. Set $X = X_1/b$. Then X and Y are independent, and $[X, Y] = Y$. So any non-commutative two-dimensional Lie algebra is isomorphic to the Lie algebra of the group of positive affine transformations on the reals.

These examples should give an idea of the relation between Lie groups and Lie algebras. The theory of Lie algebras helps to understand a rather tricky issue for densities of limit high risk scenarios or excess measures in dimension $d > 3$. Suppose the density g is continuous and invariant under two *one-parameter groups* $\gamma_i^t = e^{tC_i}$, $t \in \mathbb{R}$, $i = 0, 1$. Then g is constant along orbits of these two groups, and also along any continuous path which consists of a finite number of arcs from these orbits:

$$g(z) = g(z_0), \quad z = \gamma(z_0),\ \gamma = \gamma_1^{t_k}\gamma_0^{s_k}\gamma_1^{t_{k-1}}\cdots\gamma_0^{s_1}.$$

The Lie algebra $\mathfrak{g}$ generated by C_0 and C_1 may have dimension $m > 2$. This will be the case if the Lie bracket $[C_0, C_1]$ is linearly independent of the elements C_0 and C_1. It may be shown that any Lie algebra $\mathfrak{g}$ generates a Lie group $\mathcal{G}$, which locally has the dimension of the Lie algebra, by the Campbell–Hausdorff formula. The dimension of the closure of $\mathcal{G}$ may be larger still. Even though the group $\mathcal{G}$ is generated by two one-parameter subgroups, the orbits of $\mathcal{G}$ will typically be m-dimensional manifolds. The algebra forces a continuous density which is constant along the curves $\gamma_i^t(z_0)$, $t \in \mathbb{R}$, $i = 0, 1$, to be constant on the set $\mathcal{G}z_0$.

On compact groups and their orbits one may define the *uniform distribution*. A random affine transformation U is uniformly distributed on the closed group $\mathcal{G}$ß$\mathcal{A}$ if $\mathbb{P}\{U \in \mathcal{G}\} = 1$, and if $\gamma(U) = U$ in distribution for each $\gamma \in \mathcal{G}$. The distribution of U is called the *Haar measure* on $\mathcal{G}$. If $M = \mathcal{G}z$ is an orbit of $\mathcal{G}$ then the vector $Z = Uz$ is uniformly distributed on M in the sense that $\gamma(Z) = \gamma(Uz) = \gamma(U)z = Uz = Z$ in law for each $\gamma \in \mathcal{G}$.

Theorem 18.72. *Let $\mathcal{G} ß \mathcal{A}$ be a compact group. There exists a unique probability measure μ on $\mathcal{G}$, the Haar measure, such that $\gamma(\mu) = \mu$ for all $\gamma \in \mathcal{G}$.*

Proof. See Halmos [1950]. □

The importance of orthogonal groups becomes clear from the theorem below.

Lemma 18.73. *Suppose Z is a random vector and γ an affine transformation such that $\gamma(Z) = Z$ in law. If Z is standardized, then γ is orthogonal.*

Proof. Let $\gamma(z) = b + Az$. Then $b = \mathbb{E}(b + AZ) = \mathbb{E}Z = 0$, hence $AA^T = \mathbb{E}AZZ^T A^T = \mathbb{E}ZZ^T = I$. □

Theorem 18.74. *A compact subgroup of $\mathcal{A}$ is a closed subgroup of the group* O *of orthogonal linear transformations in appropriate coordinates.*

Proof. Let U be uniformly distributed on the compact group $\mathcal{G}$, and let Z_0 be uniformly distributed over the $d + 1$ points $0, e_1, \dots, e_d$, where $e_1, \dots, e_d$ are independent vectors in $\mathbb{R}^d$. Set $Z = UZ_0$. Then Z has compact support, Z is nondegenerate, and $\gamma(Z) = Z$ in law for each $\gamma \in \mathcal{G}$. Choose coordinates such that Z is standardized, and apply the lemma above. □

Theorem 18.75. *If ρ is a finite measure and full, the symmetry group is compact.*

Proof. First note that symmetries are measure preserving. We may assume that $\rho(\mathbb{R}^d) = 1$. Let Z_n have distribution ρ for $n = 1, 2, \dots$. The symmetry group $\mathcal{G}$ is a closed subgroup of $\mathcal{A}$, hence locally compact. If $\mathcal{G}$ is not compact, there is a sequence (γ_n) which diverges. Trivially $\gamma_n(Z_n) \Rightarrow Z$ and $Z_n \Rightarrow Z$. By the Convergence of Types Theorem (Theorem 1 in the Preview), the sequence (γ_n) is relatively compact. □

The converse problem of constructing a probability measure for a given *compact group* of symmetries is discussed in Meerschaert & Veeh [1995].

Excess measures are defined in terms of one-parameter groups of affine transformations. It is possible that the actual *symmetry group* of the excess measure ρ is larger. In that case there is a non-trivial group $\mathcal{S}$ of measure preserving symmetries. The group $\mathcal{S}$ is a closed normal subgroup of the closed symmetry group $\mathcal{G}$ since it is the kernel of a continuous homomorphism to $(0, \infty)$, see Theorem 18.28. The group $\mathcal{S}$ for excess measures for linear expansions is *compact* by Theorem 16.11. For excess measures associated with exceedances over *horizontal thresholds* the group $\mathcal{S}$ need not be compact.

Example 18.76. In the case of *Lebesgue measure* on the paraboloid $y > x_1^2 + x_2^2$ in $\mathbb{R}^3$ the symmetry group $\mathcal{G}$ has dimension $m = 4$. The measure preserving symmetries form a three-dimensional subgroup $\mathcal{S}$, isomorphic to the group of affine Euclidean transformations on the plane. ◊

What does the theory of Lie groups tell us about closed subgroups $\mathcal{G}$ of GL? If there is an open set in GL which contains exactly one element of $\mathcal{G}$, then $\mathcal{G}$ is discrete and each element is an isolated element of $\mathcal{G}$. In particular there exists $\varepsilon > 0$ such that $(I + \varepsilon B) \cap \mathcal{G} = \{I\}$. If $\mathcal{G}$ is not discrete it contains a non-trivial Lie group. The proof is simple:

Proposition 18.77. *A closed subgroup of the general linear group is discrete or it contains a one-parameter subgroup.*

Proof. Suppose there is a sequence $\alpha_n \in \mathcal{G}$ which converges to I with $\alpha_n \neq I$. Write $\alpha_n = e^{A_n}$ with $A_n = \log(\alpha_n)$ for $\|\alpha_n - I\| < 1$. Then $A_n \to 0$. Choose integers $m_n \to \infty$ such that $\|m_n A_n\| \to 1$. By taking an appropriate subsequence we may assume that $C_n = m_n A_n \to C$. Then $\|C\| = 1$. Let $t \in \mathbb{R}$. Set $k_n = [tm_n]$. Then $k_n A_n \to tC$ and $e^{k_n A_n} \to e^{tC}$. Since $\mathcal{G}$ is closed, and contains the elements $\alpha_n^{k_n} = e^{k_n A_n}$, it contains the limit e^{tC}. □

The matrix C is a *generator of a Lie group* for the closed subgroup $\mathcal{G}$ if $\mathcal{G}$ contains the elements e^{tC} for $t \in \mathbb{R}$. This is the case if there is a sequence $\alpha_n = e^{A_n} \in \mathcal{G}$ and a sequence of integers $m_n \to \infty$ such that $m_n A_n \to C$. Let $\mathfrak{g}$ denote the set of generators. It is not difficult to show that

1) $A, B \in \mathfrak{g}$ implies $A + B \in \mathfrak{g}$;

2) If $\alpha^t = e^{tA} \in \mathcal{G}$ for all $t \in \mathbb{R}$, and also $\beta^t = e^{tB}$, then $\log(\alpha^t \beta^t \alpha^{-t} \beta^{-t})/t^2 \to [A, B]$ for $t \to 0$.

It follows from 1) that $\mathfrak{g}$ is a linear space, and from 2) that it is a Lie algebra. Moreover, locally, the closed Lie group $\mathcal{G}$ is the image under the exponential map of an open centered ball in $\mathfrak{g}$:

Proposition 18.78. *There exists* $\varepsilon > 0$ *such that* $\log(\gamma) \in \mathfrak{g}$ *for* $\gamma \in \mathcal{G}$, $\|\gamma - I\| < \varepsilon$.

The result follows from the more technical statement in the lemma below.

Lemma 18.79. *Let* $\mathbf{L}$ *be a linear space of matrices* M*, and* $\mathcal{G}$ *a group of matrices which contains the matrices* e^M, $M \in \mathbf{L}$*. Let* $\gamma_n \in \mathcal{G}$, $\gamma_n \to I$*, and suppose* $C_n = \log \gamma_n \notin \mathbf{L}$*. Then there exists a subsequence* γ_{k_n}*, a null sequence* $B_n \in \mathbf{L}$*, a sequence* $r_n \to \infty$*, and a matrix* $A \notin \mathbf{L}$ *such that*

$$(e^{-B_n}\gamma_{k_n})^{[r_n t]} \to e^{tA}, \quad t \in \mathbb{R}.$$

Proof. Let $a = \|A\|$ and $b = \|B\|$. Then $e^{A+B} - e^B e^A = O(ab)$ for $a, b \to 0$ by comparing the power series expansions. Hence $e^{-B}e^{A+B} = e^A + Q$ with $Q = O(ab)$. Differentiability of the power series for log around $X = I$ in e^A gives $\log(e^{-B}e^{A+B}) = A + Q'$ with $Q' = O(ab)$. Write $C_n = A_n + B_n$ with $A_n \perp \mathbf{L}$. Set $s_n = 1/a_n$. Then $\|s_n A_n\| = 1$. So there is a convergent subsequence: $s_{k_n} A_{k_n} \to A$. The bound above implies $s_{k_n} t \log(e^{-B_{k_n}} e^{A_{k_n} + B_{k_n}}) \to tA$. □

Let $\mathcal{G}_0$ be the connected *component of the identity* in the closed group $\mathcal{G}$. By definition $\mathcal{G}_0$ is the set of all $\gamma \in \mathcal{G}$ which may be linked to the identity by a continuous curve in $\mathcal{G}$. Each such γ is a finite product of elements $\alpha_k = e^{A_k}$ with $A_k \in \mathfrak{g}$ and $\|\alpha_k - I\| < \varepsilon$.

Proposition 18.80. *Let $\mathcal{G}$ be a closed subgroup of* GL, *and $\mathcal{G}_0$ the connected component of the identity. The set $\mathcal{G}_0$ is a group. It is closed in* GL. *It is open in $\mathcal{G}$. It is a normal subgroup of $\mathcal{G}$.*

Proof. If $\xi(t)$ is a curve connecting id to α, then $\beta\xi(t)$ connects β to $\beta\alpha$. Hence one can reach $\beta\alpha$ if one can reach α and β. So $\mathcal{G}_0$ is a group. If $W = I + \varepsilon B$ satisfies $W \cap \mathcal{G}_0 = W \cap \mathcal{G}$ then this also holds for αW for any $\alpha \in \mathcal{G}_0$. So $\mathcal{G}_0$ is open in $\mathcal{G}$. Hence so are the left cosets, and so is the complement $\mathcal{G} \setminus \mathcal{G}_0$ as a union of left cosets. So $\mathcal{G}_0$ is closed in $\mathcal{G}$, and since $\mathcal{G}$ is closed in GL, the set $\mathcal{G}_0$ is closed in GL. Two elements α and α' lie in the same left coset, if and only if they may be connected. If $\alpha' \in \alpha\mathcal{G}_0$, and $\beta' \in \beta\mathcal{G}_0$ then $\alpha'\beta'$ may be connected to $\alpha\beta$, and hence $\alpha'\beta' \in \alpha\beta\mathcal{G}_0$. So $\mathcal{G}_0$ is the kernel of the quotient map $\alpha \mapsto \alpha\mathcal{G}_0$, which is a homomorphism. □

Corollary 18.81. *$\mathcal{G}_0$ is the Lie group generated by* $\mathfrak{g}$.

So closed subgroups of $\mathcal{A}$ and GL have a simple structure. They fall apart into a finite or countable collection of copies of a connected closed group $\mathcal{G}_0$, which is a manifold modelled on the Lie algebra $\mathfrak{g}$.

Example 18.82. The group $\mathcal{A}_0^h$ of all affine transformations which map H_+ onto itself consists of two connected parts diffeomorphic to $\mathbb{R}^h \times \mathbb{R}^h \times \mathrm{GL}^+(h) \times (0, \infty)$. ◊

A closed group may be isomorphic to a proper closed subgroup of itself: The group $2\mathbb{Z}$ of even integers is isomorphic to $\mathbb{Z}$. For compact groups in $\mathcal{A}$ this is not possible.

Proposition 18.83. *If $\mathcal{G}$ is a closed subgroup of $\mathcal{A}$ with finitely many connected components, and $\mathcal{H}$ a closed subgroup of $\mathcal{G}$, isomorphic to $\mathcal{G}$, then $\mathcal{H} = \mathcal{G}$.*

Proof. First assume $\mathcal{G}$ is connected. Then so is $\mathcal{H}$, and the Lie algebra $\mathfrak{h}$ is a linear subspace of $\mathfrak{g}$, and with the same dimension as $\mathfrak{g}$ (and $\mathcal{H}$ and $\mathcal{G}$). So $\mathfrak{h} = \mathfrak{g}$ and $\mathcal{H} = \mathcal{G}$ since each element of $\mathcal{G}$ is a finite product of elements in $e^{\mathfrak{h}}$. In general the isomorphism preserves the number of components. So $\mathcal{H}$ fills out $\mathcal{G}$. □

We conclude with a few words on homomorphisms. Excess measures use continuous homomorphisms $\Theta\colon \mathcal{G} \to (0, \infty)$. Such homomorphisms correspond to certain linear functions on the Lie algebra $\mathfrak{g}$.

Proposition 18.84. *For a Lie group $\mathcal{G}$ with Lie algebra $\mathfrak{g}$ a continuous homomorphism $\Theta\colon \mathcal{G} \to (0, \infty)$ has the form $\Theta(e^{tA}) = e^{t\theta A}$ for a linear function $\theta\colon \mathfrak{g} \to \mathbb{R}$.*

Proof. Define $\theta(A) = \log \Theta(e^A)$. Then $\theta(tA) = t\theta(A)$ since this holds for $t = -1$, $t = 1/m$, and in general for $t \in \mathbb{Q}$. Set $C_n = \log(e^{A/n + B/n})$. The power series of the exponential function gives $n(e^{A/n+B/n} - I) \to A + B$, and hence $nC_n \to A + B$. This gives

$$\theta A + \theta B = \log \Theta(e^A) + \log \Theta(e^B) = n \log(\Theta(e^{A/n} e^{B/n})) = n \log \Theta(e^{C_n})$$
$$= n\theta(C_n) = \theta(nC_n) \to \theta(A + B).$$

Hence $\theta A + \theta B = \theta(A + B)$. The equality holds for arbitrary A and B in $\mathfrak{g}$. □

Corollary 18.85. *If $\theta(A) = \log \Theta(e^A)$ then $\theta([A, B]) = 0$ for $A, B \in \mathfrak{g}$.*

Proof. Set $\varphi(t) = e^{tA} e^{tB} e^{-tA} e^{-tB}$. Then $\Theta\varphi(t) = 1$. Set $C_n = \log \varphi(1/\sqrt{n})$ for $n \geq n_0$. Then $\theta(C_n) = 0$, hence $\theta(nC_n) = 0$ and $nC_n \to [A, B]$ gives $\theta[A, B] = 0$. □

The results above also hold if Θ is only defined on a neighbourhood of the identity.

Proposition 18.86. *Let $A \in \mathfrak{g}$ have complex diagonal Jordan form with diagonal entries on the imaginary axis. If $\Theta\colon \mathscr{G} \to (0, \infty)$ is a continuous homomorphism then $\theta A = 0$.*

Proof. The closure $\mathscr{H}$ of the one-parameter group e^{tA}, $t \in \mathbb{R}$, is a compact group. Hence $\Theta(\mathscr{H})$ is a compact subgroup of $(0, \infty)$. This implies $\Theta(\mathscr{H}) = \{1\}$. □

Our description of Lie groups is ambivalent. Symmetry groups of excess measures are closed. So for our purpose it is natural to think of a Lie group as a connected closed subgroup of GL or $\mathcal{A}$. In the general theory one prefers to think of a Lie group as a manifold modeled on a Lie algebra. The exposition given above for the closed subgroups of GL is in the spirit of von Neumann [1929]. A simple introduction to matrix Lie groups is Curtis [1979]. More advanced texts on Lie groups are Bump [2004], Duistermaat & Kolk [2000] and Varadarajan [1974]. Lorente & Gruber [1972] gives a good impression of the complexity of the system of multidimensional subgroups of GL(d) for $d \leq 6$; Winternitz [2004] describes the subgroups of SL(3).

18.14 An example. In these notes three methods for handling extremes of multivariate samples have been developed. In each case one looks at exceedances. In each case there is a class of convex open sets which determines which points of the sample are to be termed extreme. Extreme points fall outside some convex set from this class. The convex sets are open halfspaces, open ellipsoids, and translates of the open negative orthant. The edge of the normalized sample cloud is described by a limiting Poisson point process. The mean measure of this Poisson point process is the *excess measure*. There are situations where all three approaches may be applied. Do the three approaches then yield the same excess measure?

A distribution may lie in the domain of attraction of many different excess measures. In the example below, high risk scenarios on halfspaces converge for every direction. The limit distribution depends on the direction; but not continuously. In addition to these directional limit laws for exceedances over linear thresholds, there is a limit law for exceedances over elliptic thresholds, a limit law for exceedances over translates of the negative orthant, and one for coordinatewise maxima. In our example the corresponding excess measures are different in each case.

We shall construct a probability distribution π on the plane, and excess measures ρ_i on open sets O_i for $i = 0, \dots, m$, with the same *scalar expansions*:

$$\gamma^t(\rho_i) = e^t \rho_i, \quad \gamma^t : w \mapsto e^t w, \quad t \in \mathbb{R},\ i = 0, \dots, m, \tag{18.28}$$

such that for $i = 0, \dots, m$

$$(e^t/t^i)\gamma^{-t}(\pi) \to \rho_i \ \text{ vaguely on } O_i, \quad t \to \infty, \tag{18.29}$$

and such that for any halfplane $J = \{au + bv \geq 1\}$, $a^2 + b^2 = 1$, the scaled high risk scenarios converge:

$$Z^{H_r}/r \Rightarrow W^J, \quad H_r = rJ,\ r \to \infty,$$

where W^J is a vector on J with density g^J. Convergence $J_n \to J$ does not imply $W^{J_n} \Rightarrow W^J$. It is possible that J_1 and J_2 intersect and

$$\mathbb{P}\{W^{J_2} \in J_1\} > 0, \quad \mathbb{P}\{W^{J_1} \in J_2\} = 0. \tag{18.30}$$

Example 18.87. Let $C_1 \supset \cdots \supset C_m$ be a strictly decreasing finite sequence of proper closed cones (sectors) in $\mathbb{R}^2$. We assume that $C_1 \cap [0, \infty)^2 = \{(0,0)\}$. Set $C_0 = \mathbb{R}^2$ and $C_{m+1} = \{(0,0)\}$. Let g_0 be a continuous positive function on $\mathbb{R}$, periodic modulo 2π. The density (in polar coordinates)

$$g(r, \varphi) = g_0(\varphi)/r^3, \quad r > 0,\ \varphi \in [0, 2\pi)$$

determines excess measures ρ_k, with densities $g_k = g 1_{C_k \setminus C_{k+1}}$, which satisfy (18.28). The probability distribution π of Z has density f, which, outside some large disk, has the form

$$f(r, \varphi) = (\log r)^k g(r, \varphi) \ \text{ on } C_k \setminus C_{k+1}, \quad k = 0, \dots, m.$$

It follows that (18.29) holds on $O_i = C_{i+1}^c$, and that W^J has density

$$g^J = g 1_{C_k \setminus C_{k+1}} / \rho_k(J),$$

where $k = k_J$ is the maximal index for which J intersects C_k. The two probabilities in (18.30) are positive only if $k_{J_1} = k_{J_2}$, and then $\rho_k(J_1) g^{J_1} = \rho_k(J_2) g^{J_2}$. The integer-valued function $J \mapsto k_J$ is not continuous, hence neither is the map from J to the distribution of W^J.

The vector Z has heavy tails. The asymptotics for exceedances over elliptic thresholds focuses on the global behaviour, and will see only the heaviest tails. The excess measure is $d\rho^\infty = 1_{C_m} d\rho_m$ on $\mathbb{R}^d \setminus \{0\}$. The vector Z conditioned to lie outside the ball rB converges for $r \to \infty$ in distribution to a vector W with distribution $1_{B^c} d\rho^\infty / \rho^\infty(B^c)$:

$$Z^r / r \Rightarrow W, \quad r \to \infty.$$

For *componentwise maxima* one looks at the sample points outside translates of the negative quadrant. For heavy tails it is customary to replace the vector Z by $Z^+ = (X \vee 0, Y \vee 0)$ on $[0, \infty)^2$. The part of π in $(-\infty, 0]^2$ does not contribute. The asymptotic behaviour is determined by the cones C_j which extend into the second or fourth quadrant. The limiting excess measure for exceedances over translates of the negative quadrant will be denoted by ρ^Q. This measure lives on

$$\mathcal{X} = [-\infty, \infty)^2 \setminus [-\infty, 0]^2.$$

The measure ρ^Q and the *exponent measure* $\rho^\vee$ on $[0, \infty]^2 \setminus \{0\}$ of the max-stable limit distribution are related: $\rho^\vee$ is the image of ρ^Q under the non-linear projection $z \mapsto z^+$. These limit measures are determined by the asymptotics for exceedances over horizontal and vertical thresholds. If C_1 lies in $(-\infty, 0]^2$ then ρ^Q is the restriction of ρ to $\mathcal{X}$. The exponent measure $\rho^\vee$ charges the positive quadrant and the two boundary halflines. If C_1 extends into the second or fourth quadrant then $\rho^\vee$ lives on the boundary of the positive quadrant, and, under appropriate diagonal normalization, the sample maxima converge to a limit vector with iid components, with *Fréchet* distribution $e^{-1/v^{1/\tau}}$ on $[0, \infty)$. The excess measure ρ^Q shows greater variety. It is the restriction of ρ_j to $\mathcal{X}$ if C_j extends into both the second and fourth quadrant and C_{j+1}ß$(-\infty, 0]^2$; otherwise it lives on two positive halflines, one vertical the other horizontal, one in zero the other in $-\infty$. If ρ^Q charges $\{-\infty\} \times (0, \infty)$ this just means that for any $\varepsilon > 0$ for points $W_{nk} = (U_{nk}, V_{nk})$ in the normalized sample with $V_{nk} > \varepsilon$ the horizontal coordinate U_{nk} tends to lie far out on the negative axis for large n. ◊

The example is rather artificial. Even in the simple case of heavy tails and a limit measure with scalar symmetries, the conditions for convergence in the case of $\mathcal{D}^h(\rho)$, $\mathcal{D}^\vee(\rho)$ and $\mathcal{D}^\infty(\rho)$ are so different that in general a probability distribution will typically lie in only one of these domains. This is clear if one looks at the normalizations which are allowed in the different situations. The class of normalizations is largest for $\mathcal{D}^\infty$. The normalizations for $\mathcal{D}^\infty$ need not preserve the coordinate axes nor the "horizontal" plane. The normalizations for $\mathcal{D}^h$ preserve the "horizontal" plane; those for $\mathcal{D}^\vee$ preserve the coordinate axes. On the other hand $\mathcal{D}^\infty$ looks at the global asymptotics, and hence convergence may be destroyed by perturbations of the distribution in directions which are not noticed in $\mathcal{D}^h$ or $\mathcal{D}^\vee$.

Now suppose the vector Z has non-negative components and *heavy tails*. Then $Z \in \mathcal{D}^\vee(\rho)$ implies $Z \in \mathcal{D}^\infty(\rho)$. If the d marginal tails $T_i = 1 - F_i$ are equal, then exceedances over linear thresholds will converge for all directions $\theta \in [0, \infty)^d$. In general convergence of exceedances over linear thresholds for one direction does not imply convergence for any other direction. If T_i / T_1 fails to have a positive finite limit in ∞ for one of the coordinates, then convergence over linear thresholds with direction $\theta \in (0, \infty)^d$ will fail too. We refer to Section 15.4 for a discussion on the relation between the domains $\mathcal{D}^\vee(\rho)$ and $\mathcal{D}^h(\rho)$.

One of our aims in writing these notes has been to clarify the relation between the different domains of attraction.

V Open problems

The reader will have noticed that there is considerable variation in the size of the individual lectures. For standard topics we have presented the basic material and given references for further reading. Where the material is new we have not hesitated to include less basic results. This may sometimes give the impression that the road splits into many smaller paths that wander off without a definite goal. That impression is not incorrect. In writing a book one has to decide what to include. Our aim was to give an impression of the state of the theory at this moment. By diminishing the number of examples and technical results we could have reduced the size of the book by a hundred pages. However, the book, like most books in mathematics, is not intended to be read, except for the Introduction, the Preview, and the introductory pages of the chapters and of some of the longer sections. It is directed at probabilists, statisticians and risk analysts who want to see whether particular problems with which they are confronted can be clarified, and perhaps even solved by using a more geometric approach. Such a person wants to know what results are available on a specific topic, say heavy tails with scalar symmetries and non-scalar normalizations. To find out she will have to read the introductory pages of Section 16, skim through the remainder of that section, and then zoom in on Subsection 17.2.

You are served a Preview, which opens up a new vision on multivariate extremes, and an Introduction, which, in a lighter vein, describes the relevance of this theory to financial mathematics and insurance and risk analysis. The main course is a novel theory of multivariate Generalized Pareto Distributions. Side dishes are a limit theory for exceedances over linear thresholds, and an extension of the classic univariate theory of exceedances for heavy tailed distributions to the multivariate setting. As a bonus the reader gets a nice introduction to Poisson point processes on open subsets of Euclidean space, and to univariate and coordinatewise multivariate extreme value theory. On top of that we offer you an introduction to univariate and multivariate regular variation, including the Meerschaert Spectral Decomposition Theorem, to Lie groups of affine transformations and the Jordan form in linear algebra, and to multivariate stable distributions and Lévy processes. We zoom in on a number of special situations which may be of particular interest to risk theory: light tails in all directions, or only in one particular direction, heavy tails with the same rate of decrease in all directions, or with different rates of decrease.

As the topics become more specific, the presentation becomes more technical. This can not be avoided. It is hoped that the detailed subdivision of the material, the use of starred subsections to indicate non-main stream material, the many examples illustrating unexpected behaviour, the extensive index, and the back-references in the bibliography will help to open up the book to the serious investigator.

The foregoing chapters gave a description of work in progress. There we told what we know about high risk scenarios; in this chapter we talk about what we do not know. We shall discuss open problems in the analysis of high risk scenarios, both from the probabilistic background of the models, and in the statistical analysis of data in terms of these models.

The numbers at the end of each item refer to sections which contain relevant material.

19 The stochastic model

1) Are the multivariate *GPD*s the only limit laws for high risk scenarios? One can show that these are the only limit laws in dimension $d = 2$, and the only cylinder symmetric limit laws for any dimension. The proofs are rather algebraic. [13.1]

2) An XS-measure is an excess measure ρ, with halfspace J_0 of unit mass, whose symmetry group $\mathcal{G}$ is so large that $\mathcal{G}(J_0)$ is open in $\mathcal{H}$. Examples are the multivariate *GPD*s, and *Lebesgue measure* on the quadratic cone $\{v > \|u\|\}$ or on $(0, \infty)^d$. Also the heavy-tailed densities $1_{Q^c}(u, v)/(v + u^T u/2)^{(d+1)/2+\lambda}$ with $\lambda > 0$ and $Q = \{v \le -u^T u/2\}$. These measures have the *tail property* to an extreme degree. Determine the class of all *XS-measures*, and their domains of attraction. There are seven classes of cylinder symmetric XS-measures, see Balkema [2006]. [12.2; 14.1]

3) *The domain of the heavy tailed multivariate GPDs.* Let ρ have density $1/\|w\|^{d+1/\tau}$ with $\tau > 0$. Then $Z \in \mathcal{D}^\infty(\rho)$ implies $Z \in \mathcal{D}(\tau)$ by a two-step conditioning, first on the complement of an ellipsoid, then on a halfspace supporting this ellipsoid. Does the inclusion $\mathcal{D}(\tau)$ß$\mathcal{D}^\infty(\rho)$ hold? [12.1; 17.2]

4) Do *unimodal* densities describe the global behaviour of distributions in the domain of the *Gauss-exponential* law? Suppose π lies in the domain of attraction of the Gauss-exponential law. Does there exist a density $f \in \mathrm{U}_0$ which converges with the same normalizations as π, such that $d\pi = f d\mu$ where μ is a *roughening of Lebesgue measure* for f? The measure π determines the normalizations α_H. So we may use (8.11) to determine the asymptotic value of f in the point $z = \alpha_H(0)$. The function $f \circ \alpha_H / f(z)$ will be close to $w \mapsto e^{-(u^T u/2+v)}$ on bounded sets. This suggests that it might be possible to construct a continuous function f in U_0 by pasting pieces of the Gauss-exponential density together. [11.4]

5) *Rates of convergence.* How does the rate of convergence behave with increasing dimension?

   High risk limit laws do not lie in their own domain of attraction. However there are simple spherical distributions in the domain of attraction of the multivariate *GPD*s. The standard *Gauss distribution* lies in the *Gauss-exponential* domain, the *spherical Student* distributions lie in the domains of the heavy tailed limits, and the spherical beta distributions lie in the domains of the bounded limit vectors. For these classical densities one may compute the $\mathbf{L}^1$ distance δ between the density of the normalized high risk scenario and of the limit vector. Choose the optimal normalization. The number δ then depends on the dimension d, on the risk level p, and on the shape parameter $\tau \in [-1/2h, \infty)$. The dependence on the dimension is not crucial. For $\tau = 0$ there is no dependence on d. With hindsight, this is not surprising, since the horizontal coordinate already has the correct limit distribution. [8.3]

6) *The domain of the bounded multivariate GPDs.* Describe $\mathcal{D}(\rho)$ where ρ is *Lebesgue measure* on the paraboloid $v < -u^T u/2$. [12.2]

7) Many of our results are on $\mathbf{L}^1$ convergence of densities, rather than weak convergence of probability measures. What is the bonus for this stronger form of convergence? [9.4; 10.3]

8) *Skeletons for densities in the Gauss-exponential domain.* Let $f = e^{-\varphi} \in \mathcal{D}(0)$ be *unimodal*. The increasing sequence of bounded convex sets $\{\varphi < n\}$ determines the asymptotic behaviour of f in the same way in which a df $F \in \mathcal{D}^+(0)$ is determined by a sequence $y_n \uparrow y_\infty$ for which $1-F(y_n) \sim e^{-n}$. Such a sequence is a *discrete skeleton*.

   In the multivariate case the sets $\{\varphi < n+1\} \cap H_n$, where H_n are half-spaces supporting the level set $\{\varphi < n\}$, are asymptotically *parabolical caps*. For any decreasing sequence of such halfspaces H_n, the hyperplanes ∂H_n are asymptotically equidistant. Are these conditions sufficient to ensure $f \in \mathcal{D}(0)$?

   In the univariate setting a sequence of positive reals $a_0, a_1, \ldots$ with $a_{n+1}/a_n \to 1$ determines a skeleton $(y_n)$for given y_0 by the relation $y_n = y_{n-1} + a_n$. Hence up to asymptotic equality (and a translation) the sequence (a_n) determines the tail of a df in the domain of the exponential limit law. In the multivariate setting a similar constructive approach to the domain of the Gauss-exponential limit law might be developed using skeletons consisting of increasing sequences of bounded open strictly convex sets O_n containing the origin, and having a C^1 boundary ∂O_n, as in the definition of U_0 in Section 10.1.

   In particular this approach might enable us to answer the following questions:

i) Given a sequence of rotund sets D_n does there exist a function $f \in \mathrm{U}_0$ and increasing sequences $k_n \to \infty$ and $r_n \to \infty$ such that $\{f > e^{-k_n}\} = r_n D_n$?

ii) Given an increasing sequence of balls $B_n = z_n + r_n B$ with center z_n and radius $r_n > \|z_n\|$, what conditions on the centers z_n and radii r_n ensure that there exists a function $f \in \mathrm{U}_0$ such that all level sets $\{f > c\}$ are balls, and $\{f > e^{-n}\} = B_n$?

iii) Does there exist an $f \in \mathrm{U}_0$ such that $\{f > 0\}$ is an open simplex, an open paraboloid, or some other given open convex set? [9.4; 11.4]

9) How close can the distribution get to the exponent measure in the theory of max-stable distributions? Let R be the df of an exponent measure $\rho \in \mathcal{MSE}$. Does there exist a copula C such that

$$C(e + z) = 1 - R(z), \quad e = (1, \ldots, 1),\ z \in (-\infty, 0]^d,\ \|z\| < \varepsilon?$$

10) For $\tau \to 0+$ the heavy tailed multivariate *GPD* on H_+ with shape parameter τ (properly normalized) tends to the *Gauss-exponential* distribution. For the domains of attraction there seems to be a discontinuity in $\tau = 0+$. Nice densities in the domain of the Gauss-exponential law have rotund *level sets*, but in the domain of the multivariate *Euclidean Pareto* distribution the level sets $\{f > c\}$ are elliptic for $c \to 0+$. For heavy tails the distribution has to satisfy a structural condition (asymptotically elliptic distribution); for light tails the condition is a local smoothness condition (continuously varying positive curvature). What happens for $\tau > 0$ and small? Are the asymptotics based on $\tau = 0$ valid as a reasonable approximation in this situation? [13.1]

11) For heavy tailed distributions a similar discontinuity in the domains $\mathcal{D}^\infty(\rho)$ occurs. Take a spectral measure with a fixed non-constant density on the unit circle. Let ρ_n be the excess measure with this spectral measure and with generator $C_n = \mathrm{diag}(1, 1 + 1/n)$. In the limit the generator is scalar and distributions in $\mathcal{D}^\infty(\rho)$ may be twisted. [16.1; 18.4; 17.2]

12) Suppose ρ_1 and ρ_2 are full measures on $\mathbb{R}^d$, α_1 and α_2 affine transformations, and $C_1, C_2 > 1$ such that $\alpha_i(\rho_i) = C_i \rho_i$ for $i = 1, 2$. If ρ_1 and ρ_2 agree outside some bounded set, does this imply that $\rho_1 = \rho_2$? [18.10]

13) What can one say about the convex hull C_n of the sample cloud N_n from a distribution in the *Gauss-exponential* domain? If the distribution has a density Lf with $f \in \mathrm{U}_0$ and L flat for f then the convex hull locally may be described by the convex hull of the Poisson point process M with intensity $e^{-(u^T u/2+v)}$. The convex hull of M is roughly parabolic. What can one say about the global behaviour of the convex hulls C_n?

Define the core sets K_n for a distribution π as the intersection of all halfspaces H of mass $\pi(H) \geq 1 - 1/n$. How much does the convex hull C_n differ from the core set? How many points of the sample cloud N_n fall outside the core set K_n? In any boundary point $p \in \partial K_n$ one may introduce coordinates so that p is the origin and $n\pi$ is close to the Gauss-exponential measure ρ. So one can measure how far out the points of N_n outside the core set are. What can one say about the fluctuations in this landscape? Global peaks are simple. What about the valleys?

There are some results for spherically symmetric distributions. It would be interesting to see whether these hold when the condition of *spherical symmetry* is relaxed. A clear exposition of the basic ideas in the bivariate situation is given in Nagaev [1995]. The results in Hueter [1999] are very general; the proofs rather sketchy. See Finch & Hueter [2004] for more references. [11.5; 5.7]

14) *The relation to coordinatewise extreme value theory.* For heavy tails the relation between coordinatewise extremes and exceedances over elliptic thresholds is clear. For light tails there are many dark areas. In particular the domains of exponent measures with exponential marginals have received no attention in these notes. [Preview; 15.3; 15.4; 17.3]

15) If the vector Z has exponentially thin tails, the moment generating function M exists, and one may use exponential tilting to emphasize the measure in certain directions, replacing $d\pi$ by $d\pi^\zeta(z) = e^{\zeta z} d\pi(z)/M(\zeta)$. Not much is known about the limit theory in this situation. See Barndorff-Nielsen & Klüppelberg [1999] and Wiedmann [1997] for multivariate Gaussian limits, and Balkema, Klüppelberg & Resnick [2001] and Nagaev & Zaigraev [2000] for more general limit distributions. We defined a scenario as a change of measure. So the Esscher transform π^ζ may be treated as a high risk scenario. If one allows ζ to diverge in any direction the basic theory is similar to that for high risk scenarios: finite-dimensional families of limit laws with large symmetry groups. Note though that these scenarios are only defined for distributions π with thin tails. [15.2]

16) What happens if one conditions on orthants $[z, \infty)$, $z \in \mathbb{R}^d$, rather than halfspaces, and assumes that $\mathbb{P}\{Z \geq z\} \to 0$? See Balkema & Qi [1998].

17) Roughening Lebesgue measure should not affect weak convergence. $\mathbf{L}^1$ convergence on H_+ is preserved if we multiply $f \in \mathrm{U}_0$ by a flat function. In Balkema & Embrechts [2004] it is shown that weak convergence is preserved if we roughen Lebesgue measure, provided the density is rotund-exponential. Does this also hold for densities in U_0? [10]

18) Suppose $f \in \mathrm{U}_0$ and L is flat for f. Is Lf asymptotic to an element $h \in \mathrm{U}_0$? [10.1]

19) It is simple to generate random vectors from the *uniform distribution* on a *rotund* set. By viewing *rotund-exponential* densities as mixtures of scaled copies of such distributions, random vectors from rotund-exponential densities may be simulated in a two-step procedure. How does one efficiently simulate high risk scenarios from such a density? [9.2]

20) Suppose our sample comes from a *unimodal* density in the domain of the *Gauss-exponential* law. We have estimated a number of rotund level sets, say at the levels c, $c/2$ and $c/4$. How do we fill in the tails? For rotund-exponential densities the level sets are scaled copies of a fixed rotund set, and this is no problem. We then only have to estimate a univariate tail. But if the shape of the level sets changes, or the barycenter of the sets, it is not clear what the next *level set* (at level $c/8$, say) should look like. [11.4]

21) How much variation is possible in the tails in different directions for *unimodal* densities in the domain of the *Gauss-exponential* high risk limit law? [11.4]

22) Pancheva [1985] and Mohan & Ravi [1992] have looked at convergence of non-negative random variables under power norming:

$$U = e^b X^a$$

with $a > 0$ and $b \in \mathbb{R}$. This corresponds to positive affine normalizations of the logarithm of X. For exceedances over horizontal thresholds of random vectors $Z = (X, Y) \in \mathbb{R}^{h+1}$ one may perform such power scaling surreptitiously by an initial non-linear transformation $\tilde{Z} = (X, \log Y)$. In how far does this procedure allow us to generalize the limit theory for heavy tails developed in Section 16?

23) A measure ρ on an open set Oß$\mathbb{R}^d$ is *sign-invariant* if $-O = O$ and if

$$\iota(\rho) = \rho \quad \text{where } \iota(w) = -w.$$

If ρ is sign-invariant then so is $T(\rho)$ for any linear transformation T. If ρ is a Radon measure on $O = \mathbb{R}^d \setminus \{0\}$ then $\rho^o = (\rho + \iota(\rho))/2$ is sign-invariant. A measure ρ^* on the unit sphere ∂B in $\mathbb{R}^d$ is *standard* if it is a probability measure with zero expectation and *covariance* matrix I/d. The *uniform distribution* on ∂B is standard. Suppose ρ is an excess measure for scalar expansions. Then so is $T(\rho)$ for any linear transformation T. Can one choose T so that the spectral measure of $(T\rho)^o$ is standard? [16.1]

24) Describe the behaviour of the convex hull C_n of an n-point sample cloud from a density $f = Le^{-\psi \circ n_D}$ as in Theorem 9.1 with L flat. Do there exist constants $c_n > 0$ such that $C_n/c_n \to D$ a.s.? [11.1]

25) There exists an elegant theory of weak convergence for increasing functions on an open interval and for multivariate dfs. Develop a similar theory for multivariate *unimodal distributions*. [10.1]

26) In $\mathbb{R}^\infty$ one may define a standard Gaussian distribution, a *Gauss-exponential* distribution and a Gauss-exponential point process. Develop the corresponding infinite-dimensional theory of *GPD*s. (For $\tau < 0$ there is no infinite-dimensional theory because of the bound $\tau \geq -2/(d-1)$.) [13.1]

27) Develop a high risk theory for exceedances beyond cones or cylinders (rather than ellipsoids, or halfspaces). [18.8; 17.7; 16]

28) We have concentrated on symmetry groups with simple generators, multiples of the identity or diagonal matrices. In these cases there is a basis of eigenvectors in $\mathbb{R}^d$. Complex eigenvalues may be represented as rotations in real coordinates. Generators which do not have a basis of complex eigenvectors have non-zero off-diagonal elements in their Jordan form. In dimension $d = 2$ such generators yield a group of *shear* transformations along the one-dimensional eigenspace. An investigation of the domain of attraction for shear expansions in $\mathbb{R}^2$ would be of interest. [17]

29) *Regular variation.* Suppose $\beta\colon [0, \infty) \to \mathcal{A}$ is continuous and varies like e^{tC}. How much information do we need to reconstruct the curve β up to asymptotic equality? If the generator C is given it suffices to know the sequence $\beta(n)$. If the γ^t are linear expansions it suffices to know the linear part of $\beta(t)$; if the γ^t are translations in $\mathbb{R}$, and $\beta(t)(v) = a(t)v + b(t)$, it suffices to know the function $t \mapsto b(t)$, or the curve $\{(b(t), a(t)) \mid t \geq 0\}$ as a subset of $\mathbb{R}^2$. [18.1; 15.5]

30) *Regular variation.* Suppose $\alpha_t(w) = A_t w + a_t$ varies like γ^t, and γ is linear. It may be possible to replace α by $\beta \sim \alpha$ where β is linear. This is the case if γ is a linear expansion (with all eigenvalues outside the unit circle in $\mathbb{C}$). In that case $A_t^{-1}\alpha_t \to \mathrm{id}$. If all eigenvalues of γ lie inside the unit circle then $a_t \to a$ and $\alpha_t \sim \beta_t$ where $\beta_t(w) = A_t(w - a) + a$.

    Take a geometric point of view, and think of γ^t, $t \in \mathbb{R}$, as a one-parameter group of affine transformations on $\mathbb{R}^d$, and of α_t as maps from $\mathbb{R}^d$ to the d-dimensional affine space L. The vectors Z and the high risk scenarios Z^H live on L. They are normalized by affine maps $\alpha^{-1}\colon L \to \mathbb{R}^d$. Under what conditions can one choose coordinates in L, and $\beta \sim \alpha$, so that in these coordinates the $\beta(t)$ belong to the same subgroup as the γ^t? [18.4; 18.1]

31) Meerschaert & Scheffler [2001], Theorem 6.1.19, give a Convergence of Types Theorem for excess measures for exceedances over elliptic thresholds. Does there exist a similar result for exceedances over *horizontal thresholds*? [Preview]

32) Give a geometric description of the curve of ellipsoids, $t \mapsto E_t = \alpha(t)(B)$, where α varies like γ^t. For scalar expansion groups, $\gamma^t(w) = e^{\tau t}w$, there is a simple criterium, (19) in the Preview for regular variation of ellipsoids. [17.2]

33) Consider a continuous density $f(x, y) = f_y(x)\tilde{f}(y)$ in the domain $\mathcal{D}^h(\rho)$ of the *Gauss-exponential* measure ρ with the property that the conditional densities f_y are Gaussian. Such a *typical density* may be constructed, starting from a sequence $b_n = b_0 + a_1 + \cdots + a_n$ with $a_{n+1} \sim a_n > 0$, a sequence of centered ellipsoids $E_n \sim E_{n-1}$, and a sequence of vectors $\mu_n = \mu_0 + \nu_1 + \cdots + \nu_n$ in $\mathbb{R}^h$ where $\varphi_{n+1} - \varphi_n = o(E_n/a_n)$ for $\varphi_n = \nu_n/a_n$. What conditions on the sequence (a_n, E_n, ν_n) will ensure that the convex hulls converge? For what measures μ is the limit relation preserved, and convergence of the convex hulls, if we replace $f d\lambda$ by $f d\mu$? [15.2; 16.7]

34) Typical densities exist for exceedances over horizontal thresholds, exceedances over elliptic thresholds, and for exceedances beyond translates of the negative orthant, as for coordinatewise maxima. Now replace horizontal thresholds by thresholds that are asymptotically parallel. In algebraic terms this means that $\beta\colon [0, \infty) \to \mathcal{A}$ is a continuous curve which varies like γ^t in $\mathcal{A}^h$. We drop the condition that $\beta(t) \in \mathcal{A}^h$. We seem to run into difficulties here if we try to construct typical densities. [18.8]

35) Explore *tail self-similar* limit laws. Here one replaces the one-parameter group of symmetries in the definition of an excess measure by a single symmetry relation: $\gamma(\rho) = a\rho$ for some $a > 1$. Obviously this implies $\gamma^k(\rho) = a^k\rho$ for all $k \in \mathbb{Z}$. The theory is simple (as long as there is only one symmetry), and for distributions in the domain of attraction of a tail self-similar distribution the recipe in the Preview will work. All one has to do is replace the spectral measure by a period of the tail self-similar distribution. Regular variation reduces to the study of sequences $\alpha_n = \alpha_0\gamma_1 \dots \gamma_n$ with $\gamma_n \to \gamma$. In the univariate situation the measure ρ with mass $1/2^{n+1}$ in the points $n \in \mathbb{Z}$ restricted to $[0, \infty)$ yields a geometric distribution. Such measures occur as exponent measures for maxima when one considers subsequences, the maximum of r_n independent observations, where $r_{n+1}/r_n \to 2$. There is a well-developed theory in the univariate case, but there are few applications. See Pancheva [1992] for basic results on max-semistability, Temido & Canto e Castro [2002] for extension to stationary sequences, and van den Brandhof [2008] for a nice application.

The corresponding theory for sums of random vectors has been developed in MS, Section 6.2 and 7.1. [14.4]

36) Develop the theory of regular variation and excess measures on infinite-dimensional linear spaces. See Hult & Lindskog [2006]. [18.1; 18.7]

37) If ρ is an excess measure on $\mathbb{R}^d \setminus \{0\}$ for linear expansions, can one choose coordinates such that the unit ball B is *adapted* and also invariant under the linear transformations which preserve ρ? [16.2]

20 The statistical analysis

The statistician in extreme value theory is faced with three interrelated problems:

1) she is only interested in the tail behaviour of the distribution;

2) she has to estimate an infinite measure;

3) this measure ρ has a continuous non-compact group of symmetries.

Let us briefly discuss each of these three points:

1) Since she is only interested in the asymptotic behaviour of the tail – she may be asked to use the sample to estimate probabilities of regions which lie outside the convex hull of the sample cloud – the statistician starts by throwing away at least 90% of the sample, in the univariate case all points below a certain threshold. The crucial question then becomes: How does one determine the threshold? In the univariate case the threshold is a real number, but in the multivariate case the situation is more complex. In general she deletes all sample points in a convex set: a halfspace, a lower orthant, or an ellipsoid.

This procedure raises a more basic issue. Should the statistician give equal weight to all points in the sample? By deleting a number of sample points the answer is clearly no. Do there exist convincing arguments for giving each of the remaining sample points the same weight? There is a variance-bias argument: The statistical advantage of an increase in sampling diversity if the sample point just above the threshold were to share some of its mass with the sample point just below the threshold may outweigh the slight decrease in dependability.

2) In the theory of coordinatewise maxima the statistician has to estimate the exponent measure; in the theory of exceedances she has to estimate the excess measure. These measures are infinite Radon measures ρ on open subsets Oß$[-\infty, \infty)^d$.

For the exponent measure there are two approaches. She may estimate the df $R(z) = \rho([-\infty, z]^c)$, or she may estimate the measure ρ itself in its upper endpoint together with the $2^d - 2$ lower dimensional marginals. See Proposition 7.9. In the case of exceedances over horizontal thresholds it suffices to estimate the probability measure $d\rho^J = 1_J d\rho/\rho(J)$ where J is a suitable horizontal halfspace; for exceedances over elliptic thresholds the recipe is similar. For the measures associated with multivariate GPDs the situation is different.

3) The infinite Radon measure ρ on O above has a one-parameter group of symmetries:

$$\gamma^t(\rho) = e^t\rho, \quad t \in \mathbb{R}. \tag{20.1}$$

There exists a base set AßO, a finite measure ρ^* on a Borel set A_0ß∂A, and a homeomorphism $T\colon O \to A_0 \times \mathbb{R}$ such that as in Section 18.9

1) $T(\rho) = \rho^* \times e^{-t}dt$ is a product measure on $A_0 \times \mathbb{R}$;

2) orbits map into vertical lines: $T(\gamma^t(z)) = (T(z), t)$, $z \in A_0$;

3) A is the inverse image of the halfspace $A_0 \times [0, \infty)$.

The measure ρ^* is called the *spectral measure*. Since one may replace the Radon measure ρ by a positive multiple of itself, one may assume that the spectral measure ρ^* is a probability measure, the distribution of a vector U^*. Given the one-parameter group γ^t, $t \in \mathbb{R}$, and the section A_0, the spectral measure determines the infinite Radon measure ρ. For exceedances over horizontal thresholds the base set A is the horizontal halfspace J_0 and A_0 the bounding horizontal hyperplane. For exceedances over ellipsoids A is the complement of the unit ball and A_0 the unit sphere – in appropriate coordinates. The spectral measure depends on the choice of coordinates. For the exponent measures associated with coordinatewise maxima of heavy-tailed vectors several base sets have been suggested: The complement of a cube, a ball, or a simplex. There is no canonical choice.

Excess measures (the Radon measures associated with exceedances) have the advantage over exponent measures (associated with maxima) that the limit distribution coincides with the excess measure on the halfspace J_0 on which it lives.

The limit distribution for exceedances over horizontal or elliptic thresholds lies in its own domain of attraction. If the sample comes from the limit distribution, and she knows the generator of the stability group γ^t, $t \in \mathbb{R}$, she may project the sample points onto a horizontal hyperplane (or an elliptic boundary) along orbits of the stability group, to obtain a sample from the spectral measure ρ^*. Essentially she is in the situation where one has to estimate a spherically symmetric density. The representation $W = \gamma^T(U^*, 0)$ transforms the data points W_i into pairs (U_i^*, T_i) with independent components. Since the variable T is standard exponential it suffices to estimate the distribution ρ^* of U^*.

For high-dimensional data sets the reduction of the dimension from d to h is not impressive. The real gain is that U^* may have any distribution on $\mathbb{R}^h$ (as long as the distribution of W is non-degenerate).

If the underlying distribution is not the limit distribution, but lies in the domain of attraction, then the statistician faces the same problems as in the univariate case.

In addition she has to determine the form of the generator of the stability group, see Section 14.9, and the values of the parameters in these matrices. Finally she has to estimate the spectral distribution. It is not clear whether these tasks may be carried out successively, if she wants good estimates.

Estimating the multivariate *GPD*s is a different matter. On the one hand she may consider exceedances over thresholds in a particular direction. The spectral measure then is known up to a finite number of parameters, so that the statistician is in the classical situation of a finite-dimensional parameter space. On the other hand she may consider convergence of the high risk scenarios in all directions. Spherical distributions are replaced by distributions based on the gauge function of a rotund set. Between the local and the global approach there is a range of questions which the statistician may have to decide. Assume the spectral measure is Gaussian. Possible questions are: Is convergence robust under slight changes in the direction of the halfspace? For which halfspaces does the Gauss-exponential limit still apply? Does the convex hull of the normalized sample cloud converge to the convex hull of the limiting Gauss-exponential point process? See Example 9.14.

We now list a number of concrete problems which the statistician may encounter in applying the limit theory for high risk scenarios as sketched in Chapters III and IV. As before the square brackets at the end refer to relevant subsections in the text.

1) Exhibit real life data sets which fit the theory. [8.5]

2) Provide synthetic data sets for a given rotund-exponential density. [9]

3) How does one associate a rotund set with a sample cloud? A bivariate sample cloud may give a clear visual impression of an underlying spherical or elliptic distribution, but it may also suggest that the level sets are egg-shaped rather than elliptic. How does one construct the egg? [9.2]

4) *Choice of the threshold.* In the theory of extremes we want to estimate the mean measure of the limiting Poisson point process. This measure satisfies certain symmetry relations. Suppose it is a mixture of one-dimensional *Lebesgue measure* on rays. Let A denote the set above the threshold. The statistician restricts the sample to this set. With each of the k points x in the subsample she associates a multiple $c(x)/k$ of Lebesgue measure on the ray through x where $1/c(x)$ is the length of the interval which A cuts off from the ray. What conditions on A will ensure that this positive linear combination of one-dimensional Lebesgue measures converges to ρ in probability? [7.4]

5) There is a growing expertise on the statistical analysis of distributions in the domain of max-stable distributions. The results mentioned in the literature should be applicable to the theory for exceedances over horizontal thresholds.

As an example we mention *sample copulas*. These describe the relative rank of the coordinates of the points in the sample cloud, see Section 7.3. For samples from a df in the domain of a max-stable law one may form the sample copula – an n-point subset of $\{1, \dots, n\}^d$ which projects onto $\{1, \dots, n\}$ in each of the d coordinates. Together with the univariate marginal samples the sample copula determines the original sample. In the limit the normalized sample cloud from a distribution in the domain of a max-stable law with standard exponential marginals on $(-\infty, 0)$ converges to a Poisson point process with standard Poisson marginals on $(-\infty, 0)$, a so-called *Poisson copula*. The associated sample copulas shifted to the negative orthant converge to the corresponding simple point process on $\{\dots, -2, -1\}$. How much information is lost by using the sample copula to estimate the copula of the max-stable limit law? This problem has recently been solved (under some extra conditions). The loss does not vanish in the limit, but is of the same order of magnitude as the statistical error due to sampling. Both may be described by a Gaussian process based on a common Wiener process. See Einmahl, de Haan & Li [2006]. [7.5]

6) For a multivariate thin tailed distribution how does one decide that the level sets are rotund, but not elliptical? [9.2]

7) For a heavy-tailed distribution how does one decide that the tails in all directions have the same exponent? [16.1]

8) How does one choose the origin for heavy tailed distributions? Suppose the distribution is a mixture of a Cauchy distribution, spherically symmetric around an unknown point ζ_0, and an unknown asymmetric light tailed or bounded distribution. How does one determine ζ_0? For the asymptotics it does not matter where one chooses the origin; for finite samples it does make a difference. In the univariate case the average of the upper and lower quartile might be a good estimate. Is there a general procedure for distributions with *balanced* heavy tails? [16.1]

9) How does one choose the coordinates? In the heavy-tailed situation one can perform an initial affine coordinate transformation to change the elliptic cloud into a spherical cloud. Now cut out all sample points within a ball of radius r_0, leaving only a fraction of the original sample. Project this subsample radially (or rather along orbits) onto the unit sphere and check whether the sample of unit vectors is centered and has a *covariance* matrix which is a multiple of the identity. If not, how does one improve on the initial affine transformation? [16.1; 8.5]

10) How does one test for *unimodality*? And how does one estimate the unimodal density? [10.1]

11) If the excess measure ρ is heavy tailed and symmetric for scalar expansions there are three models for the domain of attraction, depending on the normalizations allowed: scalar normalizations, diagonal matrices or arbitrary matrices. For extremes one starts by throwing away the central points, thus reducing the sample size by a factor ten. So the simple model of scalar expansions should do unless one has a huge data set. What is the advantage of the larger models? [17]

12) How does one handle *contamination*? A contamination of π may be defined as a mixture of two probability measures $(1-p)\pi + p\tilde{\pi}$ with $0 < p \ll 1$, where $\tilde{\pi}$ is more heavy tailed than π. In particular π may live on a cone and $\tilde{\pi}$ on the complement, or π may have exponential tails, whereas $\tilde{\pi}$ has heavy tails in certain directions.

    In the univariate case sea level distributions determined by storms, currents and tides may be contaminated by tsunamis, or by flood waves caused by meteorite impacts. Life time distributions may be contaminated by pandemics or wars. Credit risk distributions may be contaminated by economic breakdown due to political revolutions. In the multivariate case we refer to Example 9.14, and the discussion in Subsection 15.4. In robust statistics contamination is seen in outliers. The task of the statistician is to detect and delete such anomalies in the sample. In risk theory outliers have to be treated with extreme care. Do we deal with errors which may cause some jitter but have no long term effect, or do the outliers indicate real risks? In the latter case one may try to incorporate the outliers in the basic model, or stick to the light tailed model, and embed that in a larger model in which shocks (earthquakes, wars and system breakdowns) are also taken into account. See Nešlehová, Embrechts & Chavez-Demoulin [2006] and Dell'Aquila & Embrechts [2006]. [15.3]

Bibliography

The numbers in square brackets at the end of each item refer to the pages on which the respective work is cited.

Aczél, J. (1969), *On applications and theory of functional equations*. Birkhäuser, Basel. [297] 297

Agrachev, A. A. and Y. L. Sachkov (2004), *Control theory from the geometric viewpoint*. Encyclopaedia Math. Sci. 87, Springer-Verlag, Berlin. [298] 298

Alink, S. (2007), *Copulas and extreme values*. PhD Thesis, University of Nijmegen. [111] 111

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999), Coherent measures of risk. *Math. Finance* **9**, 203–228. [6]

Balkema, A. A. (1973, *Monotone transformations and limit laws*. PhD Thesis, Mathematical Centre Tracts 45. Mathematisch Centrum, Amsterdam. [2, 297] 297

Balkema, A. A. and L. de Haan (1974), Residual life time at great age. *Ann. Probability* **2**, 792–804. [101] 101

Balkema, A. A., C. Klüppelberg, and S. I. Resnick (2001), Stability for multivariate exponential families. *J. Math. Sci.* **106**, 2777–2791. [352] 352

Balkema, A. A. and Y.-C. Qi (1998), Strictly stable laws for multivariate residual lifetimes. *J. Math. Sci.* **92**, 3873–3880. [176, 352] 176, 352

Balkema, G. (2006), Cylinder symmetric measures with the tail property. *Publ. Instit. Math.* (*Beograd*) **80 (94)**, 7–27. [158, 316, 349] 158, 316, 349

Balkema, G. and P. Embrechts (2004), Multivariate excess distributions. http://www.math.ethz.ch/ baltes/ftp/guuspe08Jun04.pdf [123, 142, 169, 352] 123, 142, 169, 352

Banner, A., E. Fernholz, and I. Karatzas (2006), Atlas models for equity markets. *Ann. Appl. Probab.* **15**, 2296-2330. [8]

Barbour, A. D. and O. Chryssaphinou (2001), Compound Poisson approximation: a user's guide. *Ann. Appl. Probab.* **11**, 964–1002. [57] 57

Barbour, A. D., L. Holst, and S. Janson (1992), *Poisson approximation*. Oxford Stud. Probab. 2, The Clarendon Press, Oxford. [57] 57

Barndorff-Nielsen, O. E. and C. Klüppelberg (1999), Tail exactness of multivariate saddlepoint approximations. *Scand. J. Statist.* **26**, 253–264. [352] 352

Baryshnikov, Y. (2000), Supporting-points processes and some of their applications. *Probab. Theory Related Fields* **117**, 163–182. [36, 129] 129

Basel Committee on Banking Supervision (2005), Amendment to the capital accord to incorporate market risks. Revised version. Bank for International Settlements, Basel. [10]

Basrak, B., R. A. Davis, and T. Mikosch (2002), A characterization of multivariate regular variation. *Ann. Appl. Probab.* **12**, 908–920. [267] 267

Bedford, T. and R. Cooke (2001), *Probabilistic risk analysis: foundations and methods*. Cambridge University Press, Cambridge. [30]

Beirlant, J., Y. Goegebeur, J. Teugels, and J. Segers (2004), *Statistics of extremes*. John Wiley & Sons, Chichester. [100] 100

Bellman, R. (1969), *Stability theory in differential equations*. Dover Publications, New York. [298] 298

Bensalah, Y. (2002), Asset allocation using extreme value theory. Preprint, Bank of Canada. [10]

Berman, S. M. (1992), *Sojourns and extremes of stochastic processes*. Wadsworth & Brooks, Pacific Grove, CA. [36]

Billingsley, P. (1966), Convergence of types in k-space. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **5**, 175–179. [14]

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989), *Regular variation*. Encyclopedia Math. Appl. 27, Cambridge University Press, Cambridge. [158, 170, 297] 158, 170, 296

Bingham, N. H. and R. Kiesel (2002), Semi-parametric modelling in finance: theoretical foundations. *Quant. Finance* **2**, 241–250. [5]

Bloom, S. (1976), A characterization of B-slowly varying functions. *Proc. Amer. Math. Soc.* **54**, 243–250. [108] 108

Boyd, S. and L. Vandenberghe (2004), *Convex optimization*. Cambridge University Press, Cambridge. [86] 86

Boyle, P. and F. Boyle (2001), Derivatives. The tools that changed finance. Risk Waters Group, London 2001. [4]

Bräker, H., T. Hsing, and N. H. Bingham (1998), On the Hausdorff distance between a convex set and an interior random convex hull. *Adv. in Appl. Probab.* **30**, 295–316. [36]

Brandhof, A. van den (2008), Limit laws for the width of binary trees. [355] To appear. 355

Brouwer, L. E. J. (1911), Über Abbildung von Mannigfaltigkeiten. *Math. Ann.* **71**, 97–115; reprinted in *Collected Works*, vol. 2, 454–472, 1976. Elsevier, Amsterdam. [142] 142

Brozius, H. and L. de Haan (1987), On limiting laws for the convex hull of a sample. *J. Appl. Probab.* **24**, 852–862. [36, 172] 172

Bump, D. (2004), *Lie groups*. Graduate Texts in Math. 225, Springer-Verlag, New York. [344] 344

Cabo, A. J. and P. Groeneboom (1994), Limit theorems for functionals of convex hulls. *Probab. Theory Related Fields* **100**, 31–55. [129, 175] 129, 175

Chavez-Demoulin, V., P. Embrechts, and J. Nešlehová (2006), Quantitative models for operational risk: extremes, dependence and aggregation. *J. Banking Finance* **30**, 2635–2658. [6]

Cochrane, J. (2001), *Asset Pricing*. Princeton University Press, Princeton, NJ. [4]

Coles, S. (2001), *An introduction to statistical modeling of extreme values*. Springer-Verlag, London. [9, 100] 100

Coles, S. G. and J. A. Tawn (1991), Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53**, 377–392. [9]

Coles, S. and J. Tawn (1994), Statistical methods for multivariate extremes: an application to structural design. *J. Roy. Statist. Soc. Ser. C* **43**, 1–48. [21, 36] 21

Curtis, M. L. (1979), *Matrix groups*. Springer-Verlag, New York. [338, 344] 338, 344

Daley, D. J. and D. Vere-Jones (2003), *An introduction to the theory of point processes*. Vol. I: Elementary theory and methods. 2nd ed., Springer-Verlag, New York. [47, 53] 47

Davis, R. and S. Resnick (1986), Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533–558. [36]

Dees, S., F. di Mauro, M. Pesaran, and L. Smith (2007), Exploring the international linkages of the Euro area: a global VaR analysis. *J. of Appl. Ecconom.* **22**, 1–38. [8]

Deheuvels, P. (1981), Multivariate tests of independence. In *Analytical methods in probability theory* (Oberwolfach, 1980), Lecture Notes in Math. 861, Springer-Verlag, Berlin, 42–50. [116] 116

Dell'Aquila, R. and P. Embrechts (2006), Extremes and robustness: a contradiction? *Financ. Markets Portfolio Manage.* **20**, 103–118. [360] 360

Demarta, S. (2007), *The copula approach to modelling multivariate extreme values*. PhD Thesis, ETH Zurich. [111] 111

Dharmadhikari, S. and K. Joag-Dev (1988), *Unimodality, convexity, and applications*. Academic Press, Boston. [147] 147

Dieker, A. (2006), Extremes and fluid queues. PhD Thesis, University of Amsterdam. [36]

Drees, H. (2003), Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**, 617–657. [111] 111

Dugundji, J. (1966), *Topology*. Allyn and Bacon, Boston. [338] 338

Duistermaat, J. J. and J. A. C. Kolk (2000), *Lie groups*. Springer-Verlag, Berlin. [338, 344] 338, 344

Eddy, W. F. (1980), The distribution of the convex hull of a Gaussian sample. *J. Appl. Probab.* **17**, 686–695. [27]

Eddy, W. F. and J. D. Gale (1981), The convex hull of a spherically symmetric sample. *Adv. in Appl. Probab.* **13**, 751–763. [36]

Einmahl, J. H. J., L. de Haan, and D. Li (2006), Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist.* **34**, 1987–2014. [116, 359] 116, 359

Einstein, A. (1906), Zur Theorie der Brownschen Bewegung. *Ann. Physik* (4) **19**, 371–381; translated as Chapter II in Investigations on the Theory of Brownian Motion, Dover 1956. [30]

Embrechts, P. (1979), Karakterisatiestellingen en Subexponentiële Verdelingsfuncties: Toepassingen in Kanstheorie. PhD Thesis, KU Leuven. [2]

Embrechts, P. (2004), Extremes in economics and the economics of extremes. In *Extreme values in finance, telecommunications, and the environment*, (Eds Finkenstädt and Rootzén). Chapman & Hall/CRC, Boca Raton, FL, 169–183. [100] 100

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997), *Modelling extremal events for insurance and finance*. Appl. Math. 33, Springer-Verlag, Berlin. [14, 39, 100, 103, 104, 113, 126] 100

Embrechts, P., A. J. McNeil, and D. Straumann (2002), Correlation and dependence in risk management: properties and pitfalls. In *Risk management: value at risk and beyond* (Cambridge, 1998), (Ed. M.A.H. Dempster). Cambridge University Press, Cambridge, 176–223. [5]

Falk, M., J. Hüsler, and R.-D. Reiss (2004), *Laws of small numbers: extremes and rare events*. 2nd ed., Birkhäuser, Basel. [57, 100] 57, 100

Feller, W. (1971), *An introduction to probability theory and its applications*. Vol. II, 2nd edition, John Wiley & Sons, New York. [296] 296

Fergusson, K. and E. Platen (2006), On the distributional characterization of log-returns of a world stock index. *Appl. Math. Finance* **13**, 19–38. [8]

Fernholz, E. R. (2002), *Stochastic portfolio theory*. Appl. Math. 48, Springer-Verlag, New York. [4, 8]

Finch, S. and I. Hueter (2004), Random convex hulls: a variance revisited. *Adv. in Appl. Probab.* **36**, 981–986. [36, 129, 352] 129, 352

Finkenstädt, B. and H. Rootzén (eds.) (2004), *Extreme values in finance, telecommunications, and the environment*. Monographs Statist. Appl. Probab. 99, Chapman & Hall/CRC, Boca Raton, FL. [36, 100] 100

Fisher, R. A. and L. H. C. Tippett (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.* **24**, 180–190. [100] 100

Fisz, M. (1954), A generalization of a theorem of Khintchin. *Studia Math.* **14**, 310–313. [14]

Fougères, A.-L. (2004), Multivariate extremes. In *Extreme values in finance, telecommunications, and the environment*, (Eds. Finkenstädt and Rootzén), Chapman & Hall/CRC, Boca Raton, FL, 373–388. [9]

Galambos, J. (1978), *The asymptotic theory of extreme order statistics*. Wiley, New York. [111, 116, 122] 111, 116, 122

Geluk, J. L. and L. de Haan (1987), *Regular variation, extensions and Tauberian theorems*. CWI Tract 40, Stichting Mathematisch Centrum voor Wiskunde en Informatica, Amsterdam. [108, 297] 108, 297

Giné, E., M. G. Hahn, and P. Vatan (1990), Max-infinitely divisible and max-stable sample continuous processes. *Probab. Theory Related Fields* **87**, 139–165. [36]

Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.* (2) **44**, 423–453. [101] 101

Greenspan, A. (1996), Remarks on risk measurement and systemic risk. In *Risk measurement and systemic risk: proceedings of a joint central bank research conference, November 16-17, 1995*. Board of Governors of the Federal Reserve System (U.S.), issue Nov., Washington, DC, 11–16. [2]

Groeneboom, P. (1988), Limit theorems for convex hulls. *Probab. Theory Related Fields* **79**, 327–368. [36]

Haan, L. de (1970), *On regular variation and its application to the weak convergence of sample extremes*. Mathematical Centre Tracts 32, Mathematisch Centrum, Amsterdam. [101, 303] 101, 303

Haan, L. de (1985), Extremes in higher dimensions: the model and some statistics. In *Proceedings of the 45th session of the International Statistical Institute, Vol. 4* (Amsterdam, 1985), *Bull. Inst. Internat. Statist.* **51**, 185–192. [36]

Haan, L. de and A. Ferreira (2006), *Extreme value theory. An introduction*. Springer Ser. Oper. Res. Financ. Eng., Springer-Verlag, New York 2006. [36, 100, 111, 231] 100, 111, 231

Haan, L. de and T. Lin (2003), Weak consistency of extreme value estimators in $C[0,1]$. *Ann. Statist.* **31**, 1996–2012. [36]

Haan, L. de and E. Omey (1984), Integrals and derivatives of regularly varying functions in $\mathbf{R}^d$ and domains of attraction of stable distributions. II. *Stochastic Process. Appl.* **16**, 157–170. [24]

Haan, L. de and S. Resnick (1987), On regular variation of probability densities. *Stochastic Process. Appl.* **25**, 83–93. [24]

Haan, L. de and S. I. Resnick (1977), Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **40**, 317–337. [111] 111

Haan, L. de and J. de Ronde (1998), Sea and wind: multivariate extremes at work. *Extremes* 1, 7–45. [36]

Hall, P. and N. Tajvidi (2000), Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* **6**, 835–844. [111] 111

Halmos, P. R. (1950), *Measure Theory*. D. Van Nostrand Company, New York. [341] 341

Hirsch, M. W. and S. Smale (1974), *Differential equations, dynamical systems, and linear algebra*. Pure Appl. Math. 60, Academic Press, New York. [321, 334] 321, 334

Huang, X. (1992), *Statistics of bivariate extreme values.* Tinbergen Institute Research Series 22, Thesis Publishers, Amsterdam. [111] 111

Hueter, I. (1999), Limit theorems for the convex hull of random points in higher dimensions. *Trans. Amer. Math. Soc.***351**, 4337–4363. [146, 352] 146, 352

Hult, H. and F. Lindskog (2002), Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. in Appl. Probab.* **34**, 587–608. [9]

Hult, H. and F. Lindskog (2005), Extremal behavior of regularly varying stochastic processes. *Stochastic Process. Appl.* **115**, 249–274. [36]

Hult, H. and F. Lindskog (2006), Regular variation for measures on metric spaces. *Publ. Inst. Math.* (*Beograd*) **80 (94)**, 121–140. [356] 356

Hult, H. and F. Lindskog (2006a), On Kesten's counterexample to the Cramér-Wold device for regular variation. *Bernoulli* **12**, 133–142. [267] 267

Jacobsen, M. (2006), *Point process theory and applications*. Birkhäuser, Boston. [47] 47

Jessen, A. and T. Mikosch (2006), Regularly varying functions. *Publ. Inst. Math.* (*Beograd*) **80 (94)**, 171–192. [297] 297

Joe, H. (1997), *Multivariate models and dependence concepts*. Monogr. Statist. Appl. Probab. 73, Chapman & Hall, London. [100, 111, 116] 100, 111, 116

Jurek, Z. J. and J. D. Mason (1993), *Operator-limit distributions in probability theory*. John Wiley & Sons, New York. [298] 298

Kallenberg, O. (2002), *Foundations of modern probability*. 2nd ed., Springer-Verlag, New York. [47, 55, 76, 284] 47, 55, 76, 284

Karr, A. (1991), *Point processes and statistical inference*. 2nd ed., Marcel Dekker, New York. [47] 47

Kingman, J. F. C. (1993), *Poisson processes*. Oxford Stud. Probab. 3, The Clarendon Press, Oxford. [47] 47

Korn, R. (1997), *Optimal portfolios. Stochastic models for optimal investment and risk management in continuous time*. World Scientific, Singapore. [4]

Kotz, S. and S. Nadarajah (2000), *Extreme value distributions*. Imperial College Press, London. [100] 100

Kupiec, P. (1998), Stress testing in a value at risk framework. *J. Derivatives* **6**, 7–24. [10]

Kuratowski, C. (1948), *Topologie. I. Espaces métrisables, espaces complets*. 2nd ed., Monografie Matematyczne 20, Warszawa. [71] 71

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983), *Extremes and related properties of random sequences and processes*. Springer-Verlag, New York. [36]

Lindskog, F. (2004), Multivariate extremes and regular variation for stochastic processes. PhD Thesis, ETH Zurich. [9]

Longin, F. (2000), From value at risk to stress testing: the extreme value approach. *J. Banking Finance* **24**, 1097–1130. [10]

Lorente, M. and B. Gruber (1972), Classification of semisimple subalgebras of simple Lie algebras. *J. Math. Phys.* **13**, 1639–1663. [344] 344

Lüthi, H.-J. and G. Studer (1997), Factors at risk. Preprint, RiskLab and IFOR, ETH Zurich. [4]

Malevergne, Y. and D. Sornette (2006), *Extreme financial risks*. Springer-Verlag, Berlin. [100] 100

Marcus, M. and M. Pinsky (1969), On the domain of attraction of $e^{-e^{-x}}$. *J. Math. Anal. Appl.* **28**, 440–449. [101] 101

McNeil, A. J., R. Frey, and P. Embrechts (2005), *Quantitative risk management. Concepts, techniques and tolls*. Princeton University Press, Princeton, NJ. [4, 5, 10, 19, 30, 100, 111, 116, 162] 100, 111, 116, 162

Meerschaert, M. M. and H.-P. Scheffler (2001), *Limit distributions for sums of independent random vectors*. John Wiley & Sons, New York. [21, 22, 39, 226, 231, 234, 259, 297, 306, 310, 355] 231, 259, 306, 355

Meerschaert, M. M. and J. A. Veeh (1993), The structure of the exponents and symmetries of an operator stable measure. *J. Theoret. Probab.* **6**, 713–726. [316] 316

Meerschaert, M. M. and J. A. Veeh (1995), Symmetry groups in d-space. *Statist. Probab. Lett.* **22**, 1–6. [341] 341

Mikosch, T. (2005), How to model multivariate extremes if one must? *Statist. Neerlandica* **59**, 324–338. [172, 264] 172, 264

Mikosch, T. (2006), Copulas: Tales and facts. *Extremes* **9**, 3–20. [119] 119

Moix, P.-Y. (2001), *The measurement of market risk*. Lecture Notes in Econom. and Math. Systems 504, Springer-Verlag, Berlin. [100] 100

Mohan, N. R. and S. Ravi (1992), Max domains of attraction of univariate and multivariate p-max stable laws. *Teor. Veroyatnost. i Primenen.* 37, 709–721; English transl. *Theory Probab. Appl.* **37**, 632–643. [117, 353] 117, 353

Mulder, H. (1977), De ei-kromme. *Pythagoras* **17**, 12–15; 70. [139] 139

Nagaev, A. V. (1995), Some properties of convex hulls generated by homogeneous Poisson point processes in an unbounded convex domain. *Ann. Inst. Statist. Math.* **47**, 21–29. [352] 352

Nagaev, A. and A. Zaigraev (2000), Abelian theorems for a class of probability distributions in $\mathbb{R}^d$ and their application. *J. Math. Sci.* **99**, 1454–1462. [352] 352

Nešlehová, J., P. Embrechts, and V. Chavez-Demoulin (2006), Infinite mean models and the LDA for operational risk. *J. Operational Risk* **1**, 3–25. [6, 360] 360

Neumann, J. von (1929), Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen. *Math. Z.* 30, 3–42. [344] 344

Pancheva, E. (1985), Limit theorems for extreme order statistics under nonlinear normalization. In *Stability problems for stochastic models* (Uzhgorod, 1984), Lecture Notes in Math. 1155, Springer-Verlag, Berlin, 284–309. [353] 353

Pancheva, E. (1988), Max-stability. *Teor. Veroyatnost. i Primenen.* **33**, 167–170; *Theory Probab. Appl.* **33**, 155–158. [121] 121

Pancheva, E. (1992), Multivariate max-semistable distributions. *Teor. Veroyatnost. i Primenen.* **37**, 679–684; *Theory Probab. Appl.* **37**, 731–732. [355] 355

Parthasarathy, K. R. (1967), *Probability measures on metric spaces*. Probab. Math. Statist. 3, Academic Press, New York. [43, 76] 43, 76

Perfekt, R. (1997), Extreme value theory for a class of Markov chains with values in $\mathbf{R}^d$. *Adv. in Appl. Probab.* **29**, 138–164. [36]

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004), Modeling regional interdependencies using a global error-correcting macroeconometric model. *J. Bus. Econom. Statist.* **22**, 129–162. [8]

Pesaran, M. H. and R. Smith (2006), Macroeconomic modelling with a global perspective. *The Manchester School* **74**, 24–49. [8]

Pesaran, M. H. and P. Zaffaroni (2004), Model averaging and value-at-risk based evaluation of large multi asset volatility models for risk management. CESifo Working Paper Series No. 1358; IEPR Working Paper No. 04.3. Available at SSRN: http://ssrn.com/ abstract=642681 [8]

Pickands, III, J. (1975), Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131. [101] 101

Platen, E. (2001), A minimal financial market model. In *Mathematical finance* (Konstanz, 2000), Trends Math., Birkhäuser, Basel, 293–301. [8]

Platen, E. (2006), A benchmark approach to finance. *Math. Finance* **16**, 131–151. [8]

Poincaré, H. (1881), Mémoire sur les courbes définies par une équation différentielle, I. *J. Math. Pures Appl.* (3) **7**, 375–422; reprinted in *Œuvres* I (1928) 3–44. [142] 142

Reiss, R.-D. (1993), *A course on point processes*. Springer-Verlag, New York. [47] 47

Rényi, A. and R. Sulanke (1963), Über die konvexe Hülle von n zufällig gewählten Punkten. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **2**, 75–84. [175] 175

Resnick, S. I. (1987), *Extreme values, regular variation, and point processes*. Springer-Verlag, New York. [20, 100] 100

Resnick, S. I. (2004), On the foundations of multivariate heavy-tail analysis. *J. Appl. Probab.* **41A**, 191–212. [111, 113, 231, 298] 111, 113, 231, 298

Resnick, S. I. (2006), *Heavy-tail phenomena*. Springer Ser. Oper. Res. Financ. Eng., Springer-Verlag, New York. [36, 231] 231

Robinson, J. J. (1989), A program for generating samples of a random tesselation of the plane associated with convex hulls. MSc Thesis in Mathematics, University of Amsterdam. [129] 129

Ronan, M. (2006), *Symmetry and the monster*. Oxford University Press, Oxford. [10]

Rootzén, H. and N. Tajvidi (2006), Multivariate generalized Pareto distributions. *Bernoulli* **12**, 917–930. [298] 298

Salvadori, G., C. De Michele, N. T. Kottegoda, and R. Rosso (2007), *Extremes in nature. An approach using copulas*. Water Science and Technology Library 56, Springer-Verlag, Berlin. [100] 100

Smith, R. L., J. A. Tawn, and S. G. Coles (1997), Markov chain models for threshold exceedances. *Biometrika* **84**, 249–268. [36]

Studer, G. (1997), Maximum loss for measurement of market risk. PhD Thesis, IFOR, ETH Zurich. http://www.ifor.math.ethz.ch/publications/diss_studer [4]

Studer, G. and H.-J. Lüthi (1997), Quadratic maximum loss for risk measurement of portfolios. In *VAR: Understanding and applying value-at-risk*, Risk Publications, London, 307–316. [4]

Tajvidi, N. (1995), Multivariate generalised Pareto distributions. Report, Mathematical Statistics, Chalmers University. http://www.math.chalmers.se/ rootzen/papers/mgpd.pdf [176] 176

Temido, M. G. and L. Canto e Castro (2002), Max-semistable laws in extremes of stationary random sequences. *Teor. Veroyatnost. i Primenen.* **47**, 402–410; English translation *Theory Probab. Appl.* **47**, 365–374. [355] 355

Valentine, F. A. (1964), *Convex sets*. McGraw-Hill, New York. [165] 165

Varadarajan, V. S. (1974), *Lie groups, Lie algebras, and their representations*. Prentice-Hall, Englewood Cliffs, NJ. [330, 344] 330, 344

Walschap, G. (2004), *Metric structures in differential geometry*. Graduate Texts in Math. 224, Springer-Verlag, New York. [142] 142

Wiedmann, M. (1997), Asymptotiken für Dichten mit Gaußschen Tails in höheren Dimensionen. PhD Thesis, Universität Ulm. [352] 352

Winternitz, P. (2004), Subalgebras of Lie algebras. Example of $\mathfrak{sl}(3, \mathbb{R})$. In *Symmetry in physics* (Montréal, Canada, 2002), CRM Proc. Lecture Notes 34, Amer. Math. Soc., Providence, RI, 215–227. [344] 344

Wüthrich, M. V. (2003), Asymptotic value-at-risk estimates for sums of dependent random variables. *Astin Bull.* **33**, 75–92. [266] 266

Zmarrou, H. (2004), A probabilistic measure of risk. MSc Thesis in Mathematics, University of Amsterdam. [130, 131] 130, 131

Index

Page numbers in bold face lead to the formal definition.