Institute of Mathematical Statistics
LECTURE NOTES–MONOGRAPH SERIES

High Dimensional Probability

Proceedings of the Fourth International Conference

Evarist Giné, Vladimir Koltchinskii, Wenbo Li, Joel Zinn, Editors

Volume 51

ISBN-13: 978-0-940600-67-6
ISBN-10: 0-940600-67-6
ISSN 0749-2170

Institute of Mathematical Statistics

LECTURE NOTES–MONOGRAPH SERIES

Volume 51

High Dimensional Probability

Proceedings of the Fourth International Conference

Evarist Giné, Vladimir Koltchinskii, Wenbo Li, Joel Zinn, Editors

Institute of Mathematical Statistics
Beachwood, Ohio, USA

Institute of Mathematical Statistics
Lecture Notes–Monograph Series

Series Editor:
R. A. Vitale

The production of the *Institute of Mathematical Statistics Lecture Notes–Monograph Series* is managed by the IMS Office: Jiayang Sun, Treasurer and Elyse Gustafson, Executive Director.

Library of Congress Control Number: 2001012345

International Standard Book Number 978-0-940600-67-6,

0-940600-67-6

International Standard Serial Number 0749-2170

Printed in the United States of America

Contents

APPLICATIONS OF EMPIRICAL PROCESSES

Preface

About forty years ago it was realized by several researchers that the essential features of certain objects of Probability theory, notably Gaussian processes and limit theorems, may be better understood if they are considered in settings that do not impose structures extraneous to the problems at hand. For instance, in the case of sample continuity and boundedness of Gaussian processes, the essential feature is the metric or pseudometric structure induced on the index set by the covariance structure of the process, regardless of what the index set may be. This point of view ultimately led to the Fernique-Talagrand majorizing measure characterization of sample boundedness and continuity of Gaussian processes, thus solving an important problem posed by Kolmogorov. Similarly, separable Banach spaces provided a minimal setting for the law of large numbers, the central limit theorem and the law of the iterated logarithm, and this led to the elucidation of the minimal (necessary and/or sufficient) geometric properties of the space under which different forms of these theorems hold. However, in light of renewed interest in Empirical processes, a subject that has considerably influenced modern Statistics, one had to deal with a non-separable Banach space, namely $\mathcal{L}_\infty$. With separability discarded, the techniques developed for Gaussian processes and for limit theorems and inequalities in separable Banach spaces, together with combinatorial techniques, led to powerful inequalities and limit theorems for sums of independent bounded processes over general index sets, or, in other words, for general empirical processes.

This research led to the introduction or to the re-evaluation of many new tools, including randomization, decoupling, chaining, concentration of measure and exponential inequalities, series representations, that are useful in other areas, among them, asymptotic geometric analysis, Banach spaces, convex geometry, nonparametric statistics, computer science (e.g. learning theory).

The term High Dimensional Probability, and Probability in Banach spaces before, refers to research in probability and statistics that emanated from the problems mentioned above and the developments that resulted from such studies.

A large portion of the material presented here is centered on these topics. For example, under limit theorems one has represented both the theoretical side as well as applications to Statistics; research on dependent as well as independent random variables; Lévy processes as well as Gaussian processes; U and V-processes as well as standard empirical processes. Examples of tools to handle problems on such topics include concentration inequalities and stochastic inequalities for martingales and other processes. The applications include classical statistical problems and newer areas such as Statistical Learning theory.

Many of the papers included in this volume were presented at the IVth International Conference on High Dimensional Probability held at St. John's College, Santa Fe, New Mexico, on June 20-24, 2005, and all of them are based on topics covered at this conference. This conference was the fourteenth in a series that began with the Colloque International sur les Processus Gaussiens et les Distributions Aléatoires, held in Strasbourg in 1973, continued with nine conferences on Probability in Banach Spaces, and four with the title of High Dimensional Probability. The book *Probability in Banach Spaces* by M. Ledoux and M. Talagrand, Springer-Verlag 1991, and the Preface to the volume *High Dimensional Probability III*, Birkhäuser, 2003, contain information on these precursor conferences. More historical information can be found online at http://www.math.udel.edu/˜wli/hdp/index.html. This last reference also includes a list of titles of talks and participants of this meeting.

The participants to this conference are grateful for the support of the National Science Foundation, the National Security Agency and the University of New Mexico.

July, 2006

Evarist Giné
Vladimir Koltchinskii
Wenbo Li
Joel Zinn

Contributors to this volume

Arcones, M. A., *Binghamton University*

Berthet, P., *Université Rennes 1*
Borisov, I. S., *Sobolev Institute of Mathematics*
Bystrov, A. A., *Sobolev Institute of Mathematics*

Deheuvels, P., *L.S.T.A., Université Pierre et Marie Curie (Paris 6)*
Dobrić, V., *Lehigh University*
Dudley, R. M., *Massachusetts Institute of Technology*

Eggermont, P. P. B., *University of Delaware*

Figueroa-López, J. E., *Purdue University*

Giné, E., *University of Connecticut*

Houdré, C., *Georgia Institute of Technology*
Hush, D., *Los Alamos National Laboratory*

Koltchinskii, V., *Georgia Institute of Technology*

LaRiccia, V. N., *University of Delaware*

Mason, D. M., *University of Delaware*
Mendelson, S., *ANU & Technion I.I.T*

Nickl, R., *Department of Statistics, University of Vienna*

Ojeda, F. M., *Universidad Simón Bolívar and Lehigh University*

Peligrad, M., *University of Cincinnati*
Pinelis, I., *Michigan Technological University*

Scovel, C., *Los Alamos National Laboratory*
Steinberg, S., *University of New Mexico*
Steinwart, I., *Los Alamos National Laboratory*

Umarov, S., *University of New Mexico*
Utev, S., *University of Nottingham*

Vershynin, R., *University of California, Davis*

Wu, D., *Department of Statistics*
Wu, W. B., *University of Chicago*

Xiao, Y., *Probability, Michigan State University*

Zinn, J., *Texas A&M University*

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 1–17

DOI: 10.1214/074921706000000725

Stochastic integrals and asymptotic analysis of canonical von Mises statistics based on dependent observations

Igor S. Borisov[1,*] **and Alexander A. Bystrov**[1,†]

Sobolev Institute of Mathematics

Abstract: In the first part of the paper we study stochastic integrals of a nonrandom function with respect to a nonorthogonal Hilbert noise defined on a semiring of subsets of an arbitrary nonempty set.

In the second part we apply this construction to study limit behavior of canonical (i.e., degenerate) Von Mises statistics based on weakly dependent stationary observations.

1. Stochastic integrals of non-random kernels for non-orthogonal noises.

1.1. Introduction and statement of the main result.

Let $\{\Omega, \Theta, \mathbf{P}\}$ be a probability space, $\mathfrak{X}$ be an arbitrary nonempty set, and $\mathfrak{M}$ be a semiring with identity of its subsets (i.e., $\mathfrak{X} \in \mathfrak{M}$ and, for all $A, B \in \mathfrak{M}$, we have $A \cap B \in \mathfrak{M}$ and $A \setminus B = \sum_{i \le n} C_i$, where $C_i \in \mathfrak{M}$). We call a random process $\{\mu(A),\ A \in \mathfrak{M}\}$ *an elementary stochastic measure or a noise* if $\mu(A) \in L_2(\Omega, \mathbf{P})$ for all $A \in \mathfrak{M}$ (i.e., $\mu(\cdot)$ is a *Hilbert process*) and (N1) $\quad \mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$ a.s. if only $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 \in \mathfrak{M}$.

A noise μ is called *orthogonal* if
(N2) $\quad \mathbf{E}\mu(A_1)\mu(A_2) = m_0(A_1 \cap A_2)$,
where m_0 a finite measure (*the structure function*) on $\sigma(\mathfrak{M})$ [14].

Typical Examples.

(i) Consider the following semiring of subsets of a closed interval $[0, T]$:

$$\mathfrak{M} = \{(t, t+\delta];\ 0 < t < t+\delta \le T\} \bigcup \{[0, \delta];\ 0 < \delta \le T\}.$$

A random process $\xi(t)$ defined on $[0, T]$ generates the noise

$$\mu((t, t+\delta]) := \xi(t+\delta) - \xi(t),$$

where, in the case $t = 0$, this formula defines the measure of the closed interval $[0, \delta]$. If $\xi(t)$ is a process with independent increments then μ is an orthogonal noise.

[1]630090, Novosibirsk, Russia, acad. Koptyug avenue, 4, Sobolev Institute of Mathematics, e-mail: sibam@math.nsc.ru; bystrov@ngs.ru

*Supported in part by the Russian Foundation for Basic Research Grants 05-01-00810 and 06-01-00738 and INTAS Grant 03-51-5018.

†Supported in part by the Russian Foundation for Basic Research Grant 06-01-00738.

AMS 2000 subject classifications: primary 60H05, 60F05; secondary 62G20.

Keywords and phrases: stochastic integral, nonorthogonal noise, canonical U-statistics, canonical von Mises statistics, empirical process, dependent observations, ψ-mixing.

(ii) To construct multiple stochastic integrals the semiring $\mathfrak{M}^k := \mathfrak{M} \times \cdots \times \mathfrak{M}$ is considered, where $\mathfrak{M}$ is defined above, and the following *multiple noise* is defined by increments of a random process $\xi(t)$:

$$\mu((\bar{t}, \bar{t} + \bar{\delta}]) = \prod_{i \le k} (\xi(t_i + \delta_i) - \xi(t_i)),$$

where $(\bar{t}, \bar{t} + \bar{\delta}] = (t_1, t_1 + \delta_1] \times \cdots \times (t_k, t_k + \delta_k]$.
It worth noting that, in general, in the second example the noise μ does not satisfy condition (N2) even if the process $\xi(t)$ has independent increments.

We note some significant results in the area under consideration:

I. *Univariate stochastic integrals based on orthogonal noises*:
N. Wiener, 1923.
A. N. Kolmogorov, H. Cramér, 1940.
I. I. Gikhman and A. V. Skorokhod, 1977.

II. *Multiple stochastic integral with a multiple noise generated by a process with independent increments*:
N. Wiener, 1938, 1958.
K. Itô, 1951.
P. Major, 1981.

III. *Univariate stochastic integral with a noise generated by increments of a Hilbert process on the real line* (*nonorthogonal noise*):
M. Loève, 1960.
S. Cambanis and S. Huang, 1978.
V. Pipiras and M. S. Taqqu, 2000.

IV. *Multiple stochastic integral with a multiple noise generated by increments of a Gaussian process on the real line* (*nonorthogonal noise*):
S. Cambanis and S. Huang, 1978.
A. Dasgupta and G. Kallianpur, 1999.

General Case. We begin to study stochastic integrals with nonorthogonal noises defined on semirings of subsets of an arbitrary measurable space. We follow the generality considered in [14], where general stochastic integrals with orthogonal noises were studied. Complete proofs of some statements in the first Section of the paper are published in [4].

Introduce the function $m(A \times B) := E\mu(A)\mu(B)$ indexed by elements of $\mathfrak{M}^2$.

Main Assumption. The function m is a finite σ-additive signed measure (covariance measure) on $\mathfrak{M}^2$.

Example. Let $\Phi(t, s) := \mathbf{E}\xi(t)\xi(s)$ be the covariance function of a centered Hilbert random process $\xi(t)$ defined on a closed interval $[0, T]$. We say that the function $\Phi(t, s)$ possesses a bounded variation if, for a constant C,

$$\sup_{\{t_i, s_i\}} \sum_{i,j} |\Delta\Phi(t_i, s_j)| \le C,$$

where $\Delta\Phi(t_i, s_i) := \Phi(t_{i+1}, s_{j+1}) + \Phi(t_i, s_j) - \Phi(t_{i+1}, s_j) - \Phi(t_i, s_{j+1})$ (*the double difference*); $0 = t_0 < t_1 < \cdots < t_n = T$, $0 = s_0 < s_1 < \cdots < s_l = T$ are arbitrary finite partitions of the interval $[0, T]$, and the supremum is taken over all such partitions (see also [9]).

Proposition 1. *If $\Phi(t,s)$ has a bounded variation then the Main Assumption for the corresponding covariance measure is valid*

It is well known that any measure of such a kind can be uniquely extended onto $\sigma(\mathfrak{M}^2)$. Moreover, $m = m^+ - m^-$ (Hahn–Jordan decomposition), where m^+ and m^- are nonnegative finite measures. Put $|m| := m^+ + m^-$ (*the total variation measure*).

Introduce the space of $\sigma(\mathfrak{M})$-measurable functions:

$$S := \{f : \int_{\mathfrak{X}^2} f(t)f(s)m(dt \times ds) < \infty\}.$$

For any $\sigma(\mathfrak{M})$-measurable functions $f, g \in S$ consider the bilinear symmetric functional

$$d(f,g) := \int_{\mathfrak{X}^2} f(t)g(s)m(dt \times ds).$$

It is clear that $d(f,f) \geq 0$. But, in general, the equation $d(f,f) = 0$ has not only zero solution. Denote $\|f\| := d(f,f)^{1/2}$ (seminorm). If S is the factor space w.r.t. to the condition $d(f,f) = 0$ then S is an Euclidean space. But, in general, S may be incomplete (i.e., is not Hilbert) ([22]).

Notice that the space S can also defined in such a way:

$$S = \{f : \int_{\mathfrak{X}^2} |f(t)f(s)||m|(dt \times ds) < \infty\}.$$

But the functional $\|f\|_* := \left(\int_{\mathfrak{X}^2} |f(t)f(s)||m|(dt \times ds)\right)^{1/2}$ may not satisfy the triangle inequality ($\|f\|_*$ is a seminorm iff $|m|$ is *nonnegatively defined*).

For an orthogonal noise μ with a structure function m_0 the following obvious equality chain is valid: $\|f\| = \|f\|_* = \|f\|_{L_2(\mathfrak{X},m_0)}$.

Theorem 1. *Let $f \in (S, \|\cdot\|)$. Then there exists a sequence of step functions*

$$f_n(x) := \sum_{k \leq n} c_k I(x \in A_k), \tag{1}$$

where $c_k \in \mathbb{R}, A_k \in \mathfrak{M}$, converging in $(S, \|\cdot\|)$ to f as $n \to \infty$. Moreover, the sequence

$$\eta(f_n) := \sum_{k \leq n} c_k \mu(A_k) \tag{2}$$

mean-square converges to a limit random variable $\eta(f)$ which does not depend on the sequence of step functions $\{f_n\}$.

Proof. (For detail see [4]). Let $\{f_n\}$ be a sequence of step functions of the form (1) converging to f in the seminorm $\|\cdot\|$. One can prove that the sequences of such a kind exist. Then

$$\|\eta(f_n) - \eta(f_k)\|^2_{L_2(\Omega,\mathbf{P})} = \sum_{i,j} (c_i^{(n)} - c_i^{(k)})(c_j^{(n)} - c_j^{(k)}) m(A_i \times A_j)$$
$$\equiv \|f_n - f_k\|^2 \to 0$$

as $n, k \to \infty$ due to the triangle inequality (i. e., $\{f_n\}$ is a Cauchy sequence in S). Without loss of generality, we may assume here that the step functions f_k and f_n are defined on a common partition $\{A_i\}$. Hence $\{\eta(f_n)\}$ is a Cauchy sequence in the Hilbert space $L_2(\Omega, \mathbf{P})$. Thus, in this Hilbert space, there exists a limit random variable $\eta(f)$ for the sequence $\{\eta(f_n)\}$. $\square$

Remark 1. Since, in general, the space S is incomplete then, in this case, one cannot construct an *isometry* (one-to-one mapping preserving distances) between the L_2-closed linear span of all integral sums and S ([22]). Existence of such isometry is a key argument of the classical construction of stochastic integrals with orthogonal noises.

Remark 2. The generality in Theorem 1 allows us to define stochastic integrals both for the univariate and multivariate cases studied by the predecessors mentioned above. In the case of Gaussian noises generated by arbitrary centered Gaussian processes on the real line, our construction differs from that in [9], where multiple stochastic integrals are defined by the corresponding tensor power of the reproducing Hilbert space corresponding to the above-mentioned initial Gaussian process. However, one can prove that, to define these multiple integrals in the Gaussian case, the descriptions of the corresponding kernel spaces in these two constructions coincide.

Remark 3. If we consider the introduced-above multiple stochastic integral with the product–noise generated by a White noise with a structure function m_0, and, moreover, the kernel vanishes on all diagonal subspaces then our construction coincides with the classical Wiener – Itô multiple construction. Notice that, in this construction, for the kernels with zero values on all diagonal subspaces, there exists the isometry mentioned in Remark 1. In this case, the space S coincides with the Hilbert space $L_2(\mathfrak{X}^k, m_0^k)$, where k is the dimension of the multiple integral.

1.2. Infinitesimal analysis of covariance measures.

We now describe some function kernel spaces to define the stochastic integrals in Theorem 1. We start with the univariate construction.

1.2.1. Univariate stochastic integral.

Consider a centered random process $\xi(t)$ with a covariance function $\Phi(t,s)$. In all the examples of Section I we put $\mathfrak{X} = [0,T]$. This process generates the elementary stochastic measure $\mu(dt) := d\xi(t)$ introduced above.

Regular covariance functions. In the above-mentioned definition of the double difference of the covariance function Φ we set $t_{j+1} := t_j + \delta$ and $s_{j+1} := s_j + \delta$. Assume that, for all $\delta > 0$ and $t_j, s_j, t_j + \delta, s_j + \delta \in [0,T]$,

$$\Delta\Phi(t_i, s_i) = \int_{t_i}^{t_i+\delta} \int_{s_i}^{s_i+\delta} q(t,s)\lambda(dt)\lambda(ds),$$

where λ is an arbitrary σ-finite measure. If

$$\int_{\mathfrak{X}^2} |f(t)f(s)q(t,s)|\lambda(dt)\lambda(ds) < \infty$$

then $f \in S$ (i.e., $\int\limits_{\mathfrak{X}} f(t)d\xi(t)$ is well defined.)

For example, the regular FBM has the covariance function

$$\Phi(t,s) = \frac{1}{2}(t^{2h} + s^{2h} - |t-s|^{2h}),$$

where $h \in (1/2, 1]$. In this case, $q(t,s) := h(2h-1)|t-s|^{2h-2}$, $t \neq s$, and $\lambda(dt) := dt$ is the Lebesgue measure. Moreover, in this case one can prove the embedding $L_{1/h}(\mathfrak{X}, dt) \subseteq S$. ([22].)

Irregular covariance functions. Consider the class of *factorizing* covariance functions

$$\Phi(t,s) = G(\min(t,s))H(\max(t,s)).$$

It is known [3] that $\Phi(t,s)$ of such a kind is the covariance function of a nondegenerate on $(0,T)$ random process iff the fraction $\frac{G(t)}{H(t)}$ is a nondecreasing positive function. In particular, any Gaussian Markov process with non-zero covariance function admits such factorization: For example, a standard Wiener process has the components $G(t) = t$ and $H(t) \equiv 1$; a Brownian bridge on $[0,1]$ has the components $G(t) = t$ and $H(t) = 1-t$; and, finally, an arbitrary stationary Gaussian process on the positive half-line has the components $G(t) = \exp(\alpha t)$ and $H(t) = \exp(-\alpha t)$, where $\alpha > 0$.

Notice that, in this case, the function $\Phi(t,s)$ is nondifferentiable on the diagonal if the components are nondegenerate functions.

Let, in addition, $G(t) \uparrow$, $H(t) \downarrow$ be monotone, positive on $(0,T)$, and absolutely continuous w.r.t. the Lebesgue measure on $[0,T]$. We prove that *supp* $m^- = [0,T]^2 \setminus D$, where $D := \{(t,s) : \ t = s\}$, is the main diagonal of the square $[0,T]^2$. Let $s < s+\delta \leq t < t+\Delta$. Then

$$\begin{aligned} & m(\ (s, s+\delta] \times (t, t+\Delta]\ ) \\ &\quad = G(s)H(t) + G(s+\delta)H(t+\Delta) - G(s)H(t+\Delta) - G(s+\delta)H(t) \\ &\quad = (G(s+\delta) - G(s))(H(t+\Delta) - H(t)) < 0. \end{aligned}$$

In other words, m^- is absolutely continuous w.r.t. the bivariate Lebesgue measure λ_2 and the corresponding Radon – Nikodym derivative is defined by the formula

$$\frac{dm^-}{d\lambda_2}(t,s) = G'(t)|H'(s)|.$$

We now calculate m^+-measure of an infinitesimal diagonal square:

$$\begin{aligned} & m^+((s, s+h] \times (s, s+h]) \\ &\quad = H(s+h)\big(G(s+h) - G(s)\big) - G(s)\big(H(s+h) - H(s)\big) \\ &\quad = \int_s^{s+h} \big(H(z)G'(z) - G(z)H'(z)\big)dz + o(h). \end{aligned}$$

Hence m^+ is absolutely continuous w.r.t. the induced Lebesgue measure on the diagonal.

Therefore, in the case under consideration, the measures m^- and m^+ are *singular*. Finally, to verify the condition $f \in S$ we need to verify existence of the following two integrals:

$$\int_{\mathfrak{X}} f(t)^2(H(t)+1)dG(t), \ \int_{\mathfrak{X}} f(t)^2(G(t)+1)dH(t).$$

Covariance functions of mixed type. Let $\{X_n;\ n \geq 1\}$ be a stationary sequence of r.v.'s satisfying φ-mixing condition. Consider a centered Hilbert process (not necessarily Gaussian!) $Y(t)$ with the covariance function which is well defined under some restrictions on the φ-mixing coefficient (see the Gaussian case in [1]):

$$\mathbf{E}Y(s)Y(t) = F(\min(s,t)) - F(t)F(s) + \sum_{j\geq 1} \left(F_j(s,t) + F_j(t,s) - 2F(s)F(t)\right),$$

where $F(t)$ is the distribution function of X_1 which is assumed absolutely continuous with a density $p(t)$, and $F_j(t,s)$ is the joint distribution functions of the pairs (X_1, X_{j+1}). Let all the functions $F_j(t,s)$, $j = 1, 2, \ldots$, have bounded densities $p_j(t,s)$. For all $t, s \in R$, we assume that the series

$$b(t,s) := \sum_{j\geq 1} [p_j(t,s) + p_j(s,t) - 2p(t)p(s)]$$

absolutely converges and the corresponding series $\sum |\cdot|$ is integrable on R^2.

We note that, under the above-mentioned restrictions, we deal with a covariance function represented as a sum of covariance functions from items A.1 and A.2. Hence we may use the infinitesimal analysis of the corresponding covariance measures from these items. Indeed,

$$\begin{aligned} &m((t,t+\Delta] \times (s,s+\Delta]) \\ &\quad = \mathbf{P}(X_1 \in (t,t+\Delta] \cap (s,s+\Delta]) + \int_t^{t+\Delta} \int_s^{s+\Delta} (b(u,v) - p(u)p(v))\, dudv. \end{aligned}$$

So, under the conditions $\int f^2(t)p(t)dt < \infty$ and $\int\int |f(t)f(s)b(t,s)|dtds < \infty$, we can correctly define the stochastic integral $\int f(t)dY(t)$.

1.2.2. Multiple stochastic integral.

We study multiple stochastic integrals (MSI) based on a product-noise defined in Example (**ii**) by increments of a *Gaussian* process. In this case, to calculate the covariance measure we use the following well-known convenient representation:

$$m((t_1, t_1 + \delta] \times \cdots \times (t_{2k}, t_{2k} + \delta]) = \sum \prod \Delta\Phi(t_i, t_j),$$

where the sum is taken over all partitions on pairs of the set $\{1, 2, \ldots, 2k\}$, and the product is taken over all pairs in a fixed such partition. Notice that, in the sequel, to define multiple stochastic integrals we use only this property of Gaussian processes. However, we may study the multiple integrals for the non-Gaussian case: For example, if the integrating process $\xi(t)$ can be represented as a polynomial transform of a Gaussian process. In this case, to define the multiple integral, we can obtain some restrictions on the kernel f close to the conditions below.

Remark 4. The Main Assumption for the covariance measure $m(A_1 \times \cdots \times A_{2k})$ introduced above follows from the Proposition 1 if only $\Phi(t,s)$ has a bounded variation. This property of the covariance function is fulfilled in items B.1 – B.3 below.

Regular covariance functions. *Conditions to define MSI*:

$$\int_{[0,T]^{2k}} |f(t_1,\ldots,t_k)f(t_{k+1},\ldots,t_{2k})| \sum\prod |q(t_i,t_j)|dt_1\cdots dt_{2k} < \infty,$$

where the sum and the product are introduced above, $q(t,s)$ is the density (the Radon–Nikodym derivative) of $\Delta\Phi$. As a consequence, we obtain the main result in [7], for the regular FBM from item A.1. In this case we should set in this condition $q(t,s) := h(2h-1)|t-s|^{2h-2}$.

Factorizing covariance functions. Let the factorizing components H and G be smooth functions.
1) If $t_i = t_j$ then, as $\delta \to 0$, we have the following asymptotic representation of the double difference of $\Phi(\cdot)$ on the infinitesimal cube $(t_i, t_i+\delta] \times (t_j, t_j+\delta]$:

$$\Delta\Phi(t_i,t_j) = \delta(H(t_i)G'(t_i) - G(t_i)H'(t_i)) + O(\delta^2),$$

2) If $t_i \neq t_j$ then, as $\delta \to 0$,

$$\Delta\Phi(t_i,t_j) = \delta^2 G'(\min(t_i,t_j)H'(\max(t_i,t_j)) + o(\delta^2).$$

Denote

$$g_1(t) := H(t)G'(t) - G(t)H'(t),$$

$$g_2(t,s) := G'(\min(t,s))H'(\max(t,s)).$$

A set $D_{(r_1,\ldots,r_l)} \subset [0,T]^{2k}$ is called a *diagonal subspace determined by variables of multiplicity* $r_1,\ldots,r_l$ ($r_i \geq 2$, $\sum r_i < 2k$) if it defines by the following l chains of equalities:

$$x_{i_{j,1}} = \cdots = x_{i_{j,r_j}} \quad j = 1,\ldots,l,$$

where $i_{j,m} \neq i_{n,d}$ for any $(j,m) \neq (n,d)$.

Proposition 2 (see Borisov and Bystrov, 2006a). *In the case under consideration any covariance measure m has zero mass on any diagonal subspace $D_{(r_1,\ldots,r_l)}$ having at least one multiplicity $r_i > 2$.*

Given a kernel $f(t_1,\ldots,t_k)$ we set $\varphi_f(t_1\cdots t_{2k}) := f(t_1,\ldots,t_k)f(t_{k+1},\ldots,t_{2k})$.
Conditions to define MSI: First, we need to verify the condition

$$\begin{aligned}\int & |\varphi_f(s_1,s_1,\ldots s_n,s_n,t_1,\ldots t_{2(k-n)})| \\ & \times \prod_{i=1}^{n} g_1(s_i) \sum\prod |g_2(t_i,t_j)| ds_1\cdots ds_n dt_1\cdots dt_{2(k-n)} < \infty \end{aligned} \tag{3}$$

for all $n = 0,1,\ldots,k$, and, second, to verify finiteness of all the analogous integrals for all permutations of $2k$ arguments of the kernel φ_f. Here, by definition, $\prod_{i=1}^{0} = 1$.

If the kernel $f(\cdot)$ is symmetric and vanishes on all diagonal subspaces, and the functions g_1 and g_2 are bounded then condition (3) is reduced to the restriction $f \in L_2(\mathfrak{X}^k, dt_1\cdots dt_k)$. In particular, if the multiple noise is defined by increments of a standard Wiener process then $g_1 \equiv 1$ and $g_2 \equiv 0$ (cf. [17, 19]).

Covariance functions of mixed type. We now define the multiple stochastic integral for a Gaussian process $Y(t)$ with the covariance introduced in A.3. Let $p(t)$ and $b(t,s)$ defined in A.3 be continuous functions. Then, as $\delta \to 0$, we have for $t_i = t_j$ (see A.3):

$$\Delta\Phi(t_i, t_j) = \delta p(t_i) + o(\delta).$$

If $t_i \neq t_j$ then

$$\Delta\Phi(t_i, t_j) = \delta^2 \left(b(t_i, t_j) - p(t_i)p(t_j)\right) + o(\delta^2).$$

So, we actually repeat the arguments from item B.2 and to define the multiple stochastic integral for $Y(t)$ we need to verify condition (3) for $g_1(t) := p(t)$ and $g_2(t,s) := b(t,s)$.

2. Asymptotics of canonical von mises statistics.

In this Section we consider some applications of the MSI construction from Section I. We study limit behavior of multivariate Von Mises functionals of empirical processes based on samples from a stationary sequence of observations.

Let $\{X_n;\ n \geq 1\}$ be a stationary sequence of $[0,1]$-uniformly distributed r.v.'s satisfying the ψ-mixing condition: $\psi(m) \to 0$ if $m \to \infty$, where

$$\psi(m) := \sup \left| \frac{\mathbf{P}(AB)}{\mathbf{P}(A)\mathbf{P}(B)} - 1 \right|, \quad m = 1, 2, \dots, \tag{4}$$

and the supremum is taken over all events A and B (having non-zero probabilities) from the respective σ-fields $\mathcal{F}_1^k$ and $\mathcal{F}_{k+m}^\infty$, where $\mathcal{F}_l^k$, $l \leq k$, is the σ-field generated by the random variables $X_l, \dots, X_k$, as well as over all natural k. This mixing condition was introduced in Blum, Hanson and Koopmans, 1963.

Introduce the normalized d-variate Von Mises statistics (or V*-statistics*)

$$V_n := n^{-d/2} \sum_{1 \leq i_1, \dots, i_d \leq n} f(X_{i_1}, \dots, X_{i_d}), \quad n = 1, 2, \dots, \tag{5}$$

where the kernel $f(\cdot)$ satisfies the degeneracy condition

$$\mathbf{E} f(t_1, \dots, t_{k-1}, X_k, t_{k+1}, \dots, t_d) = 0$$

for all $t_1, \dots, t_d \in [0,1]$ and $k = 1, \dots, d$. Such *canonical* statistics were introduced in [25], and [15], where, moreover, the so-called U*-statistics* were studied:

$$U_n := (C_n^d)^{-1/2} \sum_{1 \leq i_1 < \cdots < i_d \leq n} f_0(X_{i_1}, \dots, X_{i_d}),$$

where, as a rule, the kernel f_0 is symmetric w.r.t. all permutations of the arguments.

Notice that in the definitions above we may consider the observations $\{X_i\}$ taking values in an arbitrary measurable space and having an arbitrary distribution. If the sample distribution has no atoms then an U-statistics with a symmetric kernel can be easily reduced to a V-statistics with a new kernel having zero values on all diagonal subspaces. For the IID samples limit behavior of these statistics is well known (for reference see Korolyuk and Borovskikh, 1994). In the case of one-dimensional observations, using the corresponding quantile transforms, we may reduce these statistics to those based on samples from the $[0,1]$-uniform distribution.

Introduce the normalized empirical process $S_n(t) := \sqrt{n}(F_n^*(t) - t)$, where $F_n^*(t)$ is the standard empirical distribution function based on the above-introduced sample $X_1, \dots, X_n$ with the $[0,1]$-uniform marginal distribution. We now recall the well-known key representation of the canonical V-statistics (5) by the Lebesgue integral:

$$V_n = \int\limits_{[0,1]^d} f(x_1, \dots, x_d)\; dS_n(x_1) \cdots dS_n(x_d). \tag{6}$$

We recall that a diagonal subspace in $[0,1]^d$ is determined as

$$\begin{aligned} D_{i^*,q^*} := \{(t_1, \dots, t_d) \in [0,1]^d :\\ t_{i_1} = t_{i_2} \dots = t_{i_{q_1}}, \dots, t_{i_{q_{r-1}+1}} = t_{i_{q_{r-1}+2}} \dots = t_{i_{q_r}}\}, \end{aligned}$$

where $i^* := (i_1, i_2, \dots, i_{q_1}, \dots, i_{q_{r-1}+1}, i_{q_{r-1}+2}, \dots, i_{q_r})$ is a vector with pairwise different integer coordinates from the set $\{1, \dots, d\}$ and $q^* := (q_1, \dots, q_r)$, where $q_1 \geq 2$, $q_r \leq d$ and $q_{i+1} - q_i \geq 2$ if $1 \leq i < r$. In the sequel, to indicate diagonal subspaces, we will use natural parameter i instead of vector-valued (i^*, q^*) renumbering all the diagonal subspaces.

The subspace of $[0,1]^d$ defined by pairwise different coordinates t_i is called the *main subspace* and denoted by D_0. Obviously,

$$D_0 = [0,1]^d \setminus \bigcup_{i \geq 1} D_i.$$

Introduce the following function space

$$S_0 := \{f : \sum_{i \geq 0} \int_{D_i} f^2(t_1, \dots, t_d) dt_{i_{q_1}} \cdots dt_{i_{q_r}} < \infty\},$$

where the subscripts $i_{q_1}, \dots, i_{q_r}$ determine the corresponding diagonal (or the main) subspace D_i of dimension r and the kernel $f^2(\cdot)$ in the corresponding multiple integral $\int_{D_i}$ of this notation has only r independent arguments. For example, if $D_i = \{(t_1, t_2, t_3, t_4) \in [0,1]^4 : \; t_1 = t_3, \; t_2 = t_4\}$ then

$$\int_{D_i} f^2(t_1, t_2, t_3, t_4)\; dt_3 dt_4 = \int_{[0,1]^2} f^2(t_3, t_4, t_3, t_4)\; dt_3 dt_4.$$

Notice that, under the conditions of Theorem 2 below, the function space S_0 is a subspace of S introduced in Theorem 1 for d-fold product-noise generated by increments of the stochastic processes $Y(t)$ from items A.3 and B.3 since the series

$$b(t,s) = \sum_{k \geq 1} (p_k(t,s) + p_k(s,t) - 2)$$

from item A.3 satisfies the restrictions we need and, moreover, it is uniformly bounded on $[0,1]^2$ due to ψ-mixing condition (4) and the restrictions on the mixing coefficient in Theorem 2 below. First of all, we observe that (4) implies existence of joint densities $p_k(t,s)$ since we may consider in (4) ‘infinitesimal events” $A := \{X_1 \in (t, t + dt)\}$ and $B := \{X_{k+1} \in (s, s + ds)\}$ and take the relations $\mathbf{P}(A) = dt$ and $\mathbf{P}(B) = ds$ into account. Moreover, we obtain the upper bound

$|b(t,s)| \le \sum_{k\ge 1}\psi(k)$. This fact allow us to estimate the covariance measure of the corresponding product-noise generated by increments of $Y(t)$.

Define on the space S_0 the combined L_2-norm

$$||f||_0^2 := \sum_{i\ge 0}\int_{D_i} f^2(t_1,\dots,t_d)dt_{i_{q_1}}\cdots dt_{i_{q_r}}, \tag{7}$$

where in the case $i=0$ (i. e., for the main subspace D_0,) we put $i_{q_k}=k$ and $r=d$ in the corresponding summand on the right-hand side of (7). Notice that this norm is stronger than $||\cdot||$ since in the case under consideration we have $g_1(t) := p(t) \equiv 1$ on $[0,1]$ and $g_2(t,s) := b(t,s)$ is a bounded function as well (see B.3 and the upper bounds in (20) and (21) below). In other words, in the case under considerations, the linear normed space $(S_0, ||\cdot||_0)$ is embedded to the corresponding space $(S, ||\cdot||)$ in Theorem 1.

Theorem 2. *Under the above-mentioned restrictions let $f \in S_0$ and*

$$\Psi(d) := \sum_{k\ge 1}\psi(k)k^{2d-2} < \infty. \tag{8}$$

Then, as $n\to\infty$,

$$V_n \xrightarrow{d} \int_{[0,1]^d} f(t_1,\dots,t_d)dY(t_1)\cdots dY(t_d),$$

where $Y(t)$ is a centered Gaussian process with the covariance

$$\boldsymbol{E}Y(t)Y(s) = \min(t,s) - ts + \sum_{k\ge 1}\left(F_k(t,s) + F_k(s,t) - 2ts\right).$$

Remark 5. In the IID case ($\psi(\cdot)\equiv 0$, $b(\cdot,\cdot)\equiv 0$, and $F_k(t,s) = ts$) the conditions of Theorem 2 coincide with the traditional well-known restrictions in Mises, 1947, Filippova, 1959, 1962, Borisov and Sakhanenko, 2000, and others. In this case, the process $Y(t) = W^0(t)$ is a Brownian bridge. Notice also that, in the IID case, the multiple stochastic integral w.r.t the product-noise generated by $W^0(t)$ in Theorem 2 coincides in distribution with the analogous integral w.r.t. a White noise (generated by a standard Wiener process $W(t)$) due to degeneracy of the kernel and the well-known representation $W^0(t) \stackrel{d}{=} W(t) - tW(1)$.

If a kernel $f(t_1,\dots,t_d)$ vanishes on all diagonal subspaces then sufficient conditions for f to be an element of S_0 is as follows: $f\in L_2([0,1]^d, dt_1\cdots dt_d)$. The kernels of such a kind are used to describe limit behavior of canonical U-statistics (for example, see Major, 1994, and the corresponding comment above). Moreover, to describe the Itô – Wiener construction of multiple integrals the kernels vanishing on all diagonal subspaces provide the above-mentioned isometry between closed linear span of multiple integral sums and the corresponding function space.

Remark 6. From the beginning, in the IID case, to describe limit behavior of canonical U and V-statistics another representation of the limit distribution was used. For instance, in the case $d=2$, for a canonical U-statistics with a symmetric kernel the limit random variable can be represented as the random series

$$\sum_{k=1}^{\infty}\lambda_k(\tau_k^2-1), \tag{9}$$

where $\{\tau_k\}$ are i.i.d. standard Gaussian random variables, $\{\lambda_k\}$ are eigenvalues of the integral operator in $L_2[0,1]$ with the kernel $f(s,t)$ (see [25]). Later this result was extended to the general case $d \geq 2$ (see [23]). In the general case the limit random variable can be represented as a polynomial transform of $\{\tau_k\}$ in terms of the Hermite–Chebyshev polynomials. Finally, using analogous representations of multiple Wiener integrals noted in [27], and in [17], one can rewrite the limit random variables as the corresponding multiple stochastic integral w.r.t. to a multiple White noise (see, for example, Dynkin and Mandelbaum, 1983).

In the non-IID case an analog of (9) was obtained in [11], where $\{\lambda_k\}$ are the above-mentioned eigenvalues which are additionally assumed to be summable, and $\{\tau_k\}$ is a Gaussian sequence with the covariance function depending of these eigenvalues as well as of the covariance function of the initial stationary sequence $\{X_i\}$ satisfying φ-mixing condition.

In the special case when the observations are defined by a nonrandom transform of a Gaussian stationary sequence under another dependency restriction, limit behavior of canonical U-statistics were investigated in [8].

Proof of Theorem 2. We divide the proof into three steps (for detail see [5]). Let $\{f_k(x_1,\ldots,x_d)\}$ be a sequence of step functions of the form

$$f_k(x_1,\ldots,x_d) := \sum_{j_1,\ldots,j_d \leq k} f_{j_1,\ldots,j_d} \prod_{i=1}^{d} I(x_i \in A_{j_i}) \tag{10}$$

converging in $(S_0, ||\cdot||_0)$ to a function $f(x_1,\ldots,x_d)$ as $k \to \infty$. Hence the functions $\{f_k\}$ converge to f in the corresponding linear normed space $(S, ||\cdot||)$ in Theorem 1. Then

S t e p I. For every fixed k, due to the multivariate central limit theorem for finite-dimensional distributions of the empirical measure based on slowly dependent observations (see, for example, [1]), as $n \to \infty$,

$$\int_{[0,1]^d} f_k(x_1,\ldots,x_d)dS_n(x_1)\cdots dS_n(x_d) \xrightarrow{d} \int_{[0,1]^d} f_k(x_1,\ldots,x_d)dY(x_1)\cdots dY(x_d).$$

Actually, this weak convergence is valid under weaker dependency conditions than those in Theorem 2.

S t e p II. As $k \to \infty$, by the above-mentioned construction of multiple stochastic integrals (see Theorem 1 and the comment above) we have the following mean square convergence:

$$\int_{[0,1]^d} f_k(x_1,\ldots,x_d)dY(x_1)\cdots dY(x_d) \xrightarrow{L_2} \int_{[0,1]^d} f(x_1,\ldots,x_d)dY(x_1)\cdots dY(x_d).$$

S t e p III. As $k \to \infty$, we should prove the following mean square convergence *uniformly* on n:

$$\int_{[0,1]^d} f_k(x_1,\ldots,x_d)dS_n(x_1)\cdots dS_n(x_d) \xrightarrow{L_2} \int_{[0,1]^d} f(x_1,\ldots,x_d)dS_n(x_1)\cdots dS_n(x_d).$$

It is easy to see that the assertion of Theorem 2 follows from items I – III.

So, we should prove only item III. To establish the last convergence, for *every* step function g_k of the form (10), it suffices to prove the estimate

$$\sup_n \mathbf{E}\left(\int_{[0,1]^d} g_k(x_1,\ldots,x_d)dS_n(x_1)\cdots dS_n(x_d)\right)^2 \leq C||g_k||_0^2, \tag{11}$$

where the constant C depends on d and $\Psi(d)$ only. Indeed, this estimate implies the analogous upper bound for the step function $f_k - f_l$ (see the corresponding comment in the proof of Theorem 1), where f_k and f_l are arbitrary elements of the above-mentioned sequence $\{f_m\}$ converging to the kernel f in the norm $||\cdot||_0$. Therefore, as $l \to \infty$, the corresponding L_2-limit of the integral sums $\int f_l(x_1,\ldots,x_d)dS_n(x_1)\cdots dS_n(x_d)$ exists and coincides with the Lebesgue integral in (6) since this limit does not depend on a sequence of step functions approximating to f. Thus, as $k \to \infty$,

$$\sup_n \mathbf{E}\left(\int f_k(x_1,\ldots,x_d)dS_n(x_1)\cdots dS_n(x_d)\right.$$
$$\left. - \int f(x_1,\ldots,x_d)dS_n(x_1)\cdots dS_n(x_d)\right)^2 \to 0$$

for every sequence of step functions $\{f_k\}$ converging to the kernel f in $(S_0, ||\cdot||_0)$ if only relation (11) is valid.

To prove (11) we need the following auxiliary statements.

Lemma 1. *(see [21]) Let ξ and η be random variables measurable w.r.t. the $\mathcal{F}_1^k$ and $\mathcal{F}_{k+m}^\infty$ $(m \ge 1)$ respectively. If $\boldsymbol{E}|\xi| < \infty$ and $\boldsymbol{E}|\eta| < \infty$ then*

$$|\boldsymbol{E}\xi\eta - \boldsymbol{E}\xi\boldsymbol{E}\eta| \le \psi(m)\boldsymbol{E}|\xi|\boldsymbol{E}|\eta|. \tag{12}$$

Introduce the following notation: $\widetilde{I}_k(A) := I(X_k \in A) - P(A), \quad A \in \mathcal{A}$, where the marginal P is the Lebesgue measure on $[0,1]$. The following assertion is easy proved by induction (for detail see [5]).

Lemma 2. *For any natural numbers q and $l_1,\ldots,l_q$, and for any pairwise disjoint measurable subsets $A_1,\ldots,A_q$ the following moment inequality is valid:*

$$\boldsymbol{E}|\widetilde{I}_k^{l_1}(A_1)\cdots\widetilde{I}_k^{l_q}(A_q)| \le (q+1)P(A_1)\cdots P(A_q). \tag{13}$$

From Lemmas 1 and 2 we deduce the following upper bound.

Lemma 3. *Let $q_1 < \cdots < q_s$ be arbitrary natural numbers and $q_0 = 0$. Consider s collections of sets $\{A_1,\ldots A_{q_1}\},\ldots,\{A_{q_{s-1}+1},\ldots A_{q_s}\}$, where the measurable subsets A_i inside of every collection are pairwise disjoint. Put*

$$\nu_{k_i} := \widetilde{I}_{k_i}^{l_{q_{i-1}+1}}(A_{q_{i-1}+1})\cdots\widetilde{I}_{k_i}^{l_{q_i}}(A_{q_i}).$$

Then, for any natural numbers $k_1 < \cdots < k_s$ and $l_1,\ldots,l_{q_1},\ldots,l_{q_s}$, the following estimate is valid:

$$\boldsymbol{E}|\nu_{k_1}\cdots\nu_{k_s}| \le C(\psi(1),s,q_s)P(A_1)\cdots P(A_{q_s}), \tag{14}$$

where the constant $C(\cdot)$ depends only on the arguments indicated.

Proof. By (13) we have $\mathbf{E}|\nu_{k_i}| \le (q_i - q_{i-1} + 1)P(A_{q_{i-1}+1})\cdots\mathbf{P}(A_{q_i})$. It is clear that the random variables $\{\nu_{k_i}\}$ satisfy ψ-mixing condition. Hence, by (12) and (13) we obtain

$$\begin{aligned}\mathbf{E}|\nu_{k_1}\cdots\nu_{k_s}| &\le \prod_{j=1}^{s-1}(1+\psi(k_{j+1}-k_j))\mathbf{E}|\nu_{k_1}|\cdots\mathbf{E}|\nu_{k_s}| \\ &\le \prod_{j=1}^{s-1}(1+\psi(k_{j+1}-k_j))\prod_{j=1}^{s-1}(q_j - q_{j-1}+1)P(A_1)\cdots P(A_{q_s}) \\ &\le C(\psi(1),s,q_s)P(A_1)\cdots P(A_{q_s}).\end{aligned}$$

The Lemma is proved. □

The main assertion to prove (11) is as follows:

Lemma 4. *Let $l_1, \dots, l_q$ be natural numbers such that $l_1 + \dots + l_q = 2d$, and let $A_1, \dots, A_q$ be a collection of pairwise disjoint measurable sets. Under condition* (8) *the following upper bound is valid:*

$$|\mathbf{E}S_n^{l_1}(A_1)\cdots S_n^{l_q}(A_q)| \le CP(A_1)\cdots P(A_q). \tag{15}$$

where the constant C depends only on d and $\Psi(d)$.

Remark 7. The statement of Lemma 4 is also contained in Sen, 1972. However, the corresponding constant in this paper contains the factor $(1+\psi(0))^2$. It is easy to see that if the marginal distribution has a continuous component (for example, if it is the Lebesgue measure on $[0,1]$) or infinitely many atoms then $\psi(0) = \infty$. To verify this property we may put in (4) $A = B$, where A is an event from the σ-field $\mathcal{F}_1^1$. So, under the above-mentioned restrictions on the marginal P

$$\psi(0) \ge \sup_{A\in\mathcal{F}_1^1,\ P(A)>0} (1/P(A) - 1) = \infty.$$

In the sequel we will denote by the symbols C or C_i various positive constants depending only on d and $\Psi(d)$.

Proof of Lemma 4. We first write the simple estimate

$$\begin{aligned}&|\mathbf{E}S_n^{l_1}(A_1)\cdots S_n^{l_q}(A_q)|\\ &\quad\le n^{-d}\sum_{k_1,\dots,k_{2d}\le n}|\mathbf{E}\widetilde{I}_{k_1}(A_1)\cdots\widetilde{I}_{k_{l_1}}(A_1)\cdots\widetilde{I}_{k_{2d-l_q+1}}(A_q)\cdots\widetilde{I}_{k_{2d}}(A_q)|.\end{aligned}$$

The initial sum on the right-hand side of this inequality can be estimated by a finite sum of the following diagonal subsums

$$\sum_{k_1<\dots<k_r\le n}|\mathbf{E}\nu_{k_1}\cdots\nu_{k_r}|, \tag{16}$$

where $\nu_{k_i} := \widetilde{I}_{k_i}^{s_1(i)}(A_1)\cdots\widetilde{I}_{k_i}^{s_q(i)}(A_q)$, and the integers $s_j(i)$ are defined by the corresponding diagonal subspace of subscripts in the initial multiple sum and satisfy the conditions $0 \le s_j(i) \le l_j$ for all $i \le r$ and $j \le q$, and $\sum_{i\le r}\sum_{j\le q} s_j(i) = 2d$.

Let $r \le d$. Estimating by (14) each summand in (16) and taking the normalized factor n^{-d} and the number of summands in (16) into account we obtain the upper bound we need.

Let now $r > d$. We call the random variable ν_{k_i} *short product* if $\sum_{j\le q} s_j(i) = 1$, i. e., $\nu_{k_i} = \widetilde{I}_{k_i}(A_{q_i})$ for some $q_i \le q$. Notice that if ν_{k_i} is a short product then $\mathbf{E}\nu_{k_i} = 0$.

We now consider the auxiliary multiple sum consisting of the random variables ν_{k_i} defined in (16) for the fixed diagonal subspace of subscripts:

$$\sum_{k_{v_1}<\dots<k_{v_2}\le n}|\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_{v_2}}|, \tag{17}$$

where $1 \le v_1 < v_2 \le r$ and the value $v := v_2 - v_1 + 1$ is the dimension of the corresponding multiple sum. Introduce the following notation: $e_j(i) := \min\{1, s_j(i)\}$.

We first prove the following assertion: If, in the summands in (17), there are at least m shorts products, where $0 \le m \le v$, then the following upper bound is valid:

$$\sum_{k_{v_1}<\cdots<k_{v_2}\le n} |\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_{v_2}}| \le Cn^{v-m/2}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)}, \tag{18}$$

where $\alpha_j(v_1,v_2) := \sum\limits_{i=v_1}^{v_2} e_j(i)$. Notice that the set function $\alpha_j(a,b)$ is additive on intervals $[a,b]$. We prove this assertion by induction on m for all v_1 and v_2 such that $v \ge m$ and $v \le r$. Let $m=1$, i. e., the expectations in (17) contain at least one short product. Denote it by ν_{k_l}, where $k_{v_1} \le k_l \le k_{v_2}$. First we note that, in terms of the notation above, we can rewrite the statement of Lemma 3 for the absolute moment of each random product in (17) in such a way:

$$\mathbf{E}|\nu_{k_{v_1}}\cdots\nu_{k_{v_2}}| \le C\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)}.$$

Taking this estimate into account we evaluate by (12) every summand in (17) setting $\xi := \nu_{k_{v_1}}\cdots\nu_{k_l}$ and $\eta := \nu_{k_{l+1}}\cdots\nu_{k_{v_2}}$:

$$\begin{aligned}
&\sum_{k_{v_1}<\cdots<k_{v_2}\le n} |\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_{v_2}}| \\
&\le \sum_{k_{v_1}<\cdots<k_{v_2}\le n} \psi(k_{l+1}-k_l)\mathbf{E}|\nu_{k_{v_1}}\cdots\nu_{k_l}|\mathbf{E}|\nu_{k_{l+1}}\cdots\nu_{k_{v_2}}| \\
&\quad + \sum_{k_{v_1}<\cdots<k_l\le n} |\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_l}| \sum_{k_{l+1}<\cdots<k_{v_2}\le n} \mathbf{E}|\nu_{k_{l+1}}\cdots\nu_{k_{v_2}}| \\
&\le C_1 n^{v-1}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)}\sum_{i\ge 1}\psi(i) \\
&\quad + C_2 n^{v_2-l}\prod_{j\le q} P(A_j)^{\alpha_j(l+1,v_2)} \\
&\quad \times \sum_{k_{v_1}<\cdots<k_l\le n} \psi(k_l-k_{l-1})\mathbf{E}|\nu_{k_1}\cdots\nu_{k_{l-1}}|\mathbf{E}|\nu_{k_l}| \\
&\le (C_3 n^{v-1} + C_4 n^{v_2-l}n^{l-v_1})\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)} \\
&\le C_5 n^{v-1/2}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)}
\end{aligned}$$

which required. In this chain of relations the second inequality is valid due to (12) and the equality $\mathbf{E}\nu_{k_l}=0$ as well.

We now assume that the upper bounds

$$\sum_{k_{v_1}<\cdots<k_{v_2}\le n} |\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_{v_2}}| \le Cn^{v-z/2}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)}$$

are true for all integers $z < m$, where z is the minimal possible number of short products in the expectations under consideration, and for all possible dimensions $v:\ z \le v \le r$ of multiple sums of the form (17), and, moreover, the moments in (17) contain at least m shorts products. Denote these products by $\nu_{k_{l_1}},\ldots,\nu_{k_{l_m}}$.

Consider the following $m-1$ pairs of neighboring products: $\nu_{k_{l_s}}\nu_{k_{l_s+1}}$, $s = 1\dots m-1$. Denote by $t_1, \dots, t_{m-1}$ differences between the subscripts in these pairs. We have

$$\sum_{k_{v_1}<\dots<k_{v_2}\le n} |\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_{v_2}}| \le R_1 + \dots + R_{m-1},$$

where the subsum R_s is taken over the set of subscripts

$$I_s := \{(k_{v_1}, \dots, k_{v_2}):\ k_{v_1} < \dots < k_{v_2} \le n,\ t_s = \max t_i\}.$$

We now estimate by (12) each summand in R_s setting $\xi := \nu_{k_{v_1}}\cdots\nu_{k_{l_s}}$ and $\eta := \nu_{k_{l_s+1}}\cdots\nu_{k_{v_2}}$:

$$\begin{aligned} R_s \quad &\le \quad \sum_{I_s}\psi(k_{l_s+1} - k_{l_s})\mathbf{E}|\nu_{k_{v_1}}\cdots\nu_{k_{l_s}}|\mathbf{E}|\nu_{k_{l_s+1}}\cdots\nu_{k_{v_2}}| \\ &\quad + \sum_{k_{v_1}<\dots<k_{l_s}\le n}|\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_{l_s}}|\sum_{k_{l_s+1}<\dots<k_{v_2}\le n}\mathbf{E}|\nu_{k_{l_s+1}}\cdots\nu_{k_{v_2}}|. \end{aligned} \tag{19}$$

Consider the first sum on the right-hand side of (19):

$$\begin{aligned} &\sum_{I_s}\psi(k_{l_s+1} - k_{l_s})\mathbf{E}|\nu_{k_{v_l}}\cdots\nu_{k_{l_s}}|\mathbf{E}|\nu_{k_{l_s+1}}\cdots\nu_{k_{v_2}}| \\ &\qquad \le C\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)}\sum_{I_s}\psi(t_s) \\ &\qquad \le C\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)}n^{v-(m-1)}\sum_{t_i:\,t_i\le t_s}\psi(t_s) \\ &\qquad \le C\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)}n^{v-m/2}\Psi\left(\frac{m}{2}\right). \end{aligned}$$

Notice that the last inequality is valid for $m \ge 2$. Consider now the product of the sums on the right-hand side of (19). Let the summands of the first sum contain m_1 short products indicated above, and, in the summands of the second sum, there are $m - m_1$ short products indicated above. By the construction we have $1 \le m_1 \le m-1$. Hence, for these sums we can use the induction proposition. Finally, we have

$$\begin{aligned} &\sum_{k_{v_1}<\dots<k_{l_s}\le n}|\mathbf{E}\nu_{k_{v_1}}\cdots\nu_{k_{l_s}}|\sum_{k_{l_s+1}<\dots<k_{v_2}\le n}|\mathbf{E}\nu_{k_{l_s+1}}\cdots\nu_{k_{v_2}}| \\ &\qquad \le Cn^{l_s-m_1/2}n^{v-l_s-(m-m_1)/2}\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)} \\ &\qquad = Cn^{v-m/2}\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)} \end{aligned}$$

which required. Thus, for R_s we obtained the upper bound we need. It means that the analogous estimate is valid for whole sum (17). The induction is over.

To prove the assertion of Lemma 4 we should note that, first, by the definition, $\alpha_j(1, r) \ge 1$ for all $j \le q$, and, second, in the case $r > d$, the summands in (16) contain at least $2(r-d)$ short products. So, we should put in (18) $v_1 := 1$, $v_2 := r$, $m := 2(r-d)$ and $v := r$. It means that, for the sum in (16), the following upper bound is valid:

$$\sum_{k_1<\dots<k_r\le n}|\mathbf{E}\nu_{k_1}\cdots\nu_{k_r}| \le Cn^d\prod_{j\le q}P(A_j)^{\alpha_j(1,r)} \le Cn^dP(A_1)\cdots P(A_q).$$

The Lemma is proved. □

In conclusion we deduce from Lemma 4 the upper bound (11). For every step function g_k of the form (10) we have

$$
\begin{aligned}
&\sup_n \mathbf{E}\left(\int_{[0,1]^d} g_k(x_1,\dots,x_d)dS_n(x_1)\cdots dS_n(x_d)\right)^2 \\
&\quad = \sup_n \mathbf{E}\int_{[0,1]^{2d}} g_k(x_1,\dots x_d)g_k(x_{d+1},\dots x_{2d})dS_n(x_1)\cdots dS_n(x_{2d}) \\
&\quad \le \sup_n \sum_{i_1,\dots,i_{2d}} |g_{i_1,\dots,i_d}||g_{i_{d+1},\dots,i_{2d}}|\mathbf{E}|S_n(A_{i_1})\cdots S_n(A_{i_{2d}})| \\
&\quad \le C\sum_{i\ge 0}\int_{D_i^{(2)}} |g_k(x_1,\dots x_d)||g_k(x_{d+1},\dots x_{2d})|P(dx_{i_{q_1}})\cdots P(dx_{i_{q_r}}),
\end{aligned}
\tag{20}
$$

where $D_i^{(2)}$ is a diagonal (or the main) subspace in $[0,1]^{2d}$ of dimension r which is defined by the integers $q_1,\dots,q_r$ by analogy with D_i in $[0,1]^d$. Further, by the Cauchy – Bunyakovskii inequality we estimate every multiple integral on the right-hand side of (20):

$$
\begin{aligned}
&\int_{D_i^{(2)}} |g_k(x_1,\dots x_d)||g_k(x_{d+1},\dots x_{2d})|P(dx_{i_{q_1}})\cdots P(dx_{i_{q_r}}) \\
&\quad \le \Bigg(\int_{D_{j_1}} g_k^2(x_1,\dots x_d)P(dx_{i_{v_1}})\cdots P(dx_{i_{v_l}}) \\
&\quad \times \int_{D_{j_2}} g_k^2(x_1,\dots x_d)P(dx_{i_{s_1}})\cdots P(dx_{i_{s_m}})\Bigg)^{1/2} \\
&\quad \le ||g_k||_0^2,
\end{aligned}
\tag{21}
$$

where D_{j_1} and D_{j_2} are the corresponding diagonal (or the main) subspaces in $[0,1]^d$ of respective dimensions l and m, where $l+m\ge r$, defined by the integers $\{v_j\}$ and $\{s_j\}$. Thus, the upper bound in (11) is proved. □

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* Willey. *New York–London–Sydney–Toronto.* MR233396

[2] Blum, J. R., Hanson, D. L. and Koopmans, L. H. (1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrscheinlichkeitstheor. und verw. Geb.* **2**(1) 1–11. MR161369

[3] Borisov, I. S. (1982). On a criterion for Gaussian random process to be Markovian. *Theory Probab. Appl.*, **27** 863–865. MR681474

[4] Borisov, I. S. and Bystrov, A. A. (2006a). Constructing a stochastic integral of a nonrandom function without orthogonality of the noise. *Theory Probab. Appl.*, **50**(1) 53–74.

[5] Borisov, I. S. and Bystrov, A. A. (2006b). Limit theorems for canonical Von Mises statistics based on dependent observations. *Siberian Math. J.* **47**(6) (in Russian, to appear).

[6] Borisov, I. S. and Sakhanenko, L. A. (2000). The central limit theorem for generalized canonical Von Mises statistics. *Siberian Adv. Math.* **10**(4) 1–14. MR1817633

[7] Dasgupta, A. and Kallianpur, G. (1999). Multiple fractional integrals. *Probab. Theory Relat. Fields* **115**(4) 505–525. MR1728919

[8] Dehling, H., Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17**(4) 1767–1783. MR1026312

[9] Cambanis, S. and Huang, S. T. (1978). Stochastic and multiple Wiener integrals for Gaussian processes. *Ann. Probab.* **6**(4) 585–614. MR496408

[10] Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes and multiple Wiener integrals. *Ann. Statist.* **11**(3) 739–745. MR707925

[11] Eagleson, G.K. (1979). Orthogonal expansions and U-statistics. *Austral. J. Statist.* **21**(3) 221–237. MR561949

[12] Filippova, A. A. (1959). Von Mises' theorem on limit behavior of functionals of the empirical distribution functions. *Doklady Acad. Nauk USSR* **129**(1) 44–47 (in Russian). MR124946

[13] Filippova, A. A. (1962). Von Mises' theorem on limit behavior of functionals of the empirical distribution functions and its statistical applications. *Theory Probab. Appl.* **7**(1) 26–60. MR150870

[14] Gihman, I. I. and Skorohod, A. V. (1974). *The theory of stochastic processes.* I. Springer-Verlag. Berlin–Heidelberg–New York. MR346882

[15] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325. MR26294

[16] Hoeffding, W. (1961). The strong law of large numbers for U-statistics. Inst. Statist. Mimeo Ser., No. 302, pp. 1–10.

[17] Itô, K. (1951). Multiple Wiener Integral. *J. Math. Soc. Jap.* **3**(1) 157–169. MR44064

[18] Korolyuk, V. S. and Borovskikh, Yu. V. (1994). *Theory of U-statistics.* Kluwer Academic Publ. Dordrecht. MR1472486

[19] Major, P. (1981). Multiple Wiener–Itô integrals. *Lecture Notes in Math.* **849** Springer–Verlag, Berlin. MR611334

[20] Major, P. (1994). Asymptotic distribution for weighted U-statistics. *Ann. Probab.* **22**(3) 1514–1535. MR1303652

[21] Philipp, W. (1969). The central limit problem for mixing sequences of random variables. *Z. Wahrscheinlichkeitstheor. verw. Geb.* **12**(2) 155–171. MR246356

[22] Pipiras, V. and Taqqu, M. (2000). Integration questions related to fractional Brownian motion. *Probab. Theory Relat. Fields* **118**(2) 251–291. MR1790083

[23] Rubin, H. and Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8**(1) 165–170. MR557561

[24] Sen, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrscheinlichkeitstheor. verw. Geb.* **25**(1) 71–82. MR329003

[25] Von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348. MR22330

[26] Wiener, N. (1938). The homogeneous chaos. *Amer. J. Math.* **60** 897–936. MR1507356

[27] Wiener, N. (1958). *Nonlinear Problems in Random Theory.* New York. Wiley. MR100912

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 18–32

DOI: 10.1214/074921706000000734

Invariance principle for stochastic processes with short memory

Magda Peligrad[1,*] **and Sergey Utev**[2]

University of Cincinnati and University of Nottingham

Abstract: In this paper we give simple sufficient conditions for linear type processes with short memory that imply the invariance principle. Various examples including projective criterion are considered as applications. In particular, we treat the weak invariance principle for partial sums of linear processes with short memory. We prove that whenever the partial sums of innovations satisfy the L_p–invariance principle, then so does the partial sums of its corresponding linear process.

1. Motivation: linear processes

Encountered in various applications, the linear time series, moving averages, provide a reach class of examples that are widely studied. For a stationary sequence of innovations $(\xi_i)_{i\in Z}$ and a sequence of constants $(a_j)_{j\in\mathbb{Z}}$, the linear process is defined by

$$X_k = \sum_{j=-\infty}^{\infty} a_j \xi_{k-j}. \tag{1}$$

Of course, one needs to add some conditions in order for the process to be well defined. In the classical time series analysis $(\xi_i)_{i\in Z}$ is assumed i.i.d. with $\mathbb{E}(\xi_0) = 0$ and $\mathbb{E}(\xi_0^2) < \infty$. Then, X_k is well defined when $\sum_{j=-\infty}^{\infty} a_j^2 < \infty$ and the central limit theorem (CLT) holds for $S_n/stdev(S_n)$ where $S_n = \sum_{j=1}^{n} X_j$ [10]. This CLT is not restricted to i.i.d sequences and the theorem was extended in [19, 21] to martingales and other related structures. However, without additional assumptions on the sequence of constants, the CLT for the general linear process cannot be extended to an invariance principle, not even for independent innovations, as pointed out in [15, 23] and in [16]. A related example is given in Proposition 10 below.

To deal with the problem of the invariance principle, several authors imposed certain regularity conditions on the sequence of constants together with restrictions on the dependence structure of the innovations. Various invariance principles on this line are given for example in [23, 24] and also in [15] among others.

In this paper we shall discuss the case when the coefficients a_j are absolutely summable,

$$\sum_{i\in\mathbb{Z}} |a_i| < \infty. \tag{2}$$

*Supported in part by a Charles Phelps Taft Memorial Fund grant and NSA grant, H98230-05-1-0066.

[1]Department of Mathematical Sciences, University of Cincinnati, PO Box 210025, Cincinnati, OH 45221-0025, USA.

[2]School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK.

AMS 2000 subject classifications: primary 60F17, 60F17; secondary 60K99, 60G48, 60G10.

Keywords and phrases: stationary process, linear processes, Brownian motion, invariance principle, weakly dependent sequences, sample.

This case is referred to as short memory, or sometimes as short range dependence. For this situation [11] proved that the central limit theorem (CLT) is preserved under the linear transformation, i.e. the CLT for partial sums of innovations $S_n^{(\xi)} = \sum_{j=1}^n \xi_j$, properly normalized, implies the CLT for S_n under practically the same normalization. This raises the question whether the invariance principle is also preserved under the linear transformation with short memory, that is whenever it holds for the innovations then it also holds for the corresponding linear process.

There are numerous papers dealing with this problem for particular time series with short range dependence and dependent innovations. All these results are asserting that if the innovations have a certain dependence structure, (particular classes of mixing, associated sequences, negative associated sequences, martingales, martingale-like sequences and so on) and satisfy the invariance principle, so does the short range dependent linear process.

To give an example of a result of this type, we mention the well known case when the innovations form a stationary martingale difference sequence [4, 9]. Then, the invariance principle holds,

$$\frac{S_{[nt]}}{\sqrt{n}} \Longrightarrow \eta A W(t) \quad \text{where} \quad A = \sum_{i\in\mathbb{Z}} a_i \ , \ S_k = \sum_{j=1}^{k} X_j, \tag{3}$$

where η is a random variable measurable with respect to the invariant σ–field of the stationary sequence $(\xi_i)_{i\in Z}$, denoted by $\mathcal{I}$, $W = \{W(t)\ ;\ 0 \le t \le 1\}$ is a standard Brownian motion independent of the invariant σ–field $\mathcal{I}$, $[x]$ denotes the integer part of x and $\Longrightarrow$ denotes weak convergence in $D[0,1]$, the space of cadlag functions on $[0,1]$ endowed with the uniform topology.

In this paper we show that for the short range dependent case the dependence structure is not important and, whenever the innovations satisfy the invariance principle the corresponding short range dependent process is also convergent, provided a certain condition is imposed to $E(\max_{1\le j\le n} |\sum_{k=1}^{j} \xi_k|)$. We also address the question of the L_p- invariance principle ($p \ge 1$) that is

$$\mathbb{E}(\sup_{0\le t\le 1} |\frac{S_{[nt]}}{\sqrt{n}} - \eta A W(t)|^p) \to 0 \text{ as } n \to \infty \tag{4}$$

and we show that if the innovations satisfy an L_p- invariance principle so does the linear process.

For dealing with this problem and other related facts, we develop first a general device to compare the linear combination of processes to one of the initial processes. Another general result will allow to prove the weak convergence for a linear combination of processes by studying only a finite sum. The theory is not restricted to real valued stochastic processes and it can be used for random fields and Hilbert and Banach space valued processes. We then apply the general results to study the asymptotic behavior of linear processes with short memory and to extend the celebrated projective criterion of Hannan [9] to the nonstationary case. Various examples will comment on the optimality of our results. We shall also point out that, for a more general linear transformation with short memory, the invariance principle for innovations does not imply in general, the invariance principle for the linear process.

In this paper we shall use the following notations. The space $B(T)$ of bounded functions on $T \subset R^d$ is equipped with the supremum norm $\|x\|_T = \sup_{t\in T} |x(t)|$. The notation $\|X\| = \sqrt{E(X^2)}$ stays for the L_2-norm. As already specified in

(3), the notation $S_n = X_1 + \cdots + X_n$ will be reserved for the partial sums of the key sequence of interest $(X_i)_i$. Throughout the paper, without changing the distributions we shall redefine the innovations on a probability space rich enough to support a standard Brownian motion.

2. The general device

A well-known powerful tool in the proofs of weak convergence results consists of approximating the underlying sequence (W_n) by a double indexed sequence $(W_n^{(m)})$, such that for any m, the sequence $W_n^{(m)} \to W^{(m)}$ and then, we deduce the existence of the limit of the original sequence (W_n) by studying the limit of $(W^{(m)})$ as $m \to \infty$. (see for example, Theorem 3.2. in [1]).

In this section we shall establish two variants of this type of approximation, aimed to study the linear combination of stochastic processes.

First, we present a coupling type approximation lemma. We consider a process that can be expressed as a linear combination with short memory of stochastic processes, and approximate it by an individual summand. The context is general enough to be applied to processes indexed by countable sets allowing also the treatment of random fields and, just a modification of language will lead us to higher dimensional spaces and operators.

Lemma 1. *Let $\psi^{(n)} = \psi^{(n)}(t)$, $t \in T$, $n = 1, 2, \ldots$ be a sequence of stochastic processes that admits the following representation*

$$\psi^{(n)} = \sum_{j \in I} a_j U_j^{(n)} \quad \textit{with} \quad \sum_{j \in I} |a_j| < \infty, \tag{5}$$

where I is a countable set, $(a_j)_{j \in I}$ is a real valued sequence and $U_j^{(n)} = U_j^{(n)}(t)$, $t \in T$, $j \in I$, $n = 1, 2, \ldots$ is a double sequence of stochastic processes, satisfying the following conditions.

$$\sup_n \sup_j \mathbb{E}(\|U_j^{(n)}\|_T) \ < \ \infty \tag{6}$$

and, for each pair $i, j \in I$,

$$\|U_i^{(n)} - U_j^{(n)}\|_T \ \to 0 \qquad \textit{in probability as } n \to \infty. \tag{7}$$

Fix $e \in I$. Then, with the notation $A = \sum_{j \in I} a_j$, we have

$$\|\psi^{(n)} - AU_e^{(n)}\|_T \to 0 \quad \textit{in probability as } n \to \infty.$$

Moreover, if for a certain $p \geq 1$ we have $\sup_n \sup_j \mathbb{E}(\|U_j^{(n)}\|_T^p) \ < \ \infty$ and the convergence in (7) is in L_p then

$$\mathbb{E}(\|\psi^{(n)} - AU_e^{(n)}\|_T^p) \to 0 \ \textit{as } n \to \infty.$$

Proof of Lemma 1. First we find a sequence $(I_m)_{m \geq 0}$ of subsets of I such that $I_0 = \emptyset$, $I = \cup I_m$, the set I_m contains exactly m elements and $I_m \subset I_{m+1}$. Observe that for any positive integer m and $e \in I$,

$$\begin{aligned}\psi^{(n)} - AU_e^{(n)} &= \sum_{j \in I} a_j (U_j^{(n)} - U_e^{(n)}) \\ &= \Big(\sum_{j \in I - I_m} a_j (U_j^{(n)} - U_e^{(n)}) \Big) + \Big(\sum_{j \in I_m} a_j (U_j^{(n)} - U_e^{(n)}) \Big).\end{aligned}$$

Hence, by the triangle inequality,

$$\|\psi^{(n)} - AU_e^{(n)}\|_T \le \Big(\sum_{j\in I-I_m} |a_j|(\|U_j^{(n)}\|_T + \|U_e^{(n)}\|_T\Big) + \Big(\sum_{j\in I}|a_j|\Big)\Big(m\max_{j\in I_m}\|U_j^{(n)} - U_e^{(n)}\|_T\Big)$$

and so, for $\epsilon > 0$,

$$P(\|\psi^{(n)} - AU_e^{(n)}\|_T \ge \epsilon) \le \frac{4}{\epsilon}\Big(\sum_{j\in I-I_m}|a_j|\Big)\sup_n\sup_j \mathbb{E}(\|U_j^{(n)}\|_T) + P\Big(m\sum_{j\in I}|a_j|\max_{j\in I_m}\|U_j^{(n)} - U_e^{(n)}\|_T > \frac{\epsilon}{2}\Big).$$

We let first $n\to\infty$, and notice that by condition (6) the second term vanishes. Then, we let $m\to\infty$ and the first term vanishes because

$$\sum_{j\in I-I_m}|a_j| \to 0 \qquad \text{as } \ m\to\infty.$$

If the convergence in (7) is in L_p and $\sup_n\sup_j \mathbb{E}(\|U_j^{(n)}\|_T^p) < \infty$ then, obviously, the convergence in the conclusion of the last part of this lemma holds also in L_p. $\square$

To comment on the conditions used in this lemma, we shall see later on that condition (6) is in a particular context also a necessary condition. To verify condition (7) one needs a certain linear structure of the process $U_j^{(n)}$. In our next lemma, we avoid this condition. Moreover we show that for stochastic processes satisfying condition (5) the asymptotic behavior is then identified by the limiting behavior of roughly a sum of m sequences, where m is arbitrary fixed positive integer.

Lemma 2. *Let $\psi^{(n)} = \psi^{(n)}(t)$, $t\in T = [0,1]^d$, $n = 1,2,\dots$ be a sequence of stochastic processes satisfying conditions* (5) *and* (6)*. Let $I = \cup_m I_m$ for some sequence of increasing subsets I_m of I and assume that for each m,*

$$\sum_{j\in I_m} a_j U_j^{(n)} \Longrightarrow Z_m \text{ as } n\to\infty \tag{8}$$

in $C[0,1]^d$ (or in $D[0,1]^d$ endowed with the uniform norm) where Z_m is a continuous stochastic process. Then, there exists a limiting stochastic process $Z_m \Longrightarrow Z$ as $m\to\infty$ in $C[0,1]^d$ and $\psi^{(n)} \Longrightarrow Z$ as $n\to\infty$ in $C[0,1]^d$.

Proof of Lemma 2. Let

$$\psi_m^{(n)} = \sum_{j\in I_m} a_j U_j^{(n)}$$

Notice that for $h\ge m$,

$$\mathbb{E}(\|\psi_h^{(n)} - \psi_m^{(n)}\|_T) \le 2\Big(\sum_{j\in I-I_m}|a_j|\Big)(\sup_n\sup_j \mathbb{E}(\|U_j^{(n)}\|_T)) \to 0$$

as $h\ge m\to\infty$.

Whence, by the Fatou lemma, the sequence of stochastic processes $(Z_m)_{m\in\mathbb{N}}$ satisfies the Cauchy criterion

$$\mathbb{E}(\|Z_m - Z_h\|_T) \to 0 \quad \text{ as } m,h\to\infty.$$

Therefore, $(Z_m)_{m\in\mathbb{N}}$ has a limiting process Z that has sample paths in $C[0,1]^d$.
We also have

$$\mathbb{E}(\|\psi^{(n)} - \psi_m^{(n)}\|_T) \le 2\Big(\sum_{j\in I-I_m} |a_j|\Big)(\sup_n \sup_j \mathbb{E}(\|U_j^{(n)}\|_T)) \to 0 \text{ as } m\to\infty$$

and thus, by Theorem 3.2. in Billingsley (1999), $\psi^{(n)} \Longrightarrow Z$ as $n\to\infty$ which proves the lemma. □

3. Examples

An i.i.d. example.

We begin by showing that somewhat surprisingly, condition (6) is not just a technical condition and it is necessary in some situations.

Proposition 3. *Let $(U_j^{(n)})_{j,n\in\mathbb{N}}$ be a double array of i.i.d. non-negative random variables with finite mean, $\mathbb{E}(U_1^{(n)}) < \infty$. The following conditions are equivalent:*

(i) $U_1^{(n)} \to^P 0$ *as* $n\to\infty$ *and* $\sup_n \mathbb{E}(U_1^{(n)}) < \infty$.
(ii) *For any non-negative sequence* $(a_j)_{j\in\mathbb{N}}$

$$\textit{if} \ \sum_{j\in\mathbb{N}} a_j < \infty \quad \textit{then} \quad \sum_{j=1}^{\infty} a_j U_j^{(n)} \to^P 0 \ \textit{as}\ n\to\infty.$$

Proof. Implication (i) → (ii) follows from Lemma 1. Now, to prove implication (ii) → (i), we notice that for the fixed n, $(U_j^{(n)})_{j\in\mathbb{N}}$ is an i.i.d. sequence and, the Kolmogorov three series theorem then implies that

$$\text{for any non-negative sequence } (a_j)_{j\in\mathbb{N}} \quad \text{with } \sum_{j\in\mathbb{N}} a_j < \infty$$

we have,

$$\sum_{j=1}^{\infty} \mathbb{E}g(a_j U_j^{(n)}) \to 0 \qquad \text{as } n\to\infty.$$

where $g(x) = xI_{(x\le 1)} + I_{(I>1)}$. Since for any positive a

$$\mathbb{E}g(a\xi) = a\int_0^{1/a} P(\xi \ge t)dt$$

then,

$$\begin{aligned}\sum_{j=1}^{\infty} \mathbb{E}g(a_j U_j^{(n)}) &= \sum_{j=1}^{\infty}\int_0^{1/a_j} P(U_1^{(n)} \ge x)a_j dx \\ &= \sum_{j=1}^{\infty}\int_0^{\infty} P(U_1^{(n)} \ge x)I_{(x\le 1/a_j)}a_j dx \\ &= \int_0^{\infty} P(U_1^{(n)} \ge x)t_a(x)dx\end{aligned}$$

where

$$t_a(x) = \sum_{j=1}^{\infty} I_{(x \le 1/a_j)} a_j .$$

On the other hand, for any non-negative non-increasing function $t(x)$ such that $t(x) \downarrow 0$ we can find $a_j \downarrow 0$ such that

$$t(x) \le t_a(x) \text{ , for all } x \ge x_0.$$

and so, for any non-negative non-increasing function $t(x)$ such that $t(x) \downarrow 0$ we have

$$\int_0^{\infty} P(U_1^{(n)} \ge x) t(x) dx \to 0 \quad \text{ as } n \to \infty \tag{9}$$

Now, we proceed by contradiction and, without loss of generality, assume that

$$\mathbb{E}(U_1^{(n)}) \to \infty \text{ as } n \to \infty. \tag{10}$$

Then, we can find two increasing sequences $n_k \nearrow \infty$, $M_k \nearrow \infty$ and a positive sequence $Q_k \to \infty$ as $k \to \infty$ such that

$$\int_{M_{k-1}}^{M_k} P(U_1^{(n_k)} \ge x) dx \ge Q_k .$$

Now, we take $t(x) = 1/Q_k$ on the half open interval $x \in (M_{k-1}, M_k]$ and $t(x) = 0$ for $x \le x_0$. Then,

$$\int_0^{\infty} P(U_1^{(n_k)} \ge x) t(x) dx \ge \frac{1}{Q_k} \int_{M_{k-1}}^{M_k} P(U_1^{(n_k)} \ge x) dx \ge 1 \not\to 0 \text{ as } k \to \infty$$

and so (10) contradicts (9). □

Linear stationary processes

Let $(\xi_i)_{i \in \mathbb{Z}}$ be a stationary sequence with $\mathbb{E}(|\xi_0|) < \infty$ and $\mathbb{E}(\xi_0) = 0$ and let $\mathcal{I}$ be its invariant σ–field. Define

$$X_k = \sum_{j=-\infty}^{\infty} a_j \xi_{k-j} \quad \text{and assume} \quad \sum_{i=-\infty}^{\infty} |a_i| < \infty. \tag{11}$$

In addition to the notations defined in (3), we remind the notation

$$S_k^{(\xi)} = \sum_{j=1}^{k} \xi_j \ .$$

Our next proposition allows to compare the partial sums of the innovations to the partial sums of the linear process with short memory.

Proposition 4. *Assume the representation* (11) *is satisfied and in addition, there is a constant $C > 0$ and a sequence of positive reals $b_n \to \infty$ such that for all n,*

$$\mathbb{E}\Big(\max_{1 \le j \le n} |S_j^{(\xi)}| \Big) \le C b_n \tag{12}$$

and

$$b_n^{-1} \max_{1 \le j \le n} |\xi_j| \to^P 0 \quad \text{as } n \to \infty \ . \tag{13}$$

Then,

$$\frac{1}{b_n} (\max_{1 \le j \le n} |S_j - AS_j^{(\xi)}|) \to^P 0 \quad \text{as } n \to \infty \ . \tag{14}$$

If the innovations are assumed in L_p, $p \ge 1$, (12) is replaced by $\mathbb{E}(\max_{1 \le j \le n} |S_j^{(\xi)}|^p) \le C b_n^p$ and the convergence in (13) holds in L_p , then

$$\frac{1}{b_n^p} E(\max_{1 \le j \le n} |S_j - AS_j^{(\xi)}|^p) \to 0 \quad \text{as } n \to \infty \ .$$

Proof. This proposition follows from Lemma 1 applied to the representation

$$\frac{1}{b_n} S_{[nt]} = \sum_{i=-\infty}^{\infty} a_i U_i^{(n)} \quad \text{where} \quad U_i^{(n)} = \frac{1}{b_n} \sum_{k=1}^{[nt]} \xi_{k-i}$$

so that conditions (5) and (6) follow from (11) and (12), while the convergence in (7) follows from (13). □

From this proposition, we derive

Theorem 5. *Assume that representation* (11) *and condition* (12) *are satisfied. Moreover assume that the innovations satisfy the invariance principle $b_n^{-1} S_{[nt]}^{(\xi)} \Longrightarrow \eta W(t)$ as $n \to \infty$ where η is $\mathcal{I}$–measurable and W is a standard Brownian motion $[0,1]$ independent on $\mathcal{I}$. Then the linear process also satisfies the invariance principle, i.e. $b_n^{-1} S_{[nt]} \Longrightarrow \eta A W(t)$ as $n \to \infty$.*

Proof. Notice that the convergence in probability in (13) follows from the invariance principle $b_n^{-1} S_{[nt]}^{(\xi)} \Longrightarrow \eta W(t)$ as $n \to \infty$, since the modulus of continuity is convergent to 0. All the conditions in Proposition 4 are then satisfied which imply the conclusion of the theorem. □

From Theorem 5 we easily derive the following useful consequence.

Corollary 6 (L_p-invariance principle). *Assume the representation* (11) *holds and $p \ge 1$. Then,*

$$\text{If} \quad \mathbb{E}(\sup_{0 \le t \le 1} |b_n^{-1} S_{[nt]}^{(\xi)} - \eta W(t)|^p) \to 0 \quad \text{then} \quad \mathbb{E}(\sup_{0 \le t \le 1} |b_n^{-1} S_{[nt]} - \eta A W(t)|^p) \to 0$$

as $n \to \infty$.

Discussion

Applications

Theorem 5 and its Corollary 6 work for many dependent structures such as surveyed in [3, 6, 17], Merlevède, Peligrad and Utev (2006). Various invariance principles can be extended from the original sequence to the linear process with short memory.

Here we mention some traditional and also recently developed dependence conditions for innovations whose partial sums satisfy both a maximal inequality and the invariance principle and therefore Theorem 5 applies.

Let us assume that $(\xi_i)_{i\in\mathbb{Z}}$ is a stationary sequence with $\mathbb{E}(\xi_0^2) < \infty$ and $\mathbb{E}(\xi_0) = 0$ and let $\mathcal{F}_a^b$ be a σ–field generated by ξ_i with indexes $a \leq i \leq b$. For all the structures bellow the family $(max_{1\leq k\leq n} S_k^2/n)_{n\geq 1}$ is uniformly integrable and the conclusion of Corollary 6 holds with $b_n = \sqrt{n}$ and with $p = 2$.

(i) Hannan [9], see also its extension to Hilbert space in [4]:

$$\sum_{n=1}^{\infty} \|P_0(\xi_n)\| < \infty \text{ and } \mathbb{E}(\xi_0|\mathcal{F}_{-\infty}) = 0 \ a.s.$$

where $P_k(X) = \mathbb{E}(X|\mathcal{F}_k) - \mathbb{E}(X|\mathcal{F}_{k-1})$ is the projection operator.

(ii) Newman and Wright [18]: $(\xi_i)_{i\in\mathbb{Z}}$ is a negatively associated sequence or, positively associated sequence with

$$\sum_{k=1}^{\infty} cov(\xi_k, \xi_0) < \infty.$$

(iii) Doukhan, Massart and Rio [7]:

$$\sum_{k=1}^{\infty} \int_0^{\tilde{\alpha}(k)} Q^2(u)du < \infty.$$

where Q denotes the cadlag inverse of the function $t \to P(|\xi_0| > t)$ and $\tilde{\alpha}(k) = \alpha(\mathcal{F}_0, \mathcal{F}_n^n) = \sup\{|P(A\cap B) - P(A)P(B)|\ ;\ A \in \mathcal{F}_0, B \in \mathcal{F}_n^n\}$ is the strongly mixing coefficient.

(iv) Dedecker and Rio [5]:

$$\mathbb{E}(X_0 S_n|\mathcal{F}_0) \text{ converges in } \mathbb{L}_1 .$$

(v) Peligrad and Utev [20], by developing Maxwell and Woodroofe[12]:

$$\sum_{n=1}^{\infty} \frac{\|\mathbb{E}(S_n|\mathcal{F}_0)\|_2}{n^{3/2}} < \infty.$$

(vi) Peligrad, Utev and Wu [22], which guarantees the L_p-invariance principle: For $p \geq 2$.

$$\sum_{n=1}^{\infty} \frac{\|\mathbb{E}(S_n|\mathcal{F}_0)\|_p}{n^{3/2}} < \infty.$$

Remarks

(a) If $\mathbb{E}(\xi_0^2) < \infty$ and $b_n \geq \sqrt{n}$, then condition (13) automatically holds.
(b) If the sequence of innovations is ergodic then there is a nonnegative constant σ such that $\eta = \sigma$ a.s.
(c) The set of indexes Z can be replaced by Z^d where d is a positive integer allowing for the treatment of random fields.
(d) A natural extension is to consider innovations with values in functional spaces that also facilitate the study of estimation and forecasting problems for several

classes of continuous time processes [2]. The linear processes are still defined by the formula (1) with the difference that now, the innovations $(\xi_k)_{k\in\mathbb{Z}}$ are Hilbert space H-valued random elements and the sequence of constants is replaced by the sequence of bounded linear operators $\{a_k\}_{k\in\mathbb{Z}}$ from H to H. In [13], was treated the problem of the central limit theorem for this case under the summability condition

$$\sum_{j=-\infty}^{\infty} \|a_j\|_{L(H)} < \infty,$$

where $\|a_j\|_{L(H)}$ denotes the usual operators. It was discovered that, if this condition is not satisfied, then the central limit theorem fails even for the case of independent innovations. The approach developed in this paper shows that the central limit theorem results stated can be strengthened to the invariance principle (some results in this direction for strongly mixing sequences are established in [14]).

Projective criteria

In this section we apply the general devise in Lemma 2 to derive a non-stationary projective criteria. Let X be an integrable random variable and $\{\mathcal{F}_j,\ j\in\mathbb{Z}\}$ be a non-decreasing filtration of σ-fields, that is $\mathcal{F}_j \subseteq \mathcal{F}_i$ for all $j \le i$. As before, define the projection operator by

$$P_k(X) = \mathbb{E}(X|\mathcal{F}_k) - \mathbb{E}(X|\mathcal{F}_{k-1}).$$

The next proposition gives a linear representation of the process $n^{-1/2}S_{[nt]}$ in terms of a linear combination of processes involving the sequence of projections.

Proposition 7. *Let $(X_i)_{i\in\mathbb{Z}}$ be a square integrable centered sequence which is adapted to the non-decreasing filtration $(\mathcal{F}_i)_{i\in\mathbb{Z}}$. Let $\mathcal{F}_{-\infty} = \bigcap_{i\in\mathbb{Z}} \mathcal{F}_i$ Assume that for all $k = 1,2,\ldots$,*

$$\mathbb{E}(X_k|\mathcal{F}_{-\infty}) = 0 \ a.s. \tag{15}$$

and for all $k = 1,2\ldots$ and $j = 0,1,2,\ldots$

$$\|P_{k-j}(X_k)\| \le p_j \quad \text{where } p_j > 0 \text{ and } \sum_{j=0}^{\infty} p_j < \infty\ . \tag{16}$$

Then, the process $n^{-1/2}S_{[nt]}$ satisfies the representation (5) with

$$a_i = p_i \text{ and } U_k^{(n)} = n^{-1/2} \sum_{i=1}^{[nt]} P_{k-i}(X_k) p_i^{-1}$$

and so $(n^{-1/2}S_{[nt]})$ satisfies the invariance principle when (8) holds.

Proof. From conditions (15) and (16) it follows the following martingale difference decomposition

$$X_k = \sum_{i=0}^{\infty} P_{k-i}(X_k) \tag{17}$$

which proves (5).

On the other hand, to check condition (6) we apply the Doob L_2–maximal inequality

$$\mathbb{E}\Big(\max_{1\le j\le n}\Big|\sum_{k=1}^{j} P_{k-i}(X_k)\Big|^2\Big) \le 4\sum_{k=1}^{n}\|P_{k-i}(X_k)\|^2 \le 4(\sqrt{n})^2(p_i)^2$$

which implies condition (6) and completes the proof of this proposition by Lemma 2. □

If we impose more restrictive degrees of stationarity the conclusion can be strengthen.

Proposition 8. *Let $(X_i)_{i\in\mathbb{Z}}$ be a centered, uniformly square integrable sequence of random variables, adapted to the non-decreasing filtration $(\mathcal{F}_i)_{i\in\mathbb{Z}}$. Assume that conditions* (15) *and* (16) *are satisfied and in addition for each $t\in[0,1]$, and all $m=0,1,2,\dots$*

$$\frac{1}{n}\sum_{j=1}^{[nt]}(P_j(S_{j+m-1}-S_{j-1}))^2 \to^P \eta_m B(t) \text{ as } n\to\infty \tag{18}$$

where $B(t)$ is a non-random non-decreasing function. Then, there is a random variable η such that $n^{-1/2}S_{[nt]}\implies \eta B(t)W(t)$ where W is a standard Brownian motion independent of η.

Proof. We employ the martingale decomposition and then, we use a standard result for martingales. By Proposition 7, we know that the limiting behavior of $n^{-1/2}S_{[nt]}$ is determined by the limiting behavior of the partial sum process

$$n^{-1/2}\sum_{k=1}^{[nt]} Y_k \quad \text{where} \quad Y_k = X_k - \mathbb{E}(X_k|\mathcal{F}_{k-m}),$$

where m is a fixed arbitrary positive integer. Notice that $\mathbb{E}(Y_j|\mathcal{F}_{j-m})=0$ almost surely and so the following variables are properly defined

$$\theta_k=\sum_{j=k}^{\infty}\mathbb{E}(Y_j|\mathcal{F}_k)=\sum_{j=k}^{k+m-1}\mathbb{E}(Y_j|\mathcal{F}_k)=Y_k+Q_k \quad a.s.$$

In particular, it is easy to see that

$$\mathbb{E}(\theta_k|\mathcal{F}_{k-1})=-Y_{k-1}+\sum_{j=k-1}^{\infty}\mathbb{E}(Y_j|\mathcal{F}_{k-1})=-Y_{k-1}+\theta_{k-1}$$

and we derive the following coboundary decomposition

$$S_n^{(Y)}=\sum_{i=1}^{n}Y_i=M_n+(Q_0-Q_n) \quad \text{where} \quad M_n=\sum_{k=1}^{n}(\theta_k-\mathbb{E}(\theta_k|\mathcal{F}_{k-1})).$$

By construction and the conditions of the proposition,

$$n^{-1/2}\max_{t\in[0,1]}|Q_0-Q_{[nt]}| \to^P 0 \text{ as } n\to\infty$$

and so, the limiting behavior of the partial sum process $n^{-1/2}S^{(Y)}_{[nt]}$ is determined by the limiting behavior of the normalized discrete time martingale $n^{-1/2}M_{[nt]}$, $t \in [0,1]$ with uniformly square integrable martingale differences

$$\theta_k - \mathbb{E}(\theta_k|\mathcal{F}_{k-1}) = P_k(S_{k+m-1} - S_{k-1}) \ .$$

Then the proposition follows easily by standard results for the functional central limit theorem for martingales (see [8], or [1]). □

We now define a stationary filtration as in [12], that is we assume that $X_i = g(Y_j, j \le i)$ where $(Y_i)_{i\in\mathbb{Z}}$ is an underlying stationary sequence. Denote by $\mathcal{I}$ its invariant sigma field and by $(\mathcal{G}_i)_{i\in\mathbb{Z}}$ an increasing filtration of sigma fields $\mathcal{G}_i = \sigma(Y_j, j \le i)$. For the case when for every i, $\xi_i = Y_i$, and $g(Y_j, j \le i) = Y_i$, then $\mathcal{G}_i$ is simply the sigma algebra generated by ξ_j, $j \le i$. From the above Proposition 8, by using the stationary ergodic theorem we easily derive the stationary projective criterion contained in the next theorem.

We shall derive a class of invariance principles for linear type statistics. The central limit theorem was treated in [19, 21].

Theorem 9. *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence with $\mathbb{E}(X_0) = 0$ and $\mathbb{E}(X_0^2) < \infty$ and stationary filtration $(\mathcal{G}_i)_{i\in\mathbb{Z}}$. Let $\mathcal{G}_{-\infty} = \bigcap_{i\in\mathbb{Z}} \mathcal{G}_i$. Assume that*

$$\mathbb{E}(X_0|\mathcal{G}_{-\infty}) = 0 \quad a.s. \quad and \quad \sum_{i\ge 1} \|P_0(X_i)\|_2 < \infty \ .$$

Then, there exists an $\mathcal{I}$–measurable positive random variable η such that for any Lipschitz function g,

$$n^{-1/2}\sum_{i=1}^{[nt]} g(i/n)X_i \Longrightarrow \eta B_g(t)W(t) \ \text{as } n \to \infty$$

where W is a standard Brownian motion independent of η and $B_g(t) = \sqrt{\int_0^t g^2(x)dx}$.

Proof. Since $g(x)$ is Lipschitz, therefore bounded, by Proposition 8, we have only to check condition (18) that reduces to establishing the convergence

$$\frac{1}{n}\sum_{j=1}^{[nt]} g^2(j/n)(P_j(S_{j+m-1} - S_{j-1}))^2 \to^P \eta_m \int_0^t g^2(x)dx \tag{19}$$

as $n \to \infty$. Let

$$G(j/n) = g^2(j/n) \ , \quad \psi_j = P_j(S_{j+m-1} - S_{j-1}).$$

Notice that by the Birkhoff-Khintchine ergodic theorem there exists the limit

$$\frac{1}{n}\sum_{j=1}^{n} \psi_j \to \eta_m \quad a.s., \text{ as } n \to \infty.$$

On the other hand, because the function $G(x) = g^2(x)$ is bounded and continuous, therefore Riemann integrable, we derive

$$\frac{1}{n}\sum_{j=1}^{[nt]} G(j/n) \to \int_0^t g(x)^2 dx \text{ as } n \to \infty,$$

and, in order to prove (19) it is enough to show that

$$\frac{1}{n}\sum_{j=1}^{[nt]} G(j/n)(\psi_j - \eta_m) \to 0 \quad a.s., \text{ as } n \to \infty. \tag{20}$$

Let

$$U_j = \sum_{i=1}^{j}(\psi_i - \eta_m)$$

and notice that

$$\sup_j(|U_j|/j) < \infty \quad \text{and} \quad \sup_{j\geq N}(|U_j|/j) \to 0 \text{ as } N \to \infty \tag{21}$$

almost surely.

We easily get the representation

$$\sum_{j=1}^{[nt]} G(j/n)(\psi_j - \eta_m) = G([nt]/n)U_{[nt]} + \sum_{j=1}^{[nt]-1}(G(j/n) - G([j+1]/n))U_j.$$

Clearly for any fixed t, $G([nt]/n)U_{[nt]}/n \to 0$ almost surely. On the other hand, since g is Lipschitz, so is $G = g^2$. Therefore there is a constant K such that $|G(j/n) - G([j-1]/n)| \leq K/n$ and thus

$$\begin{aligned}\left|\frac{1}{n}\sum_{j=1}^{[nt]-1}(G(j/n) - G([j+1]/n))U_j\right| &\leq \frac{1}{n}\sum_{j=1}^{n} j|G(j/n) - G([j+1]/n)||U_j/j| \\ &\leq \frac{1}{n}\sup_j |U_j/j| \left(\sum_{j=1}^{[\sqrt{n}]}\frac{Kj}{n}\right) + \frac{1}{n}\sup_{j\geq\sqrt{n}}|U_j/j|\left(\sum_{j=1}^{n}\frac{Kj}{n}\right) \\ &\to 0 \text{ as } n \to \infty\end{aligned}$$

almost surely, which proves (20) and thus completes the proof of the theorem. □

We notice that for stationary sequences, when the filtration is not stationary the result is not true if we only assume the summability of projective norms. We can have the decomposition in (5) but condition (6) will not be satisfied and the invariance principle will fail.

Proposition 10. *There exists a stationary Gaussian positively associated linear process*

$$X_k = \sum_{i=0}^{\infty} t_i Y_{k-i}$$

where $t_i \geq 0$ and Y_i is an i.i.d. sequence of standard normal Gaussian variables and a non-decreasing filtration $\mathcal{F}_k$ such that

(i) *X_k is adapted to $\mathcal{F}_k$.*
(ii) *$E(X_k|\mathcal{F}_{-\infty}) = 0$ a.s. for all k.*
(iii)

$$\sup_{k\geq 0}\sum_{j\in\mathbb{Z}}\|P_j(X_k)\| < \infty.$$

(iv) $\sigma_n^2 = \mathrm{Var}(S_n)$ *is not regularly varying with index* 1, *more exactly, there exists a positive* c *and a subsequence* $k_n \to \infty$ *such that* $\mathrm{Var}(S_{k_n}) \geq ck_n^2/\ln^4(k_n)$.

(v) $\sigma_n^{-1} S_{[nt]}$ *does not satisfy the invariance principle, i.e does not converge to the standard Brownian motion.*

Proof. Let $(Y_k)_{k\in\mathbb{Z}}$ be an i.i.d. sequence of standard normal variables. Let $\mathcal{G}_i = \sigma(Y_j, j \leq i)$, $\mathcal{G}_\infty = \sigma(Y_j, j \in \mathbb{Z})$. For a positive integer r, let

$$n_r = 4^r \quad \text{and} \quad u_r = (n_{r+1} - n_r)^{-1/2} r^{-4} = 1/(3r^4 2^r).$$

We take

$$t_j = u_r \quad \text{when} \quad n_r < j \leq n_{r+1}, r = 1, 2, \ldots$$

We also take $t_i = 0$ when $i < n_1$. Next, as a filtration we take

$$\mathcal{F}_i = \mathcal{G}_{-n_r} \quad \text{when} \quad -n_{r+1} < i \leq -n_r, r = 1, 2, \ldots$$

and in addition

$$\mathcal{F}_i = \mathcal{G}_\infty \quad \text{when} \quad i \geq -n_1.$$

We notice first that X_k is $\mathcal{F}_k$–measurable and so the sequence $(X_k)_{k\in\mathbb{Z}}$ is adapted to the filtration $(\mathcal{F}_k)_{k\in\mathbb{Z}}$, which proves (i).

Clearly,

$$\mathcal{F}_{-\infty} = \bigcap_{i\in\mathbb{Z}} \mathcal{F}_i = \bigcap_{r\in\mathbb{N}} \mathcal{G}_{-n_r} = \mathcal{G}_{-\infty},$$

therefore,

$$\mathbb{E}(X_k|\mathcal{F}_{-\infty}) = \mathbb{E}(X_k|\mathcal{G}_{-\infty}) = 0 \ \ a.s.$$

which proves (ii).

Now, let us compute the projection operator. For $i \geq -n_1$, we have for all X_k, since they are $\mathcal{G}_\infty$–measurable,

$$\mathbb{E}(X_k|\mathcal{F}_i) = \mathbb{E}(X_k|\mathcal{G}_\infty) = X_k \quad a.s.$$

implying that, for $i \geq -n_1$,

$$P_i(X_k) = \mathbb{E}(X_k|\mathcal{F}_i) - \mathbb{E}(X_k|\mathcal{F}_{i-1}) = 0 \quad a.s.$$

Now, for i such that $-n_{r+1} < i - 1 < i \leq -n_r$,

$$\begin{aligned} P_i(X_k) &= \mathbb{E}(X_k|\mathcal{F}_i) - \mathbb{E}(X_k|\mathcal{F}_{i-1}) \\ &= \mathbb{E}(X_k|\mathcal{G}_{-n_r}) - \mathbb{E}(X_k|\mathcal{G}_{-n_r}) = 0 \quad a.s. \end{aligned}$$

Finally, for i such that $-n_{r+1} = i - 1 < i \leq -n_r$,

$$\begin{aligned} P_i(X_k) &= \mathbb{E}(X_k|\mathcal{F}_i) - \mathbb{E}(X_k|\mathcal{F}_{i-1}) = \mathbb{E}(X_k|\mathcal{G}_{-n_r}) - \mathbb{E}(X_k|\mathcal{G}_{-n_{r+1}}) \\ &= \mathbb{E}\Big(\sum_{j=0}^{\infty} t_j Y_{k-j}|\mathcal{G}_{-n_r}\Big) - \mathbb{E}\Big(\sum_{j=0}^{\infty} t_j Y_{k-j}|\mathcal{G}_{-n_{r+1}}\Big) \\ &= \Big(\sum_{j=k+n_r}^{\infty} t_j Y_{k-j}\Big) - \Big(\sum_{j=k+n_{r+1}}^{\infty} t_j Y_{k-j}\Big) \\ &= \sum_{j=k+n_r+1}^{k+n_{r+1}} t_j Y_{k-j} \quad a.s. \end{aligned}$$

and hence, for such an i,

$$\|P_i(X_k)\|_2^2 = \sum_{j=k+n_r+1}^{k+n_{r+1}} t_j^2$$

By construction, the sequence t_j is non-increasing and so

$$\|P_i(X_k)\|_2^2 \leq \sum_{j=n_r+1}^{n_{r+1}} t_j^2 = (n_{r+1} - n_r)u_r^2 = r^{-8}$$

which implies that

$$\sup_{k\geq 0} \sum_j \|P_j(X_k)\|_2 \leq \sum_{r=1}^{\infty} r^{-4} < \infty$$

proving (iii).

To compute the variance, we observe that

$$\mathrm{Var}(S_n) = \sum_{i=-\infty}^{\infty} \Big(\sum_{k=1}^{n} t_{k-i}\Big)^2 .$$

So, for $n = n_{r+1}$, $-[n_{r+1}/3] \leq i \leq 0$ and $[n_{r+1}/2] \leq k \leq n_{r+1}$, we have $n_r < k - i$ and therefore

$$\begin{aligned}
\mathrm{Var}(S_{n_{r+1}}) &\geq \sum_{i=-[n_{r+1}/3]}^{0} \Big(\sum_{k=[n_{r+1}/2]}^{n_{r+1}} t_{k-i}\Big)^2 \\
&\geq \sum_{i=-[n_{r+1}/3]}^{0} \Big(\sum_{k=[n_{r+1}/2]}^{n_{r+1}} u_r\Big)^2 \\
&\geq (u_r n_{r+1})^2 n_{r+1}/12 = \Big(\frac{4^{r+1}}{3r^4 2^r}\Big)^2 \frac{4^{r+1}}{12} = \frac{4^{2(r+1)}}{9r^8} = \frac{n_{r+1}^2}{9r^8}
\end{aligned}$$

and so

$$\mathrm{Var}(S_{n_{r+1}}) \geq n_{r+1}^2/(9\log_4(n_{r+1}))$$

which proves (iv). We conclude that the variance is not regularly varying with index 1. It is well known however that the weak convergence of $\sigma_n^{-1} S_{[nt]}$ to $W(t)$, standard Brownian motion, implies that variance is regularly varying with index 1 and then the invariance principle cannot hold. □

References

[1] Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley, New York.

[2] Bosq, D. (2000). *Linear Processes in Function Spaces. Theory and Applications. Lecture Notes Statistics* **149**. Springer.

[3] Bradley, R. C. (2003). Introduction to strong mixing conditions, Vol. 2. Technical Report, Department of Mathematics, Indiana University, Bloomington. Custom Publishing of I.U., Bloomington.

[4] Dedecker, J. and Merlevède, F. (2003). The conditional central limit theorem in Hilbert spaces. *Stochastic Processes and Their Applications* **108** 229–262.

[5] Dedecker, J. and Rio, E. (2000). On the functional central limit theorem for stationary processes. *Ann. Inst. H. Poincaré* **34** 1–34.

[6] Doukhan, P. (1994). Mixing properties and examples. *Lecture Notes in Statistics* **85**. Springer–Verlag.

[7] Doukhan, P., Massart, P. and Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.* **30** 63–82.

[8] Hall, P. and Heyde, C. (1980). *Martingale Limit Theorem and Its Applications.* Academic Press, New York.

[9] Hannan, E. J. (1979). The central limit theorem for time series regression. *Stochastic Process. Appl.* **9** 281–289.

[10] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** 349–382

[11] Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff Publishing, Groningen. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

[12] Maxwell, M., Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *The Annals of Probability* **28** 713–724.

[13] Merlevède, F., Peligrad, M. and Utev, S. (1997). Sharp conditions for the CLT of linear processes in a Hilbert Space. *Journal of Theoretical Probability* **10** 681–693.

[14] Merlevède, F. (2003). On the central limit theorem and its weak invariance principle for strongly mixing sequences with values in a Hilbert space via martingale approximation. *Journal of Theoretical Probability* **16** 625–653.

[15] Merlevède, F. and Peligrad, M. (2006). On the weak invariance principle for stationary sequences under projective criteria. *Journal of Theoretical Probability.* Published online September 1, 2006.

[16] Merlevède, F., Peligrad, M. and Utev, S. (2006). Recent advances in invariance principles for stationary sequences. *Probability Surveys* **3** 1–36

[17] Peligrad, M. (1986). Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables. In *Dependence in Probability and Statistics (Oberwolfach, 1985), Progr. Probab. Statist.* **11**. Birkhäuser Boston, Boston, MA, pp. 193–223.

[18] Newman, C.M. and Wright, A.L. (1981). An invariance principle for certain dependent sequences. *The Annals of Probability* **9** 671–675.

[19] Peligrad, M. and Utev, S. (1997). Central limit theorem for linear processes. *The Annals of Probability* **25** 443–456.

[20] Peligrad, M. and Utev, S. (2005). A new maximal inequality and invariance principle for stationary sequences. *The Annals of Probability* **33** 798–815.

[21] Peligrad, M. and Utev, S. (2006). Central limit theorem for stationary linear processes. *The Annals of Probability* **34**, to appear. http://arxiv.org/abs/math.PR/0509682

[22] Peligrad, M., Utev S and Wu W. B. (2005). A maximal L(p)- Inequality for stationary sequences and application. *Proc. AMS*, to appear. http://www.geocities.com/irina8/lp.pdf

[23] Wu, W.B. and Woodroofe, M. (2004). Martingale approximations for sums of stationary processes. *The Annals of Probability* **32** 1674–1690.

[24] Wu, W. B. and Min, W. (2005). On linear processes with dependent innovations. *Stochastic Processes and Their Applications* **115** 939–958.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 33–52

DOI: 10.1214/074921706000000743

Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above

Iosif Pinelis[1]

Michigan Technological University

Abstract: Let $(S_0, S_1, \dots)$ be a supermartingale relative to a nondecreasing sequence of σ-algebras $H_{\le 0}, H_{\le 1}, \dots$, with $S_0 \le 0$ almost surely (a.s.) and differences $X_i := S_i - S_{i-1}$. Suppose that $X_i \le d$ and $\mathsf{Var}(X_i | H_{\le i-1}) \le \sigma_i^2$ a.s. for every $i = 1, 2, \dots$, where $d > 0$ and $\sigma_i > 0$ are non-random constants. Let $T_n := Z_1 + \dots + Z_n$, where $Z_1, \dots, Z_n$ are i.i.d. r.v.'s each taking on only two values, one of which is d, and satisfying the conditions $\mathsf{E} Z_i = 0$ and $\mathsf{Var} Z_i = \sigma^2 := \frac{1}{n}(\sigma_1^2 + \dots + \sigma_n^2)$. Then, based on a comparison inequality between generalized moments of S_n and T_n for a rich class of generalized moment functions, the tail comparison inequality

$$\mathsf{P}(S_n \ge y) \le c\, \mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y + \tfrac{h}{2}) \quad \forall y \in \mathbb{R}$$

is obtained, where $c := e^2/2 = 3.694\dots$, $h := d + \sigma^2/d$, and the function $y \mapsto \mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y)$ is the least log-concave majorant of the linear interpolation of the tail function $y \mapsto \mathsf{P}(T_n \ge y)$ over the lattice of all points of the form $nd + kh$ ($k \in \mathbb{Z}$). An explicit formula for $\mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y + \frac{h}{2})$ is given. Another, similar bound is given under somewhat different conditions. It is shown that these bounds improve significantly upon known bounds.

1. Introduction

To begin with, consider normalized Khinchin-Rademacher sums $\varepsilon_1 a_1 + \dots + \varepsilon_n a_n$, where the ε_i's are i.i.d. Rademacher random variables (r.v.'s), with $\mathsf{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$, and the a_i's are real numbers such that $a_1^2 + \dots + a_n^2 = 1$. Whittle [27] (cf. Haagerup [10]) established the sharp form

$$\mathsf{E}\, f(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n) \le \mathsf{E}\, f\big(\tfrac{1}{\sqrt{n}}(\varepsilon_1 + \dots + \varepsilon_n)\big) \le \mathsf{E}\, f(Z) \tag{1.1}$$

of Khinchin's inequality [16] for the power moment functions $f(x) = |x|^p$ with $p \ge 3$, where $Z \sim N(0,1)$. An exponential version of inequality (1.1), with the moment functions $f(x) = e^{\lambda x}$ for $\lambda \in \mathbb{R}$, follows from a result of Hoeffding [12]. An immediate corollary of that is the exponential inequality

$$\mathsf{P}\,(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n \ge x) \le e^{-x^2/2} \quad \forall x \ge 0. \tag{1.2}$$

(In fact, Hoeffding [12] obtains more general results; cf. Remark 2.2 below.) This upper bound, $e^{-x^2/2}$, invites a comparison with an "ideal" upper bound, of the

[1]Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan 49931, USA, e-mail: ipinelis@math.mtu.edu

AMS 2000 subject classifications: primary 60E15, 60G42, 60G48, 60G50.

Keywords and phrases: supermartingales, martingales, upper bounds, probability inequalities, generalized moments.

form $c\,\mathsf{P}(Z \ge x)$ for some absolute constant $c > 0$, which one might expect to have in view of (1.1). Since $\mathsf{P}(Z \ge x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ as $x \to \infty$, this comparison suggests that a factor of order $\frac{1}{x}$ (for large x) is "lost" in (1.2). It turns out that the cause of this loss is that the class of the exponential moment functions is too small. For $\alpha > 0$, consider the following class of functions $f\colon \mathbb{R} \to \mathbb{R}$:

$$\mathcal{F}_+^{(\alpha)} := \Big\{ f\colon f(u) = \int_{-\infty}^{\infty} (u-t)_+^\alpha \, \mu(\mathrm{d}t) \text{ for some Borel measure } \mu \ge 0 \text{ on } \mathbb{R} \\ \text{and all } u \in \mathbb{R} \Big\},$$

where $x_+ := \max(0, x)$ and $x_+^\alpha := (x_+)^\alpha$. Define $\mathcal{F}_-^{(\alpha)}$ as the class of all functions of the form $u \mapsto f(-u)$, where $f \in \mathcal{F}_+^{(\alpha)}$. Let $\mathcal{F}^{(\alpha)} := \{f + g\colon f \in \mathcal{F}_+^{(\alpha)}, g \in \mathcal{F}_-^{(\alpha)}\}$.

Proposition 1.1. *For every natural α, one has $f \in \mathcal{F}_+^{(\alpha)}$ iff f has finite derivatives $f^{(0)} := f$, $f^{(1)} := f', \dots, f^{(\alpha-1)}$ on $\mathbb{R}$ such that $f^{(\alpha-1)}$ is convex on $\mathbb{R}$ and $f^{(j)}(-\infty+) = 0$ for $j = 0, 1, \dots, \alpha - 1$.*

The proof of this and other statements (whenever a proof is necessary) is deferred to Section 3.

It follows that, for every $t \in \mathbb{R}$, every $\beta \ge \alpha$, and every $\lambda > 0$, the functions $u \mapsto (u-t)_+^\beta$ and $u \mapsto e^{\lambda(u-t)}$ belong to $\mathcal{F}_+^{(\alpha)}$, while the functions $u \mapsto |u-t|^\beta$ and $u \mapsto \cosh \lambda(u-t)$ belong to $\mathcal{F}^{(\alpha)}$.

Remark 1.2. Eaton [5] (cf. [7, 21]) obtained inequality (1.1) for a class of moment functions, which essentially coincides with the class $\mathcal{F}^{(3)}$, as seen from [21, Proposition A.1]. Since the class $\mathcal{F}^{(3)}$ is much richer than the class of exponential moment functions, Eaton [6] conjectured (based on asymptotics and numerics) that his inequality (1.1) for $f \in \mathcal{F}^{(3)}$ implies the inequality $\mathsf{P}\,(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n \ge x) \le \frac{2e^3}{9} \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ for all $x > \sqrt{2}$, so that the "lost" factor $\frac{1}{x}$ would be restored. A stronger form of this conjecture was proved by Pinelis [21]:

$$\mathsf{P}\,(\varepsilon_1 a_1 + \dots + \varepsilon_n a_n \ge x) \le \frac{2e^3}{9}\,\mathsf{P}(Z \ge x) \quad \forall x \in \mathbb{R}; \tag{1.3}$$

a multivariate analogue of (1.3) was also obtained there.

Later it was realized (Pinelis [22]) that it is possible to extract (1.3) from (1.1) for all $f \in \mathcal{F}^{(3)}$ because the tail function of the normal distribution is log-concave. The following is a special case of Theorem 4 of Pinelis [23]; see also Theorem 3.11 of Pinelis [22].

Theorem 1.3. *Suppose that $\alpha > 0$, ξ and η are real-valued r.v.'s, and the tail function $u \mapsto \mathsf{P}(\eta \ge u)$ is log-concave on $\mathbb{R}$. Then the comparison inequality*

$$\mathsf{E} f(\xi) \le \mathsf{E} f(\eta) \quad \forall f \in \mathcal{F}_+^{(\alpha)} \tag{1.4}$$

implies

$$\mathsf{P}(\xi \ge x) \le c_\alpha\,\mathsf{P}(\eta \ge x) \quad \forall x \in \mathbb{R}, \tag{1.5}$$

where

$$c_\alpha := \Gamma(\alpha+1)(e/\alpha)^\alpha.$$

Moreover, the constant factor c_α is the best possible one in (1.5).

A similar result for the special case when $\alpha = 1$ is due to Kemperman and is contained in the book by Shorack and Wellner [26, pages 797–799].

Remark 1.4. As follows from [22, Remark 3.13], a useful point is that the requirement of the log-concavity of the tail function $q(u) := \mathsf{P}(\eta \geq u)$ in Theorem 1.3 can be relaxed by replacing q with any (e.g., the least) log-concave majorant of q. However, then the optimality of the constant factor c_α is not guaranteed.

Note that $c_3 = 2e^3/9$, which is the constant factor in (1.3). Bobkov, Götze, and Houdre [4] discovered a simpler proof of a variant of (1.3) with a constant factor $12.0099\ldots$ in place of $2e^3/9 = 4.4634\ldots$. The value of the constant factor is obviously important in statistical applications. The upper bound in (1.3) improves Chernoff-Hoeffding's bound $e^{-x^2/2}$ in (1.2) for all $x > 1.3124\ldots$. On the other hand, the bound in [4] does so only for all $x > 4.5903\ldots$, when $\mathsf{P}(Z \geq x) < 2.22 \times 10^{-6}$. The proof in [4] was direct (rather than based on a moment comparison inequality of the form (1.4)). Of course, this does not imply that the direct methods are inferior. In fact, this author has certain ideas to combine the direct and indirect methods to further improve the constant factors.

A stronger, "discrete" version of (1.3) was obtained in [23, Theorem 5], as follows. Let $\eta_1, \ldots, \eta_n$ be independent zero-mean r.v.'s such that $|\eta_i| \leq 1$ almost surely (a.s.) for all i, and let $b_1, \ldots, b_n$ be any real numbers such that $b_1^2 + \cdots + b_n^2 = n$. Then

$$\mathsf{P}(b_1\eta_1 + \cdots + b_n\eta_n \geq x) \leq c_3\, \mathsf{P}(\varepsilon_1 + \cdots + \varepsilon_n \geq x) \tag{1.6}$$

for all values x that are taken on by the r.v. $\varepsilon_1 + \cdots + \varepsilon_n$ with nonzero probability. Clearly, (1.3) follows from (1.6) by the central limit theorem.

In this paper, we provide new upper bounds on generalized moments and tails of real-valued (super)martingales. It is well known that such bounds can be used, in particular, to obtain concentration-type results; see e.g. [17, 18, 22].

2. Upper bounds on generalized moments and tails of (super)martingales

Theorem 2.1. *Let $(S_0, S_1, \ldots)$ be a supermartingale relative to a nondecreasing sequence of σ-algebras $H_{\leq 0}, H_{\leq 1}, \ldots$, with $S_0 \leq 0$ a.s. and differences $X_i := S_i - S_{i-1}$, $i = 1, 2, \ldots$. Suppose that for every $i = 1, 2, \ldots$ there exist non-random constants $d_i > 0$ and $\sigma_i > 0$ such that*

$$X_i \leq d_i \quad \textit{and} \tag{2.1}$$

$$\mathsf{Var}(X_i | H_{\leq i-1}) \leq \sigma_i^2 \tag{2.2}$$

a.s. Then, for all $n = 1, 2, \ldots$,

$$\mathsf{E}f(S_n) \leq \mathsf{E}f(T_n) \quad \forall f \in \mathcal{F}_+^{(2)}, \quad \textit{where} \tag{2.3}$$

$$T_n := Z_1 + \cdots + Z_n \tag{2.4}$$

and $Z_1, \ldots, Z_n$ are independent r.v.'s such that each Z_i takes on only two values, one of which is d_i, and satisfies the conditions

$$\mathsf{E}Z_i = 0 \quad \textit{and} \quad \mathsf{Var}Z_i = \sigma_i^2; \quad \textit{that is,}$$

$$\mathsf{P}(Z_i = d_i) = \frac{\sigma_i^2}{d_i^2 + \sigma_i^2} \quad \textit{and} \quad \mathsf{P}\left(Z_i = -\frac{\sigma_i^2}{d_i}\right) = \frac{d_i^2}{d_i^2 + \sigma_i^2}.$$

Let us explain in general terms how such a result as Theorem 2.1 can be proved. First, it is not difficult to reduce Theorem 2.1 to the case when $(S_0, S_1, \dots)$ is a martingale with $S_0 = 0$ a.s. Then it is not difficult to reduce the situation to the case of one random summand X, so that the problem becomes to find – for any given $t \in \mathbb{R}$, $d > 0$, and $\sigma > 0$ – the maximum of $\mathsf{E} f_t(X)$ subject to the restrictions $X \le d$ a.s., $\mathsf{E}X = 0$, and $\mathsf{E}X^2 \le \sigma^2$, where $f_t(x) := (x-t)_+^2$. Then, using arguments going back to Chebyshev and Hoeffding [11, 13–15], one sees that here an extremal r.v. X takes on at most three distinct values with nonzero probability. In fact, because of a special relation between the objective-function $f_t(x) = (x-t)_+^2$ and the restriction-functions 1, x, and x^2, one can see that an extremal r.v. X takes on only two distinct values with nonzero probability; this makes the distribution of an extremal r.v. X uniquely determined by d and σ; in particular, the extremal distribution does not depend on the value of the parameter t of the objective-function f_t. Thus the result follows.

An alternative approach is based on duality [22, Eq. (4)], and such an approach is actually used in the proof of Theorem 2.1 given in Section 3 below. According to the latter approach, one should search for a tight upper bound on the objective-function $f_t(x)$ over all $x \le d$; this upper bound must be of the form $Ax^2 + Bx + C$ (which is a linear combination of the restriction-functions 1, x, and x^2), so that $f_t(x) \le Ax^2 + Bx + C$ for all $x \le d$. It is not difficult to see that, for such a tight upper bound, the equality $f_t(x) = Ax^2 + Bx + C$ is attained for at most two distinct values of $x \in (-\infty, d]$. This again implies that an extremal r.v. X takes on only two distinct values with nonzero probability, whence the result. (See the mentioned proof for details.)

Remark 2.2. In the case when (S_i) is a martingale, Theorem 2.1 is a result of Bentkus [1, 3], who used essentially the "duality" approach. Moreover, using Schur convexity arguments similar to those in Eaton [5], he also showed that, in the case $d_i \equiv d$, for every $f \in \mathcal{F}_+^{(2)}$ the right-hand side $\mathsf{E} f(T_n)$ of inequality (2.3) is maximized – under the condition $\sigma_1^2 + \cdots + \sigma_n^2 = n\sigma^2 = \text{const}$ – when the Z_i's are identically distributed; this fact was earlier established by Hoeffding [12, (2.10)] for $f(x) \equiv e^{\lambda x}$ with $\lambda > 0$. Finally, for such i.i.d. Z_i's with $d_i \equiv d$ and $\sigma_i \equiv \sigma$, Bentkus used the method given in [22] (cf. Theorem 1.3 and Remark 1.4 above) to extract the upper bound of the form

$$\mathsf{P}(S_n \ge y) \le c_2 \mathsf{P}^{\mathsf{LC}}(T_n \ge y) \quad \forall y \in \mathbb{R}, \tag{2.5}$$

where the function $y \mapsto \mathsf{P}^{\mathsf{LC}}(T_n \ge y)$ is the least log-concave majorant of the tail function $y \mapsto \mathsf{P}(T_n \ge y)$ on $\mathbb{R}$. Note that $c_2 = e^2/2 = 3.694\dots$.

Note also that the distribution of r.v. T_n in (2.5) is a shifted and re-scaled binomial distribution, concentrated on the lattice, say $L_{n,d,h}$, of all points of the form $nd + kh$ $(k \in \mathbb{Z})$, where

$$h := d + \sigma^2/d.$$

Here and henceforth, we assume that $d_i \equiv d$, unless indicated otherwise.

Since the tail function of the binomial distribution is log-concave on $\mathbb{Z}$ (see e.g. [23, Remark 13]), one has $\mathsf{P}^{\mathsf{LC}}(T_n \ge y) = \mathsf{P}(T_n \ge y)$ for all y in the lattice $L_{n,d,h}$.

Inequality (2.5) can be significantly improved. Let the function $y \mapsto \mathsf{P}^{\mathsf{Lin}}(T_n \ge y)$ denote the linear interpolation of the function $y \mapsto \mathsf{P}(T_n \ge y)$ over the lattice $L_{n,d,h}$, so that

$$\mathsf{P}^{\mathsf{Lin}}(T_n \ge y) = (1-\gamma)\mathsf{P}(T_n \ge nd + kh) + \gamma \mathsf{P}(T_n \ge nd + (k+1)h)$$

if $\gamma := (y - nd - kh)/h \in [0,1]$ for some $k \in \mathbb{Z}$. Let then the function $y \mapsto \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geq y)$ denote the least log-concave majorant of the function $y \mapsto \mathsf{P}^{\mathsf{Lin}}(T_n \geq y)$ on $\mathbb{R}$.

Theorem 2.3. *Suppose that the conditions of Theorem 2.1 hold with $d_i \equiv d$. As before, let $\sigma^2 := \frac{1}{n}(\sigma_1^2 + \cdots + \sigma_n^2)$ and $h := d + \sigma^2/d$. Then*

$$\mathsf{P}(S_n \geq y) \leq c_2 \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geq y + \tfrac{h}{2}) \quad \forall y \in \mathbb{R}. \tag{2.6}$$

Because the tail function $y \mapsto \mathsf{P}(T_n \geq y)$ decreases very rapidly, the shift $\frac{h}{2}$ in $\mathsf{P}^{\mathsf{Lin,LC}}(T_n \geq y + \frac{h}{2})$ generally provides quite a substantial improvement. As will be shown later in this paper (Proposition 2.7), in almost all practically important cases the bound (2.6) is better, or even much better, than (2.5).

In Subsection 2.1 (Proposition 2.10), we will also provide an explicit expression for $\mathsf{P}^{\mathsf{Lin,LC}}(T_n \geq y + \frac{h}{2})$.

That $(S_0, S_1, \dots)$ in Theorems 2.1 and 2.3 is allowed to be a supermartingale (rather than only a martingale) makes it convenient to use the simple but powerful truncation tool. (Such a tool was used, for example, in [20] to prove limit theorems for large deviation probabilities in Banach spaces based only on precise enough probability inequalities and without using Cramér's transform, the standard device in the theory of large deviations.) Thus, for instance, one immediately has the following corollary of Theorem 2.3.

Corollary 2.4. *Suppose that all conditions of Theorem 2.1 hold except possibly for condition* (2.1)*. Then for all $y \in \mathbb{R}$ and $d > 0$*

$$\mathsf{P}(S_n \geq y) \leq \mathsf{P}\big(\max_{1 \leq i \leq n} X_i \geq d\big) + c_2\, \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geq y + \tfrac{h}{2}) \tag{2.7}$$

$$\leq \sum_{1 \leq i \leq n} \mathsf{P}\,(X_i \geq d) + c_2\, \mathsf{P}^{\mathsf{Lin,LC}}(T_n \geq y + \tfrac{h}{2}). \tag{2.8}$$

These bounds are much more precise than the exponential bounds in [8, 9, 19].

Remark 2.5. By the Doob inequality, inequality (2.6) holds for the maximum, $M_n := \max_{0 \leq k \leq n} S_k$, in place of S_n. This follows because (i) all functions of class $\mathcal{F}_+^{(2)}$ are convex and (ii) in view of Lemma 3.1 on page 41, one may assume without loss of generality that (S_i) is a martingale. Similarly, inequalities (2.7) and (2.8) hold for M_n in place of S_n.

In a similar manner, under conditions (2.1)–(2.2) and with

$$b := \sqrt{b_1^2 + \cdots + b_n^2}, \quad \text{where} \quad b_i := \max(d_i, \sigma_i), \tag{2.9}$$

Bentkus [2, 3] obtained the following extensions of inequalities (1.6) and (1.3), respectively:

$$\mathsf{P}(S_n \geq y) \leq c_3 \mathsf{P}^{\mathsf{LC}}\Big(\sum_{i=1}^n \varepsilon_i \geq \frac{y\sqrt{n}}{b}\Big) \quad \forall y \in \mathbb{R} \quad \text{and} \tag{2.10}$$

$$\mathsf{P}(S_n \geq y) \leq c_3 \mathsf{P}(Z \geq y/b) \quad \forall y \in \mathbb{R}.$$

The upper bound in (2.10) can be improved in a similar manner, as follows.

Proposition 2.6. *Under conditions* (2.1), (2.2), *and* (2.9),

$$\mathsf{P}(S_n \geq y) \leq c_3 \mathsf{P}^{\mathsf{Lin,LC}}\Big(\sum_{i=1}^n \varepsilon_i \geq 1 + \frac{y\sqrt{n}}{b}\Big) \quad \forall y \in \mathbb{R}. \tag{2.11}$$

In (2.10) and (2.11), the d_i's are allowed to differ from one another.

Note that the expression $\mathsf{P}^{\mathsf{Lin,LC}}\big(\sum_{i=1}^n \varepsilon_i \ge 1 + \frac{y\sqrt{n}}{b}\big) = \mathsf{P}^{\mathsf{Lin,LC}}\big(\sum_{i=1}^n \frac{b}{\sqrt{n}}\varepsilon_i \ge y+\frac{b}{\sqrt{n}}\big)$ is a special case of the expression $\mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y+\frac{h}{2})$, with $\frac{h}{2} = d = \sigma = \frac{b}{\sqrt{n}}$.

From the "right-tail" bounds stated above, "left-tail" and "two-tail" ones immediately follow. For instance, if condition $X_i \le d_i$ in Theorem 2.1 is replaced by $|X_i| \le d_i$, then inequality (2.3) holds with $\mathcal{F}^{(2)}$ in place of $\mathcal{F}_+^{(2)}$ provided that $(S_0, S_1, \dots)$ is a martingale with $S_0 = 0$ a.s.

In order to present an explicit formula for the upper bound in (2.6) and compare it with the upper bound in (2.5), it is convenient to rescale the r.v. T_n, taking on values in the lattice $L_{n,d,h}$ of all points of the form $nd + kh$ ($k \in \mathbb{Z}$), so that the rescaled r.v., say

$$B_n := \frac{q}{d}T_n + np, \quad \text{where} \quad p := \frac{\sigma^2}{d^2+\sigma^2} \quad \text{and} \quad q := 1-p = \frac{d^2}{d^2+\sigma^2}, \tag{2.12}$$

is binomial with parameters n and p. Then for all $y \in \mathbb{R}$, with

$$x := \frac{q}{d}y + np, \tag{2.13}$$

one has

$$\begin{aligned}
\mathsf{P}(T_n \ge y) &= \mathsf{P}(B_n \ge x) =: Q_n(x),\\
\mathsf{P}^{\mathsf{LC}}(T_n \ge y) &= \mathsf{P}^{\mathsf{LC}}(B_n \ge x) =: Q_n^{\mathsf{LC}}(x),\\
\mathsf{P}^{\mathsf{Lin}}(T_n \ge y) &= \mathsf{P}^{\mathsf{Lin}}(B_n \ge x) =: Q_n^{\mathsf{Lin}}(x),\\
\mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y) &= \mathsf{P}^{\mathsf{Lin,LC}}(B_n \ge x) =: Q_n^{\mathsf{Lin,LC}}(x), \quad \text{so that}
\end{aligned}$$

$$\mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y + \tfrac{h}{2}) = Q_n^{\mathsf{Lin,LC}}(x + \tfrac{1}{2}). \tag{2.14}$$

Here the function $x \mapsto \mathsf{P}^{\mathsf{Lin,LC}}(B_n \ge x)$ is defined as the least log-concave majorant on $\mathbb{R}$ of the function $x \mapsto \mathsf{P}^{\mathsf{Lin}}(B_n \ge x)$, which is in turn defined as the linear interpolation of the tail function $x \mapsto \mathsf{P}(B_n \ge x)$ over the lattice $\mathbb{Z}$. Similarly, the function $x \mapsto \mathsf{P}^{\mathsf{LC}}(B_n \ge x)$ is defined as the least log-concave majorant of the function $x \mapsto \mathsf{P}(B_n \ge x)$ on $\mathbb{R}$.

Note also that $B_n \in [0, n]$ a.s.

Now one is ready to state the following comparison between the upper bounds in (2.5) and (2.6).

Proposition 2.7. *Here relation* (2.13) *between y and x is assumed.*

(i) *Equation*

$$\ln\frac{1-u}{-\ln u} - 1 - \frac{1}{2}\,\frac{(1+u)\ln u}{1-u} = 0 \tag{2.15}$$

in $u \in (0,1)$ has exactly one solution, $u = u_* := 0.00505778\dots$.

(ii) *The upper bound in* (2.6) *is no greater than that in* (2.5) *for all $x \le j_{**}$, where*

$$j_{**} := \left\lfloor \frac{n - u_{**}\frac{q}{p}}{1 + u_{**}\frac{q}{p}} \right\rfloor \quad \textit{and} \tag{2.16}$$

$$u_{**} := \frac{u_*}{1-u_*} = 0.00508349\dots; \tag{2.17}$$

since u_{**} is small, j_{**} is rather close to n unless $\frac{q}{p}$ is large.

(iii) *Moreover, the upper bound in* (2.6) *is no greater than that in* (2.5) *for all* $x \le n$ *provided that*

$$n \le \frac{p}{q}\,\frac{1}{u_{**}} = \frac{p}{q}\,196.714\ldots\,. \tag{2.18}$$

In particular, (2.6) *works better than* (2.5) *for all* $x \le n$ *if* $n \le 196$ *and* $p \ge \frac{1}{2}$.

Since $\sum_{i=1}^{n} \varepsilon_i$ is a special case of T_n (with $d = \sigma = 1$), one immediately obtains the following corollary to Proposition 2.7.

Corollary 2.8. *The upper bound in* (2.11) *is no greater than that in* (2.10) *for all* $x \le j_{**}$, *where* $x := \frac{y\sqrt{n}}{2b} + \frac{n}{2}$ *and* $j_{**} := \left\lfloor \frac{n-u_{**}}{1+u_{**}} \right\rfloor$. *Moreover, the upper bound in* (2.11) *is no greater than that in* (2.10) *for all* $x \le n$ *provided that* $n \le 196$.

Of course, the restriction $x \le j_{**}$ (even though very weak) is only sufficient, but not necessary for the upper bound in (2.6) to be no greater than that in (2.5).

Moreover, (2.6) may work very well even when p is small. For example, in Figure 1 one can see the graph of the ratio

$$r(x) := \frac{Q_n^{\mathsf{Lin,LC}}\left(x + \frac{1}{2}\right)}{Q_n^{\mathsf{LC}}(x)} = \frac{c_2\,\mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y + \frac{h}{2})}{c_2\,\mathsf{P}^{\mathsf{LC}}(T_n \ge y)}$$

of the "new" upper bound – that in (2.6), to the "old" one – that in (2.5), for $n = 30$ and $p = \frac{3}{100}$. In the same figure, one can also see the graph of the new upper bound $q(x) := \min\left(1, c_2\,Q_n^{\mathsf{Lin,LC}}\left(x + \frac{1}{2}\right)\right)$, which is a very rapidly decreasing tail function. Proposition 2.7(ii) guaranteed that, for these n and p, the new upper bound will be an improvement of the old one (that is, one will have $r(x) \le 1$) at least for all $x \in (-\infty, j_{**}] = (-\infty, 25]$, which is rather close to what one can see in the picture. Note that, by Theorem 2.3, for $x \ge 25$ one has $\mathsf{P}(S_n \ge y) \le c_2 Q_n^{\mathsf{Lin,LC}}\left(25 + \frac{1}{2}\right) \approx 3.44 \times 10^{-33}$. Thus, the new upper bound is not an improvement only if the tail probability $\mathsf{P}(S_n \ge y)$ is less than 3.45×10^{-33}. On the other hand, for instance, $r(4) \approx 0.58$; that is, the new upper bound is approximately a 42% improvement of the old one for $x = 4$; at that, the new upper bound is $\approx 0.026 = 2.6\%$, a value quite in a common range in statistical practice.

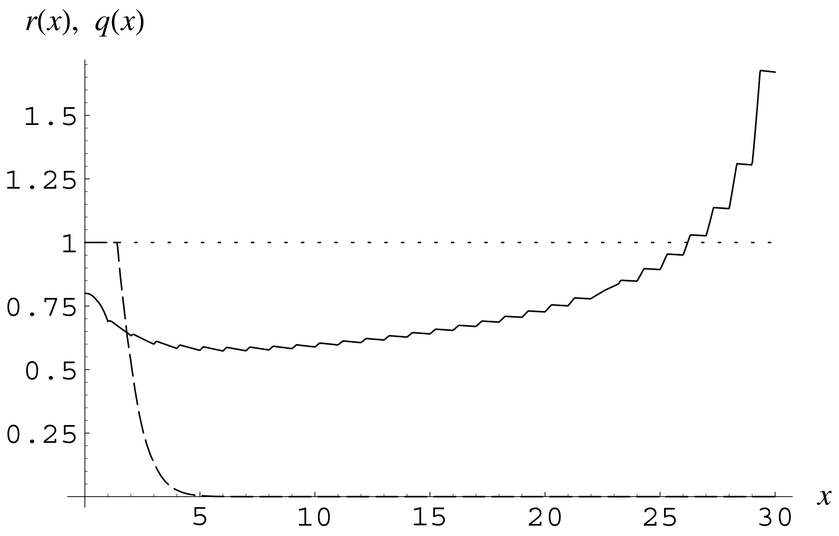

FIG 1. $r(x)$, *solid*; $q(x) = \min\left(1, c_2\,Q_n^{\mathsf{Lin,LC}}\left(x + \frac{1}{2}\right)\right)$, *long dashes.*

The upper bound in (2.6), $c_2\,\mathsf{P}^{\mathsf{Lin,LC}}(T_n \ge y + \frac{h}{2}) = c_2\,Q_n^{\mathsf{Lin,LC}}\left(x+\frac12\right)$, is rather simple to compute, as described in Proposition 2.10, given in a separate subsection, Subsection 2.1. An underlying reason for this simplicity is that the "discrete" tail function, $\mathbb{Z} \ni j \mapsto Q_n(j)$, of the binomial distribution is log-concave [23, Remark 13]; therefore, it turns out (by Propositions 2.10 and 2.9) that the value $Q_n^{\mathsf{Lin,LC}}\left(x+\frac12\right)$ can be computed locally: for a certain function $\mathcal{Q}$ which depends only on its 7 arguments, one has $Q_n^{\mathsf{Lin,LC}}\left(x+\frac12\right) = \mathcal{Q}(x,n,j_*,q_{k-1},q_k,q_{k+1},q_{k+2})$ for all $x\in\mathbb{R}$ and $p\in(0,1)$, where $j_* := \lfloor (n+1)p\rfloor + 1$, $k := \lfloor x\rfloor$, and $q_j := Q_n(j)$ for all $j\in\mathbb{Z}$.

2.1. *How to compute the expression* $Q_n^{\mathsf{Lin,LC}}\left(x+\frac12\right)$ *in* (2.14)

First here, we need to introduce some notation.

For each $j = 1,\dots,n$ such that $j > (n+1)p$, let

$$x_j := j - \frac12 + \frac{q_j}{p_j} + \frac{\frac{q_j}{p_{j-1}} - \frac{q_j}{p_j}}{\ln\frac{p_{j-1}}{p_j}}, \tag{2.19}$$

$$y_j := j - \frac12 + \frac{q_j}{p_{j-1}} + \frac{\frac{q_j}{p_{j-1}} - \frac{q_j}{p_j}}{\ln\frac{p_{j-1}}{p_j}} = x_j - q_j\left(\frac{1}{p_j} - \frac{1}{p_{j-1}}\right), \tag{2.20}$$

where, for all $j\in\mathbb{Z}$,

(2.21)

$$q_j := Q_n(j) = \mathsf{P}(B_n \ge j) \quad\text{and}\quad p_j := q_j - q_{j+1} = \mathsf{P}(B_n = j) = \binom{n}{j} p^j q^{n-j};$$

note that, for each $j = 0,\dots,n$,

$$p_{j-1} > p_j \iff j > (n+1)p \iff j \ge j_* := \lfloor (n+1)p\rfloor + 1, \tag{2.22}$$

so that, for all $j \in \mathbb{Z}\cap[j_*,n]$, one has $\ln\frac{p_{j-1}}{p_j} > 0$, and so, x_j and y_j are well defined, and $x_j > y_j$; note that $j_* \ge 1$. Let also

$$x_j := j + \tfrac12 \quad\text{and}\quad y_j := j - \tfrac12 \quad\text{for integer } j \ge n+1. \tag{2.23}$$

Proposition 2.9. *For all integer $j \ge j_*$, one has*

$$j - \tfrac32 < j - 1 < y_j \le j - \tfrac12 < x_j \le y_{j+1} \le j + \tfrac12;$$

moreover, if $j \le n$, then $y_j < j - \frac12$ and $x_j < y_{j+1}$.

By Proposition 2.9, for all integer $j \ge j_*$, the intervals

$$\delta_j := (y_j, x_j)$$

are non-empty, with the endpoints

$$y_j \in (j-1, j-\tfrac12] \subset (j-\tfrac32, j-\tfrac12] \quad\text{and}\quad x_j \in (j-\tfrac12, j+\tfrac12];$$

moreover, the intervals δ_j are strictly increasing in $j \ge j_*$: $\delta_j < \delta_{j+1}$ for all $j \ge j_*$, where we use the following convention for any two subsets A and B of $\mathbb{R}$:

$$A < B \overset{\text{def}}{\iff} (x < y \quad \forall x\in A\ \forall y\in B).$$

For all integer $j \ge j_*$ and all $x \in \delta_j$, introduce the interpolation expression

$$Q_n^{\mathsf{Interp}}(x;j) := Q_n^{\mathsf{Lin}}\left(y_j + \tfrac12\right)^{1-\delta} Q_n^{\mathsf{Lin}}\left(x_j + \tfrac12\right)^{\delta}, \quad\text{where } \delta := \frac{x - y_j}{x_j - y_j}. \tag{2.24}$$

Proposition 2.10. *For all real x,*

$$Q_n^{\mathsf{Lin,LC}}\left(x+\tfrac{1}{2}\right)=Q(x):=\begin{cases}Q_n^{\mathsf{Interp}}(x;j) & \text{if } \exists j\in\mathbb{Z}\cap[j_*,n]\ \ x\in\delta_j,\\ Q_n^{\mathsf{Lin}}\left(x+\tfrac{1}{2}\right) & \text{otherwise.}\end{cases}\tag{2.25}$$

A few comments are in order here:

- the function Q in (2.25) is well defined, because the δ_j's are pairwise disjoint;
- $Q_n^{\mathsf{Lin,LC}}\left(x+\frac{1}{2}\right)$ is easy to compute by (2.25) because, in view of Proposition 2.9, the condition $x\in\delta_j$ for $j\in\mathbb{Z}\cap[j_*,n]$ implies that j equals either $\lfloor x\rfloor$ or $\lfloor x\rfloor+1$. In particular, $Q_n^{\mathsf{Lin,LC}}\left(x+\frac{1}{2}\right)=Q_n^{\mathsf{Lin}}\left(x+\frac{1}{2}\right)=0$ for all $x\geq n+\frac{1}{2}$.

3. Proofs

Proof of Proposition 1.1. The "only if" part follows because $f\in\mathcal{F}^{(\alpha)}$ for a natural α implies that the (right) derivative $f'(u)=\alpha\int_{-\infty}^{\infty}(u-t)_+^{\alpha-1}\,\mu(\mathrm{d}t)$. Vice versa, if f satisfies the conditions listed after "iff" in the statement of Proposition 1.1, then one can use the Fubini theorem repeatedly to see that for all real u

$$\begin{aligned}f(u)&=\int_{-\infty}^{u}f'(t)\,\mathrm{d}t=\int_{-\infty}^{u}\mathrm{d}t\int_{-\infty}^{t}f''(s)\,\mathrm{d}s=\int_{-\infty}^{u}(u-s)f''(s)\,\mathrm{d}s=\dots\\&=\int_{-\infty}^{u}\frac{(u-s)^{\alpha-1}}{(\alpha-1)!}f^{(\alpha)}(s)\,\mathrm{d}s=\int_{-\infty}^{u}\frac{(u-s)^{\alpha}}{\alpha!}\,\mathrm{d}f^{(\alpha)}(s)=\int_{-\infty}^{\infty}(u-s)_+^{\alpha}\,\mu(\mathrm{d}s),\end{aligned}$$

where $f^{(\alpha)}$ is the (nondecreasing) right derivative of the convex function $f^{(\alpha-1)}$ and $\mu(\mathrm{d}s):=\mathrm{d}f^{(\alpha)}(s)/\alpha!$. ☐

Theorem 2.1 can be rather easily reduced to the case when (S_n) is a martingale. This is implied by the following lemma.

Lemma 3.1. *Let (S_i) be a supermartingale as in Theorem 2.1, so that conditions (2.1) and (2.2) are satisfied. For $i=1,2,\dots,$ let*

$$\tilde{X}_i:=(1-\gamma_{i-1})X_i+\gamma_{i-1}d_i,\quad \text{where}\quad \gamma_{i-1}:=\frac{\mathsf{E}_{i-1}X_i}{\mathsf{E}_{i-1}X_i-d_i};\tag{3.1}$$

E_j and Var_j denote, respectively, the conditional expectation and variance given $X_1,\dots,X_j$. Then $\tilde{X}_i$ is $H_{\leq i}$-measurable,

$$X_i\leq\tilde{X}_i\leq d_i,\quad \mathsf{E}_{i-1}\tilde{X}_i=0,\quad \text{and}\quad \mathsf{Var}_{i-1}\tilde{X}_i\leq\mathsf{Var}_{i-1}X_i\leq\sigma_i^2\quad \text{a.s.}$$

Proof. The conditions $d_i>0$ and $\mathsf{E}_{i-1}X_i\leq 0$ imply that that $\gamma_{i-1}\in[0,1)$. Now (3.1) and the inequality $X_i\leq d_i$ yield $X_i\leq\tilde{X}_i\leq d_i$. Moreover, (3.1) yields $\mathsf{E}_{i-1}\tilde{X}_i=(1-\gamma_{i-1})\mathsf{E}_{i-1}X_i+\gamma_{i-1}d_i=0$ and $\mathsf{Var}_{i-1}\tilde{X}_i=(1-\gamma_i)^2\,\mathsf{Var}_{i-1}X_i\leq\mathsf{Var}_{i-1}X_i$. ☐

Theorem 2.1 is mainly based on the following lemma, which also appeared as [1, (12)] and [3, Lemma 4.4(i), with condition $\mathsf{E}X=0$ missing there].

Lemma 3.2. *Let X be a r.v. such that $\mathsf{E}X=0$, $\mathsf{Var}X\leq\sigma^2$, and $X\leq d$ a.s. for some non-random constants $\sigma>0$ and $d>0$. Let $a:=\sigma^2/d^2$, and let X_a be a r.v. taking on values $-a$ and 1 with probabilities $\frac{1}{1+a}$ and $\frac{a}{1+a}$, respectively. Then*

$$\mathsf{E}f(X)\leq\mathsf{E}f(d\cdot X_a)\quad\forall f\in\mathcal{F}_+^{(2)}.\tag{3.2}$$

Proof. (Given here for the reader's convenience and because it is short and simple.) By homogeneity, one may assume that $d = 1$ (otherwise, rewrite the lemma in terms of X/d in place of X). Note that $X_a \le 1$ with probability 1, $\mathsf{E}X_a = 0$, and $\mathsf{E}X_a^2 = \sigma^2$. Let here $f_t(x) := (x-t)_+^2$ and

$$h_t(x) := \frac{(1-t)_+^2}{(1-t_a)^2}\,(x-t_a)^2, \quad \text{where} \quad t_a := \min(t, -a).$$

Then it is easy to check (by considering the cases $t \ge 1$, $-a \le t \le 1$, and $t \le -a$) that $f_t(x) \le h_t(x)$ for all $x \le 1$, and $f_t(x) = h_t(x)$ for $x \in \{1, -a\}$. Therefore, $\mathsf{E}f_t(X) \le \mathsf{E}h_t(X) \le \mathsf{E}h_t(X_a) = \mathsf{E}f_t(X_a)$ (the second inequality here follows because $h_t(x)$ is a quadratic polynomial in x with a nonnegative coefficient of x^2, while $\mathsf{E}X = \mathsf{E}X_a$ and $\mathsf{E}X^2 \le \mathsf{E}X_a^2$). Now the lemma follows by the definition of the class $\mathcal{F}_+^{(\alpha)}$ and the Fubini theorem. □

Proof of Theorem 2.1. This proof is based in a standard manner on Lemma 3.2, using also Lemma 3.1. Indeed, by Lemma 3.1, one may assume that $\mathsf{E}_{i-1}X_i = 0$ for all i. Let $Z_1, \dots, Z_n$ be r.v.'s as in the statement of Theorem 2.1, which are also independent of the X_i's, and let

$$R_i := X_1 + \cdots + X_i + Z_{i+1} + \cdots + Z_n.$$

Let $\tilde{\mathsf{E}}_i$ denote the conditional expectation given $X_1, \dots, X_{i-1}, Z_{i+1}, \dots, Z_n$. Note that, for all $i = 1, \dots, n$, $\tilde{\mathsf{E}}_i X_i = \mathsf{E}_{i-1}X_i = 0$ and $\tilde{\mathsf{E}}_i X_i^2 = \mathsf{E}_{i-1}X_i^2$; moreover, $R_i - X_i = X_1 + \cdots + X_{i-1} + Z_{i+1} + \cdots + Z_n$ is a function of $X_1, \dots, X_{i-1}$, $Z_{i+1}, \dots, Z_n$. Hence, by Lemma 3.2, for any $f \in \mathcal{F}_+^{(2)}$, $\tilde{f}_i(x) := f(R_i - X_i + x)$, and all $i = 1, \dots, n$, one has $\tilde{\mathsf{E}}_i f(R_i) = \tilde{\mathsf{E}}_i \tilde{f}_i(X_i) \le \tilde{\mathsf{E}}_i \tilde{f}_i(Z_i) = \tilde{\mathsf{E}}_i f(R_{i-1})$, whence $\mathsf{E}f(S_n) \le \mathsf{E}f(R_n) \le \mathsf{E}f(R_0) = \mathsf{E}f(T_n)$ (the first inequality here follows because $S_0 \le 0$ a.s. and any function f in $\mathcal{F}_+^{(2)}$ is nondecreasing). □

Proof of Theorem 2.3. In view of Theorem 2.1 and Remark 2.2, one has

$$\mathsf{E}f(\tilde{S}_n) \le \mathsf{E}f(B_n) \quad \forall f \in \mathcal{F}_+^{(2)}, \tag{3.3}$$

where $\tilde{S}_n := \frac{q}{d}S_n + np$ and B_n is defined by (2.12). Let U be a r.v., which is independent of B_n and uniformly distributed between $-\frac{1}{2}$ and $\frac{1}{2}$. Then, by Jensen's inequality, $\mathsf{E}f(B_n) \le \mathsf{E}f(B_n + U)$ for all convex functions f, whence

$$\mathsf{E}f(\tilde{S}_n) \le \mathsf{E}f(B_n + U) \quad \forall f \in \mathcal{F}_+^{(2)}.$$

Observe that the density function of $B_n + U$ is $x \mapsto \sum_{j=0}^n p_j\, \mathbf{I}\{|x-j| < \frac{1}{2}\}$ (where the p_j's are given by (2.21)), and so, the tail function of $B_n + U$ is given by the formula $\mathsf{P}(B_n + U \ge x) = Q_n^{\mathsf{Lin}}\left(x + \frac{1}{2}\right)$ $\forall x \in \mathbb{R}$. Now Theorem 2.3 follows by Theorem 1.3, Remark 1.4, and (2.14). □

Proof of Proposition 2.6. This proof is quite similar to the proof of Theorem 2.3. Instead of (3.3), here one uses inequality $\mathsf{E}f(\tilde{S}_n) \le \mathsf{E}f(B_n)$ $\forall f \in \mathcal{F}_+^{(3)}$, where $\tilde{S}_n := \frac{\sqrt{n}}{2b}S_n + \frac{n}{2}$. This latter inequality follows from (2.3) (with d_i and σ_i each replaced by $b_i = \max(d_i, \sigma_i)$) and the first one of the inequalities (1.1) $\forall f \in \mathcal{F}_+^{(3)}$ (recall Remark 1.2), taking also into account the inclusion $\mathcal{F}_+^{(3)} \subseteq \mathcal{F}_+^{(2)}$ (which follows e.g. from [23, Proposition 1(ii)]). □

In the following two propositions, which are immediate corollaries of results of [25] and [24], it is assumed that f and g are differentiable functions on an interval $(a,b) \subseteq (-\infty,\infty)$, and each of the functions g and g' is nonzero and does not change sign on (a,b); also, $r := f/g$ and $\rho := f'/g'$.

Proposition 3.1. (Cf. [25, Proposition 1].) *Suppose that* $f(a+) = g(a+) = 0$ *or* $f(b-) = g(b-) = 0$. *Suppose also that* $\rho \nearrow$ *or* $\searrow$*; that is,* ρ *is increasing or decreasing on* (a,b). *Then* $r \nearrow$ *or* $\searrow$*, respectively.*

Proposition 3.2. (Cf. [24, Proposition 1.9]; the condition that $f(a+) = g(a+) = 0$ or $f(b-) = g(b-) = 0$ is not assumed here.) *If* $\rho \nearrow$ *or* $\searrow$ *on* (a,b), *then* $r \nearrow$ *or* $\searrow$ *or* $\nearrow\searrow$ *or* $\searrow\nearrow$ *on* (a,b). *(Here, for instance, the pattern* $\nearrow\searrow$ *for* ρ *on* (a,b) *means that* $\rho \nearrow$ *on* (a,c) *and* $\searrow$ *on* (c,b) *for some* $c \in (a,b)$*; the pattern* $\searrow\nearrow$ *has a similar meaning.) It follows that, if* $\rho \nearrow$ *or* $\searrow$ *or* $\nearrow\searrow$ *or* $\searrow\nearrow$ *on* (a,b), *then* $r \nearrow$ *or* $\searrow$ *or* $\nearrow\searrow$ *or* $\searrow\nearrow$ *or* $\nearrow\searrow\nearrow$ *or* $\searrow\nearrow\searrow$ *on* (a,b).

Lemma 3.3. *Part* (i) *of Proposition 2.7 is true.*

Proof of Lemma 3.3. Let

$$h(u) := \ln\frac{1-u}{-\ln u} - 1 - \frac{1}{2}\,\frac{(1+u)\ln u}{1-u}, \tag{3.4}$$

the left-hand side of (2.15). Here and in rest of the proof of Lemma 3.3, it is assumed that $0 < u < 1$, unless specified otherwise. Then

$$h'(u) = r_1(u) := \frac{f_1(u)}{g_1(u)}, \tag{3.5}$$

where $f_1(u) := 2\ln^2 u + (\frac{1}{u} + 2 - 3u)\ln u + 2(u + \frac{1}{u}) - 4$ and $g_1(u) := -2(1-u)^2\ln u$. Let next

$$r_2(u) := \frac{f_1'(u)}{g_1'(u)} = \frac{f_2(u)}{g_2(u)}, \tag{3.6}$$

where $f_2(u) := (\frac{3}{u} - \frac{1}{u^2})\ln u + \frac{1}{u} - \frac{1}{u^2}$ and $g_2(u) := 4\ln u - \frac{2}{u} + 2$, and then

$$r_3(u) := \frac{f_2'(u)}{g_2'(u)} = \frac{f_3(u)}{g_3(u)}, \tag{3.7}$$

where $f_3(u) := (2-3u)\ln u + 1 + 2u$ and $g_3(u) := 2u(1+2u)$.

One has $\frac{f_3''(u)}{g_3''(u)} = -\frac{1}{8}(\frac{2}{u^2} + \frac{3}{u})$, which is increasing; moreover, $\frac{\mathrm{d}}{\mathrm{d}u}\frac{f_3'(u)}{g_3'(u)}$ tends to $-\infty < 0$ and $-29/50 < 0$ as $u \downarrow 0$ and $u \uparrow 1$, respectively. Hence, by Proposition 3.2, $\frac{f_3'(u)}{g_3'(u)}$ is decreasing (in $u \in (0,1)$).

Next, by (3.7), $r_3'(0+) = \infty > 0$ and $r_3'(1-) = -2/3 < 0$. Hence, by Proposition 3.2, $r_3 \nearrow\searrow$ (on $(0,1)$).

By (3.6), $r_2'(0+) = \infty > 0$ and $f_2(1) = g_2(1) = 0$. Hence, by Propositions 3.1 and 3.2, $r_2 \nearrow\searrow$ (on $(0,1)$).

By (3.5), $r_1'(0+) = \infty > 0$ and $r_1'(1-) = -1/4 < 0$. Hence, by Proposition 3.2, $h' = r_1 \nearrow\searrow$ (on $(0,1)$). Moreover, $h'(0+) = -\infty < 0$ and $h'(1-) = \frac{1}{2} > 0$. Hence, for some $\beta \in (0,1)$, one has $h' < 0$ on $(0,\beta)$ and $h' > 0$ on $(\beta,1)$.

Hence, $h \searrow\nearrow$ on $(0,1)$. Moreover, $h(0+) = \infty$ and $h(1-) = 0$. It follows that the equation $h(u) = 0$ has a unique root $u = u_* \in (0,1)$.

Now Lemma 3.3 follows by (3.4). $\square$

Lemma 3.4. *For all $x \le j_{**}+1$,*

$$Q_n^{\mathsf{Lin}}\left(x+\tfrac12\right) \le Q_n^{\mathsf{LC}}(x). \tag{3.8}$$

Proof of Lemma 3.4. For any given $x \le j_{**}+1$, let

$$j := j_x := \lfloor x\rfloor \quad \text{and} \quad k := k_x := \lfloor x+\tfrac12\rfloor, \quad \text{so that}$$

$$\begin{gathered} j \le x < j+1, \quad k-\tfrac12 \le x < k+\tfrac12, \\ Q_n^{\mathsf{LC}}(x) = q_j^{1-\delta} q_{j+1}^{\delta}, \quad Q_n^{\mathsf{Lin}}\left(x+\tfrac12\right) = (1-\gamma)q_k + \gamma\, q_{k+1}, \quad \text{where} \\ \delta := x-j \in [0,1) \quad \text{and} \quad \gamma := x+\tfrac12-k \in [0,1). \end{gathered}$$

There are only three possible cases: $\delta=0$, $\delta\in[\frac12,1)$, and $\delta\in(0,\frac12)$.

Case 1: $\delta = 0$. This case is simple. Indeed, here $k=j$ and $\gamma=\frac12$, so that

$$Q_n^{\mathsf{LC}}(x) = q_j \ge \tfrac12\, q_j + \tfrac12\, q_{j+1} = Q_n^{\mathsf{Lin}}\left(x+\tfrac12\right),$$

since q_j is nonincreasing in j.

Case 2: $\delta\in[\frac12,1)$. This case is simple as well. Indeed, here $k=j+1$, so that

$$Q_n^{\mathsf{LC}}(x) \ge q_{j+1} \ge (1-\gamma)\, q_{j+1} + \gamma\, q_{j+2} = Q_n^{\mathsf{Lin}}\left(x+\tfrac12\right).$$

Case 3: $\delta\in(0,\frac12)$. In this case, $k=j$ and $\gamma=\delta+\frac12$, so that inequality (3.8) can be rewritten here as $q_j^{1-\delta}q_{j+1}^{\delta} \ge (\frac12-\delta)q_j + (\frac12+\delta)q_{j+1}$ or, equivalently, as

$$F(\delta,u) := u^\delta - \left(\tfrac12-\delta\right) - \left(\tfrac12+\delta\right)u \ge 0, \tag{3.9}$$

where

$$u := \frac{q_{j+1}}{q_j} \in [0,1]; \tag{3.10}$$

note that the conditions $x\le j_{**}+1$ and $j\le x<j+1$ imply $j\le j_{**}+1\le n$ (the latter inequality takes place because, by (2.16), $j_{**}<n$); hence, $q_j\ge q_n>0$, and thus, u is correctly defined by (3.10). Moreover, because both sides of inequality (3.8) are continuous in x for all $x\le n$ and hence for all $x\le j_{**}+1$, it suffices to prove (3.8) only for $x<j_{**}+1$, whence $j\le j_{**}$, and so, by (2.16),

$$j \le \frac{n-u_{**}q/p}{1+u_{**}q/p};$$

the latter inequality is equivalent, in view of (2.21), to $\frac{p_{j+1}}{p_j} \ge u_{**}$, which in turn implies, in view of (2.17), that

$$\frac{q_j}{q_{j+1}} = 1 + \frac{p_j}{q_{j+1}} \le 1 + \frac{p_j}{p_{j+1}} \le 1 + \frac{1}{u_{**}} = \frac{1}{u_*},$$

whence, by (3.10), one obtains $u\ge u_*$. Therefore, the proof in Case 3, and hence the entire proof of Lemma 3.4, is now reduced to the following lemma. □

Lemma 3.5. *Inequality* (3.9) *holds for all $\delta\in(0,\frac12)$ and $u\in[u_*,1]$.*

Proof of Lemma 3.5. Observe first that, for every $\delta \in (0, \frac{1}{2})$,

$$\begin{aligned} &F(\delta,0) = -\left(\tfrac{1}{2}-\delta\right) < 0, \;\; F(\delta,1) = 0, \\ &(\partial_u F)(\delta,1) = -\tfrac{1}{2} < 0, \text{ and } F \text{ is concave in } u \in [0,1]. \end{aligned} \tag{3.11}$$

This implies that, for every $\delta \in (0, \frac{1}{2})$, there exists a unique value

$$u(\delta) \in (0,1)$$

such that

$$F(\delta, u(\delta)) = 0, \tag{3.12}$$

and at that

$$F(\delta, u) \geq 0 \quad \forall \delta \in \left(0, \tfrac{1}{2}\right) \; \forall u \in [u(\delta), 1], \tag{3.13}$$

and $(\partial_u F)(\delta, u(\delta))$ is strictly positive and hence nonzero. Thus, equation (3.12) defines an implicit function $(0, \frac{1}{2}) \ni \delta \mapsto u(\delta) \in (0,1)$. Moreover, since F is differentiable on $(0, \frac{1}{2}) \times (0,1)$ and $(\partial_u F)(\delta, u(\delta)) \neq 0$ for all $\delta \in (0, \frac{1}{2})$, the implicit function theorem is applicable, so that $u(\delta)$ is differentiable in δ for all $\delta \in (0, \frac{1}{2})$. Now, differentiating both sides of equation (3.12) in δ, one obtains

$$u^\delta \ln u + 1 - u + (\delta u^{\delta-1} - \tfrac{1}{2} - \delta) u'(\delta) = 0, \tag{3.14}$$

where u stands for $u(\delta)$.

Let us now show that $u(0+) = u(\frac{1}{2}-) = 0$. To that end, observe first that

$$\sup_{0<\delta<\frac{1}{2}} u(\delta) < 1. \tag{3.15}$$

Indeed, otherwise there would exist a sequence (δ_n) in $(0, \frac{1}{2})$ such that $\varepsilon_n := 1 - u(\delta_n) \downarrow 0$. But then $u(\delta_n)^{\delta_n} = (1-\varepsilon_n)^{\delta_n} = 1 - \varepsilon_n \delta_n + o(\varepsilon_n)$, so that (3.12) would imply

$$0 = 1 - \varepsilon_n \delta_n + o(\varepsilon_n) - \left(\tfrac{1}{2} - \delta_n\right) - \left(\tfrac{1}{2} + \delta_n\right)(1 - \varepsilon_n) = (\tfrac{1}{2} + o(1))\varepsilon_n,$$

which would contradict the fact that $\varepsilon_n = 1 - u(\delta_n) > 0$ for all n.

If it were not true that $u(0+) = 0$, then there would exist a sequence (δ_n) in $(0, \frac{1}{2})$ and some $\varepsilon > 0$ such that $\delta_n \downarrow 0$ while $u(\delta_n) \to \varepsilon$. But then $u(\delta_n)^{\delta_n} \to 1$, so that equation (3.12) would imply

$$u(\delta_n) = \frac{u(\delta_n)^{\delta_n} - \left(\frac{1}{2} - \delta_n\right)}{\frac{1}{2} + \delta_n} \to 1,$$

which would contradict (3.15). Thus, $u(0+) = 0$.

Similarly, if it were not true that $u(\frac{1}{2}-) = 0$, then there would exist a sequence (δ_n) in $(0, \frac{1}{2})$ and some $\varepsilon > 0$ such that $\delta_n \uparrow \frac{1}{2}$ while $u(\delta_n) \to \varepsilon$. But then equation (3.12) would imply

$$u(\delta_n) = \frac{u(\delta_n)^{\delta_n} - \left(\frac{1}{2} - \delta_n\right)}{\frac{1}{2} + \delta_n} \to \varepsilon^{1/2},$$

which would imply $0 < \varepsilon = \varepsilon^{1/2}$, so that $\varepsilon = 1$, which would contradict (3.15). Hence, $u(\frac{1}{2}-) = 0$.

Thus, $(0, \frac{1}{2}) \ni \delta \mapsto u(\delta)$ is a strictly positive continuous function, which vanishes at the endpoints 0 and $\frac{1}{2}$. Therefore, there must exist a point $\delta_* \in (0, \frac{1}{2})$ such that $u(\delta_*) \geq u(\delta)$ for all $\delta \in (0, \frac{1}{2})$. Then one must have $u'(\delta_*) = 0$.

Now equation (3.14) yields

$$u(\delta_*)^{\delta_*} = -\frac{1 - u(\delta_*)}{\ln u(\delta_*)}, \quad \text{whence} \tag{3.16}$$

$$\delta_* = \frac{\ln \frac{1-u(\delta_*)}{-\ln u(\delta_*)}}{\ln u(\delta_*)}. \tag{3.17}$$

In the expression (3.9) for $F(\delta, u)$, replace now u^δ by the right-hand side of (3.16), and then replace δ by the right-hand side of (3.17). Then, recalling (3.12) and slightly re-arranging terms, one sees that $u = u(\delta_*)$ is a root of equation (2.15).

By Lemma 3.3, such a root of (2.15) is unique in $(0, 1)$. It follows that $\max_{0<\delta<1/2} u(\delta) = u(\delta_*) = u_* = 0.00505\ldots$. In view of (3.13), this completes the proof of Lemma 3.5. □

Proof of Proposition 2.7. In this proof, we shall use Propositions 2.9 and 2.10 (which will be proved later in this paper) and the following preliminary remarks.

According to [23, Remark 13], the restriction of the tail function Q_n to the set $\mathbb{Z}$ of all integers is log-concave. Therefore, the logarithm, $\ln Q_n^{\mathsf{LC}}$, of the least log-concave majorant Q_n^{LC} of Q_n can be obtained by the linear interpolation of $\ln Q_n$ over $\mathbb{Z}$, so that

$$Q_n^{\mathsf{LC}}(x) = q_j^{1-(x-j)} q_{j+1}^{x-j} \quad \text{if} \quad j \leq x \leq j+1 \ \& \ j \in \mathbb{Z}, \tag{3.18}$$

where q_j is defined by (2.21). Here and elsewhere, $0^0 := 1$. Recall that the function Q_n^{Lin} is the linear interpolation of the function Q_n over $\mathbb{Z}$, so that

$$Q_n^{\mathsf{Lin}}(x) = (1 - (x - j))q_j + (x - j)q_{j+1} \quad \text{if} \quad j \leq x \leq j+1 \ \& \ j \in \mathbb{Z}. \tag{3.19}$$

Since the function Q_n is non-decreasing and left-continuous, one can note that $Q_n^{\mathsf{Lin}} \geq Q_n$ on $\mathbb{R}$; also, $Q_n^{\mathsf{Lin}} = Q_n$ on $\mathbb{Z}$.

Let us now proceed to the proof of Proposition 2.7.

(i) Part (i) of Proposition 2.7 follows by Lemma 3.3.

(ii) If $x \leq j_{**} + 1$ and x is not in δ_j for any $j \in \mathbb{Z} \cap [j_*, n]$, then, by (2.25) and Lemma 3.4,

$$Q_n^{\mathsf{Lin,LC}}\left(x + \tfrac{1}{2}\right) = Q(x) = Q_n^{\mathsf{Lin}}\left(x + \tfrac{1}{2}\right) \leq Q_n^{\mathsf{LC}}(x). \tag{3.20}$$

If $x \in \delta_j \subset (-\infty, j_{**} + 1]$ for some $j \in \mathbb{Z} \cap [j_*, n]$, then, taking also into account the definition (2.24), (3.20), and the log-concavity of the function Q_n^{LC}, one has, for δ as in (2.24),

$$\begin{aligned}Q_n^{\mathsf{Lin,LC}}\left(x + \tfrac{1}{2}\right) &= Q(x) = Q_n^{\mathsf{Interp}}(x; j)\\ &= Q_n^{\mathsf{Lin}}\left(y_j + \tfrac{1}{2}\right)^{1-\delta} Q_n^{\mathsf{Lin}}\left(x_j + \tfrac{1}{2}\right)^{\delta} \leq Q_n^{\mathsf{LC}}(y_j)^{1-\delta} Q_n^{\mathsf{LC}}(x_j)^{\delta} \leq Q_n^{\mathsf{LC}}(x).\end{aligned}$$

Hence,

$$Q_n^{\mathsf{Lin,LC}}\left(x + \tfrac{1}{2}\right) \leq Q_n^{\mathsf{LC}}(x) \tag{3.21}$$

for all $x \le j_{**}+1$ except maybe when $x \in \delta_j \cap (-\infty, j_{**}+1]$ for some $j \in \mathbb{Z} \cap [j_*, n]$ such that $\delta_j \not\subset (-\infty, j_{**}+1]$. The latter exceptional situation implies that $y_j < j_{**}+1 < x_j$. Hence, by Proposition 2.9,

$$j-1 < y_j < j_{**}+1 < x_j \le j+\tfrac{1}{2}, \tag{3.22}$$

whence $j < j_{**}+2$ and $j \ge j_{**}+1$, so that $j = j_{**}+1$.

It follows that (3.21) holds (at least) for all

$$x \in (-\infty, j_{**}+1] \setminus \delta_{j_{**}+1} = (-\infty, y_{j_{**}+1}] \supset (-\infty, j_{**}],$$

the latter inclusion taking place because of the first inequality in (3.22), for $j = j_{**}+1$. This completes the proof of part (ii) of Proposition 2.7.

(iii) Introduce

$$\tilde{Q}(x) := \begin{cases} Q_n^{\mathsf{LC}}(x) & \text{if } x \le n, \\ q_{n-1}^{n-x} q_n^{x-n+1} & \text{if } x \ge n-1. \end{cases} \tag{3.23}$$

Note that for $x \in [n-1, n]$ the expressions for $\tilde{Q}(x)$ in the two cases in (3.23) coincide with each other, which implies that the function $\tilde{Q}$ is log-concave on $\mathbb{R}$.

Now we need the following lemma, whose proof will be given a bit later.

Lemma 3.6. *Under condition* (2.18), *for all* $x \in [n, n+\frac{1}{2}]$,

$$Q_n^{\mathsf{Lin}}\left(x+\tfrac{1}{2}\right) \le \tilde{Q}(x). \tag{3.24}$$

Under condition (2.18), it is easy to see that $j_{**}+1 \ge n$. Therefore, by Lemma 3.4, one has (3.8) and hence (3.24) for all $x \le n$.

Also, $Q_n^{\mathsf{Lin}}\left(x+\frac{1}{2}\right) = 0$ for all $x \ge n+\frac{1}{2}$.

Thus, by Lemma 3.6, $\tilde{Q}(x) \ge Q_n^{\mathsf{Lin}}\left(x+\frac{1}{2}\right)$ for all $x \in \mathbb{R}$; at that, as noted above, the function $\tilde{Q}$ is log-concave on $\mathbb{R}$. Hence, $Q_n^{\mathsf{Lin,LC}}\left(x+\frac{1}{2}\right) \le \tilde{Q}(x)$ for all $x \in \mathbb{R}$. Now part (iii) of Proposition 2.7 follows in view of (3.23), because $\tilde{Q} = Q_n^{\mathsf{LC}}$ on the interval $(-\infty, n]$. □

Proof of Lemma 3.6. In view of the definitions (3.19) and (3.18), one can rewrite inequality (3.24) as $q_{n-1}^{n-x} q_n^{x-n+1} \ge \left(n+\frac{1}{2}-x\right) q_n + \left(x-n+\frac{1}{2}\right) q_{n+1}$, for all $x \in [n, n+\frac{1}{2}]$. Since $q_{n+1} = 0$, the latter inequality is equivalent to

$$\ln u \le r(\alpha), \tag{3.25}$$

where $u := q_{n-1}/q_n > 1$, $\alpha := x-n \in (0, \frac{1}{2})$, and

$$r(\alpha) := \frac{\ln(\frac{1}{2}-\alpha)}{-\alpha}.$$

Since $r'(0+) = -\infty$, $r'(\frac{1}{2}-) = \infty$, and $((\ln(\frac{1}{2}-\alpha))'_\alpha)/(-\alpha'_\alpha) = 1/(\frac{1}{2}-\alpha)$ is increasing in $\alpha \in (0, \frac{1}{2})$, it follows from Proposition 3.2 that there is a unique value $\alpha_* \in (0, \frac{1}{2})$ such that the function r is decreasing on $(0, \alpha_*)$ and increasing on $(\alpha_*, \frac{1}{2})$, so that $r'(\alpha_*) = 0$ and α_* is the point of minimum of function r on $(0, \frac{1}{2})$. In fact, one has $\alpha_* = 0.3133\ldots$ and $r(\alpha_*) = 5.3566\ldots$.

Therefore, inequality (3.25) can be rewritten as $u \le e^{r(\alpha_*)}$. On the other hand,

$$u = \frac{q_{n-1}}{q_n} = \frac{np^{n-1}q + p^n}{p^n} = n\,\frac{q}{p} + 1,$$

so that it suffices to check that $n\,\frac{q}{p} \le e^{r(\alpha_*)} - 1 = 211.022\ldots$. But this follows from condition (2.18). □

Proof of Proposition 2.9. Assume the condition $j \ge j_*$. For $j \ge n+1$, the inequalities of Proposition 2.9 immediately follow from (2.23).

It remains to consider the case $j \le n$, which implies $p_{j-1} > p_j$, by (2.22). Then it suffices to check four inequalities, $j-1 < y_j < j-\frac12 < x_j < y_{j+1}$. Indeed, the inequality $j-\frac32 < j-1$ is trivial, and the inequality $y_{j+1} \le j+\frac12$ will then follow — for $j=n$, from (2.23); and for $j = j_*, \dots, n-1$, from the inequalities $y_i < i-\frac12$ $\forall i = j_*, \dots, n$.

(i) Checking $j-1 < y_j$. In view of definition (2.20),

$$y_j = j - \tfrac12 + \kappa_j q_j, \tag{3.26}$$

where

$$\kappa_j := \frac{1}{p_{j-1}} + \frac{\frac{1}{p_{j-1}} - \frac{1}{p_j}}{\ln\frac{p_{j-1}}{p_j}} = \frac{1 - \frac{p_{j-1}}{p_j} + \ln\frac{p_{j-1}}{p_j}}{p_{j-1}\ln\frac{p_{j-1}}{p_j}},$$

so that

$$\kappa_j < 0, \tag{3.27}$$

in view of the condition $p_{j-1} > p_j$ and the inequality $\ln u < u-1$ for $u > 1$.

On the other hand, it is well known and easy to verify that the probability mass function (p_k) of the binomial distribution is log-concave, so that the ratio p_k/p_{k-1} is decreasing in k. Hence,

$$q_j = \sum_{k=j}^{n} p_k \le \sum_{k=j}^{\infty} p_j \left(\frac{p_j}{p_{j-1}}\right)^{k-j} = \frac{p_j}{1-\frac{p_j}{p_{j-1}}} =: \hat q_j.$$

Therefore, to check $j-1 < y_j$, it suffices to check that $d_j := \hat y_j - (j-1) > 0$, where $\hat y_j := j - \frac12 + \kappa_j \hat q_j$ (cf. (3.26). But one can see that $d_j = f(u)/(2(u-1)\ln u)$, where $u := \frac{p_{j-1}}{p_j} > 1$ and $f(u) := 2(1-u) + (1+u)\ln u$. Thus, to check $j-1 < y_j$, it suffices to show that $f(u) > 0$ for $u > 1$. But this follows because $f(1) = f'(1) = 0$ and f is strictly convex on $(1,\infty)$.

(ii) Checking $y_j < j - \frac12$. This follows immediately from (3.26) and (3.27).

(iii) Checking $j - \frac12 < x_j$. This follows because $x_j - \big(j - \frac12\big) = q_j(\ln u - 1 + 1/u)/(p_j \ln u) > 0$, where again $u := \frac{p_{j-1}}{p_j} > 1$.

(iv) Checking $x_j < y_{j+1}$. Let first $j \le n-1$, so that $p_j > p_{j+1} > 0$. In view of (2.20) and the obvious identity $1 + \frac{q_{j+1}}{p_j} = \frac{q_j}{p_j}$, one has

$$y_{j+1} = j - \frac12 + \frac{q_j}{p_j} + \frac{\frac{q_{j+1}}{p_j} - \frac{q_{j+1}}{p_{j+1}}}{\ln\frac{p_j}{p_{j+1}}},$$

so that the inequality $x_j < y_{j+1}$, which is being checked, can be rewritten as $q_j r_j > q_{j+1} r_{j+1}$ or, equivalently, as

$$\sum_{k=0}^{\infty} (p_{j+k} r_j - p_{j+1+k} r_{j+1}) > 0, \tag{3.28}$$

where $r_j := (\frac{1}{p_j} - \frac{1}{p_{j-1}})/\ln\frac{p_{j-1}}{p_j}$. Note that $r_j > 0$, $r_{j+1} > 0$, and $p_j r_j = h(v) := (1-v)/-\ln v$, where $v := p_j/p_{j-1} \in (0,1)$. By Proposition 3.1, $h(v)$ is increasing in $v \in (0,1)$. On the other hand, $v = p_j/p_{j-1}$ is decreasing in j, by the mentioned log-concavity of (p_j). It follows that $p_j r_j$ is decreasing in j. Because of this and the same log-concavity, $\frac{p_{j+k}r_j}{p_{j+1+k}r_{j+1}} > \frac{p_j r_j}{p_{j+1}r_{j+1}} > 1 \quad \forall k = 0, 1, \ldots$, which yields (3.28).

Finally, in the case when $j = n \ge j_*$, the inequality $x_j < y_{j+1}$ follows from (2.19) and (2.23), because then

$$x_n = n + \frac{1}{2} + \frac{\dfrac{p_n}{p_{n-1}} - 1}{\ln \dfrac{p_{n-1}}{p_n}} < n + \frac{1}{2} = y_{n+1}.$$

□

Proof of Proposition 2.10.

Step 1. Here we observe that the function Q defined in (2.25) is continuous. Indeed, the function $x \mapsto Q_n^{\mathsf{Lin}}\left(x + \frac{1}{2}\right)$ is defined and continuous everywhere on $\mathbb{R}$. On the other hand, for every integer $j \ge j_*$, the function $x \mapsto Q_n^{\mathsf{Interp}}(x; j)$ is defined and continuous on the interval δ_j; moreover, it continuously interpolates on the interval δ_j between the values of the function $x \mapsto Q_n^{\mathsf{Lin}}\left(x + \frac{1}{2}\right)$ at the endpoints, y_j and x_j, of the interval δ_j. Also, the intervals δ_j with $j \in \mathbb{Z} \cap [j_*, n]$ are pairwise disjoint. Thus, the function Q is continuous everywhere on $\mathbb{R}$.

Step 2. Here we show that the function Q is log-concave. To that end, introduce

$$\ell_j(x) := \ln\left((\tfrac{1}{2} + j - x)q_j + (\tfrac{1}{2} - j + x)q_{j+1}\right) \quad \forall j \in \mathbb{Z}, \quad \text{so that}$$

$$(3.29) \qquad (j - \tfrac{1}{2} \le x < j + \tfrac{1}{2} \;\&\; j \le n) \implies \ln Q_n^{\mathsf{Lin}}\left(x + \tfrac{1}{2}\right) = \ell_j(x);$$

here, the condition $j \le n$ provides for both sides of the equality in (3.29) to be defined. One can check (which is better done using Mathematica or similar software) the basic relations

$$(3.30) \qquad j \in \mathbb{Z} \cap [j_*, n] \implies \ell'_{j-1}(y_j) = \ell'_j(x_j) = \frac{\ell_j(x_j) - \ell_{j-1}(y_j)}{x_j - y_j};$$

these relations do not rely on the fact that the q_j's pertain to a binomial distribution, but only on general relations: $p_{j-1} > p_j > 0$, $p_i = q_i - q_{i+1}$ $\forall i$, and $q_{j+1} \ge 0$, as well as the inequalities $(j-1) - \frac{1}{2} < y_j < (j-1) + \frac{1}{2}$ and $j - \frac{1}{2} < x_j \le j + \frac{1}{2}$, which follow by Proposition 2.9 and ensure that ℓ_j and ℓ_{j-1} are defined and differentiable in neighborhoods of x_j and y_j, respectively. Using the latter relations together with (2.24) and (3.29), one has

$$(3.31) \qquad \frac{\mathrm{d}}{\mathrm{d}x} \ln Q_n^{\mathsf{Interp}}(x; j) = \frac{\ell_j(x_j) - \ell_{j-1}(y_j)}{x_j - y_j} \quad \forall x \in \delta_j \;\forall j \in \mathbb{Z} \cap [j_*, n].$$

Moreover, for all integer $j \le n$,

$$\ell'_j\left(j - \tfrac{1}{2}\right) = \frac{q_{j+1} - q_j}{q_j} = \frac{-p_j}{q_j} \quad \text{and} \quad \ell'_{j-1}\left(j - \tfrac{1}{2}\right) = \frac{-p_{j-1}}{q_j}.$$

Hence and by (2.22), for every integer $j \le n$ one has

$$(3.32) \qquad \ell'_j\left(j - \tfrac{1}{2}\right) \le \ell'_{j-1}\left(j - \tfrac{1}{2}\right) \iff j \le j_* - 1.$$

In view of (3.29), the function $x \mapsto \ln Q_n^{\mathsf{Lin}}\left(x + \frac{1}{2}\right)$ is concave on the interval $[j - \frac{1}{2}, j + \frac{1}{2}]$ for every integer $j \le n$. Note also that $\bigcup_{j \le j_* - 1}\left[j - \frac{1}{2}, j + \frac{1}{2}\right] = \left(-\infty, j_* - \frac{1}{2}\right]$. Hence, by (3.32) and (3.29), the function

$$(3.33) \qquad x \mapsto \ln Q_n^{\mathsf{Lin}}\left(x + \tfrac{1}{2}\right) \quad \text{is concave on the interval } \left(-\infty, j_* - \tfrac{1}{2}\right].$$

In addition to the open intervals $\delta_j = (y_j, x_j)$, introduce the closed intervals
$$\Delta_j := [x_j, y_{j+1}]$$
for integer $j \ge j_*$. Then, by Proposition 2.9 and (2.23), the intervals Δ_j are each nonempty,

$$\delta_{j_*} \cup \Delta_{j_*} \cup \delta_{j_*+1} \cup \Delta_{j_*+1} \cup \dots \cup \delta_n \cup \Delta_n = \big(y_{j_*}, n + \tfrac{1}{2}\big], \tag{3.34}$$

and
$$\delta_{j_*} < \Delta_{j_*} < \delta_{j_*+1} < \Delta_{j_*+1} < \dots < \delta_n < \Delta_n.$$
Thus, the intervals δ_j and Δ_j with $j \in \mathbb{Z} \cap [j_*, n]$ form a partition of the interval $\big(y_{j_*}, n + \frac{1}{2}\big]$. Moreover, for every $j \in \mathbb{Z} \cap [j_*, n]$, by Proposition 2.9, $\Delta_j \subseteq [j - \frac{1}{2}, j + \frac{1}{2}]$, and so, by (3.29), the function

$$x \mapsto \ln Q_n^{\mathsf{Lin}}\big(x + \tfrac{1}{2}\big) \quad \text{is concave on the interval } \Delta_j. \tag{3.35}$$

Also, by (2.24), for every $j \in \mathbb{Z} \cap [j_*, n]$, the function

$$x \mapsto \ln Q_n^{\mathsf{Interp}}(x; j) \text{ is concave (in fact, affine) on the interval } \delta_j. \tag{3.36}$$

By the definition of Q in (2.25), for all $x \in \mathbb{R}$ and all $j \in \mathbb{Z} \cap [j_*, n]$, one has

$$Q(x) = \begin{cases} Q_n^{\mathsf{Lin}}\big(x + \frac{1}{2}\big) & \text{if } x \le y_{j_*}, \\ Q_n^{\mathsf{Interp}}(x; j) & \text{if } x \in \delta_j \ \&\ j \in \mathbb{Z} \cap [j_*, n], \\ Q_n^{\mathsf{Lin}}\big(x + \frac{1}{2}\big) & \text{if } x \in \Delta_j \ \&\ j \in \mathbb{Z} \cap [j_*, n], \\ 0 = Q_n^{\mathsf{Lin}}\big(x + \frac{1}{2}\big) & \text{if } x \ge n + \frac{1}{2}. \end{cases} \tag{3.37}$$

Note also that, by Proposition 2.9, $y_{j_*} \le j_* - \frac{1}{2}$, so that $(-\infty, y_{j_*}] \subseteq (-\infty, j_* - \frac{1}{2}]$. Now it follows from (3.37), (3.33), (3.35), and (3.36) that the function $\ln Q$ is concave on each of the disjoint adjacent intervals

$$(-\infty, y_{j_*}], \delta_{j_*}, \Delta_{j_*}, \delta_{j_*+1}, \Delta_{j_*+1}, \dots, \delta_n, \Delta_n, \tag{3.38}$$

whose union is the interval $(-\infty, n + \frac{1}{2}]$. Moreover, it follows from the continuity of Q (established in Step 1) and formulas (3.30), (3.29), and (3.31) that the function $\ln Q$ is differentiable at all the endpoints $y_{j_*}, x_{j_*}, y_{j_*+1}, x_{j_*+1}, \dots, y_n, x_n$ of the intervals (3.38) except the right endpoint $y_{n+1} = n + \frac{1}{2}$ of the interval Δ_n.

Therefore, the function $\ln Q$ is concave on the interval $(-\infty, n + \frac{1}{2})$. On the other hand, $\ln Q = -\infty$ on the interval $[n + \frac{1}{2}, \infty)$. Thus, it is proved that the function $\ln Q$ is concave everywhere on $\mathbb{R}$.

Step 3. Here we show that

$$Q(x) \ge Q_n^{\mathsf{Lin}}\big(x + \tfrac{1}{2}\big) \tag{3.39}$$

for all real x. In view of (3.37) and (3.34), it suffices to check (3.39) for $x \in \delta_j$ with $j \in \mathbb{Z} \cap [j_*, n]$. By Proposition 2.9, $\delta_j \subseteq (j - 1, j + \frac{1}{2}] \subset [j - \frac{3}{2}, j + \frac{1}{2}]$, for every $j \in \mathbb{Z} \cap [j_*, n]$.

By (3.29), the function $x \mapsto \ln Q_n^{\mathsf{Lin}}\big(x + \frac{1}{2}\big) = \ell_j(x)$ is concave on the interval $[j - \frac{1}{2}, j + \frac{1}{2}]$, for every integer $j \le n$. Hence, (3.30) and (2.24) imply that, for all $x \in \delta_j \cap [j - \frac{1}{2}, j + \frac{1}{2}]$ with $j \in \mathbb{Z} \cap [j_*, n]$,

$$\begin{aligned} \ln Q_n^{\mathsf{Lin}}\big(x + \tfrac{1}{2}\big) &= \ell_j(x) \le \ell_j(x_j) + \ell_j'(x_j)(x - x_j) \\ &= \frac{x_j - x}{x_j - y_j}\, \ell_{j-1}(y_j) + \frac{x - y_j}{x_j - y_j}\, \ell_j(x_j) = \ln Q_n^{\mathsf{Interp}}(x; j) = \ln Q(x), \end{aligned}$$

so that one has (3.39) for all $x \in \delta_j \cap [j - \frac{1}{2}, j + \frac{1}{2}]$ with $j \in \mathbb{Z} \cap [j_*, n]$. Similarly (using inequality $\ell_{j-1}(x) \le \ell_{j-1}(y_j) + \ell'_{j-1}(y_j)(x - y_j)$) it can be shown that (3.39) takes place for all $x \in \delta_j \cap [j - \frac{3}{2}, j - \frac{1}{2}]$ with $j \in \mathbb{Z} \cap [j_*, n]$. This completes Step 3.

Step 4. Here we show that, if $\tilde{Q}$ is a log-concave function on $\mathbb{R}$ such that

$$\tilde{Q}(x) \ge Q_n^{\mathsf{Lin}}\left(x + \tfrac{1}{2}\right) \quad \forall x \in \mathbb{R}, \tag{3.40}$$

then $\tilde{Q} \ge Q$ on $\mathbb{R}$. In view of (3.37), it suffices to check that $\tilde{Q} \ge Q$ on δ_j for every $j \in \mathbb{Z} \cap [j_*, n]$. But, by (3.40), one has $\tilde{Q}(y_j) \ge Q_n^{\mathsf{Lin}}\left(y_j + \frac{1}{2}\right)$ and $\tilde{Q}(x_j) \ge Q_n^{\mathsf{Lin}}\left(x_j + \frac{1}{2}\right)$. Hence, taking into account the log-concavity of $\tilde{Q}$ and (2.24) and (3.37), one has, for all $x \in \delta_j$ with $j \in \mathbb{Z} \cap [j_*, n]$ and δ as in (2.24),

$$\tilde{Q}(x) \ge \tilde{Q}\left(y_j\right)^{1-\delta} \tilde{Q}\left(x_j\right)^{\delta} \ge Q_n^{\mathsf{Lin}}\left(y_j + \tfrac{1}{2}\right)^{1-\delta} Q_n^{\mathsf{Lin}}\left(x_j + \tfrac{1}{2}\right)^{\delta} = Q_n^{\mathsf{Interp}}(x; j) = Q(x).$$

The facts established in Steps 2, 3, and 4 imply that the function Q is indeed the least log-concave majorant of the function $x \mapsto Q_n^{\mathsf{Lin}}\left(x + \frac{1}{2}\right)$. Thus, Proposition 2.10 is proved. ☐

References

[1] Bentkus, V. (2002). A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Lithuanian Math. J.* **42** 262–269. MR1947624

[2] Bentkus, V. (2003). An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.* **16** 161–173. MR1956826

[3] Bentkus, V. (2004). On Hoeffding's inequalities. *Ann. Probab.* **32** 1650–1673. MR2060313

[4] Bobkov, S. G., Götze, F. and Houdré, C. (2001). On Gaussian and Bernoulli covariance representations. *Bernoulli* **7** 439–451. MR1836739

[5] Eaton, M. L. (1970). A note on symmetric Bernoulli random variables. *Ann. Math. Statist.* **41** 1223–1226. MR268930

[6] Eaton, M. L. (1974). A probability inequality for linear combinations of bounded random variables. *Ann. Statist.* **2** 609–614.

[7] Figiel, T., Hitczenko, P., Johnson, W. B., Schechtman, G. and Zinn, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349** 997–1027. MR1390980

[8] Fuk, D. H. (1971). Certain probabilistic inequalities for martingales. *Siberian Math. J.* **14** 131–137. MR0293695

[9] Fuk, D. H. and Nagaev, S. V. (1971). Probabilistic inequalities for sums of independent random variables (in Russian, English summary). *Teor. Verojatnost. i Primenen.* **16** 660–675. MR0293695

[10] Haagerup, U. (1982). The best constants in the Khinchine inequality. *Studia Math.* **70** 231–283. MR0654838

[11] Hoeffding, W. (1955). The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.* **26** 268–275. MR70087

[12] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR144363

[13] Hoeffding, W. and Shrikhande, S. S. (1955). Bounds for the distribution function of a sum of independent, identically distributed random variables. *Ann. Math. Statist.* **26** 439–449. MR72377

[14] Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York–London–Sydney. MR204922

[15] Karr, A. F. (1983). Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.* **8** (1) 74–85. MR703827

[16] Khinchin, A. (1923). Über dyadische Brüche. *Math. Z.* **18** 109–116.

[17] McDiarmid, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics, 1989 (Norwich, 1989).* London Math. Soc. Lecture Note Ser., Vol. **141**. Cambridge Univ. Press, Cambridge, pp. 148–188. MR1036755

[18] McDiarmid, C. (1998). Concentration. In *Probabilistic methods for algorithmic discrete mathematics.* Algorithms Combin., Vol. **16**. Springer, Berlin, pp. 195–248. MR1678578

[19] Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR0542129

[20] Pinelis, I. F. (1985). Asymptotic equivalence of the probabilities of large deviations for sums and maximum of independent random variables (in Russian). *Limit Theorems of Probability Theory. Trudy Inst. Mat.* **5**. Nauka Sibirsk. Otdel., Novosibirsk, pp. 144–173, 176. MR0821760

[21] Pinelis, I. (1994). Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Statist.* **22** (1), 357–368. MR1272088

[22] Pinelis, I. (1998). Optimal tail comparison based on comparison of moments. *High Dimensional Probability (Oberwolfach, 1996). Progr. Probab.* **43**. Birkhäuser, Basel, pp. 297–314. MR1652335

[23] Pinelis, I. (1999). Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. *Advances in Stochastic Inequalities (Atlanta, GA, 1997), Contemp. Math.* **234**. Amer. Math. Soc., Providence, RI, pp. 149–168. MR1694770

[24] Pinelis, I. (2001). L'Hospital type rules for oscillation, with applications. *JIPAM. J. Inequal. Pure Appl. Math.* **2** (3) Article 33, 24 pp. (electronic). MR1876266

[25] Pinelis, I. (2002). L'Hospital type results for monotonicity, with applications. *JIPAM. J. Inequal. Pure Appl. Math.* **3** (1) Article 5, 5 pp. (electronic). MR1888920

[26] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York. MR0838963

[27] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Teor. Verojatnost. i Primenen.* **5** 331–335. MR0133849

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 53–61

DOI: 10.1214/074921706000000752

Oscillations of empirical distribution functions under dependence*

Wei Biao Wu[1]

University of Chicago

Abstract: We obtain an almost sure bound for oscillation rates of empirical distribution functions for stationary causal processes. For short-range dependent processes, the oscillation rate is shown to be optimal in the sense that it is as sharp as the one obtained under independence. The dependence conditions are expressed in terms of physical dependence measures which are directly related to the data-generating mechanism of the underlying processes and thus are easy to work with.

1. Introduction

Let $\varepsilon_0', \varepsilon_i, i \in \mathbb{Z}$, be independent and identically distributed (iid) random variables on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For $k \in \mathbb{Z}$ let

$$X_k = g(\ldots, \varepsilon_{k-1}, \varepsilon_k), \tag{1.1}$$

where g is a measurable function such that X_k is a well-defined random variable. Then $\{X_k\}_{k\in\mathbb{Z}}$ forms a stationary sequence. The framework (1.1) is very general. See [12, 15, 19, 23] among others. The process (X_k) is causal or non-anticipative in the sense that X_k does not depend on future innovations $\varepsilon_{k+1}, \varepsilon_{k+2}, \ldots$. Causality is a reasonable assumption in practice. The Wiener-Rosenblatt conjecture says that, for every stationary and ergodic process X_k, there exists a measurable function g and iid innovations ε_i such that the distributional equality $(X_k)_{k\in\mathbb{Z}} =_{\mathcal{D}} (g(\ldots, \varepsilon_{k-1}, \varepsilon_k))_{k\in\mathbb{Z}}$ holds; see [13, 20]. For an overview of the Wiener-Rosenblatt conjecture see [9].

Let F be the cumulative distribution function of X_k. Assume throughout the paper that F has a square integrable density f with square integrable derivative f'. In this paper we are interested in the oscillatory behavior of empirical distribution function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \le x}, \quad x \in \mathbb{R}. \tag{1.2}$$

In particular, we shall obtain an almost sure bound for the modulus of continuity for the function $G_n(x) = \sqrt{n}[F_n(x) - F(x)]$:

$$\Delta_n(b) = \sup_{|x-y|\le b} |G_n(x) - G_n(y)|, \tag{1.3}$$

[1]Department of Statistics, University of Chicago, 5734 S. University Ave, Chicago, IL 60637, e-mail: wbwu@galton.uchicago.edu

*The work is supported in part by NSF Grant DMS-0478704.

AMS 2000 subject classifications: primary 60G10, 60F05; secondary 60G42.

Keywords and phrases: almost sure convergence, dependence, empirical process, martingale.

where $b = b_n$ is a sequence of positive numbers satisfying

$$b_n \to 0 \text{ and } nb_n \to \infty. \tag{1.4}$$

Under the assumption that X_i are iid, there exists a huge literature on the asymptotic behavior of $\Delta_n(\cdot)$; see Chapter 14 in [14] and the references cited therein. A powerful tool to deal with the empirical distribution F_n is strong approximation [1]. In comparison, the behavior of $\Delta_n(\cdot)$ has been much less studied under dependence. In this paper we shall implement the new dependence measures proposed in [23] and obtain an almost sure bound for Δ_n. For a recent account of empirical processes for dependent random variables see the monograph edited by Dehling et al [4].

The rest of the paper is structured as follows. Main results on $\Delta_n(b)$ are presented in Section 2 and proved in Section 3. Section 4 contains comparisons with results obtained under independence. Some open problems are also posed in Section 4.

2. Main results

We first introduce some notation. For a random variable Z write $Z \in \mathcal{L}^p$ $(p > 0)$ if $\|Z\|_p := (\mathbb{E}|Z|^p)^{1/p} < \infty$. Write $\|\cdot\| = \|\cdot\|_2$. Let for $k \in \mathbb{Z}$,

$$\xi_k = (\ldots, \varepsilon_{k-1}, \varepsilon_k)$$

and for $k \geq 0$ let ξ_k^* be a coupled process of ξ_k with ε_0 replaced by ε_0', i.e.,

$$\xi_k^* = (\xi_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_k).$$

We shall write $X_k^* = g(\xi_k^*)$. Let $k \geq 1$ and define the conditional cumulative distribution function $F_k(\cdot|\xi_0)$ by

$$F_k(x|\xi_0) = \mathbb{P}(X_k \leq x|\xi_0). \tag{2.1}$$

Note that (X_i, ξ_i) is stationary. Then for almost every $\xi \in \cdots \times \mathbb{R} \times \mathbb{R}$ with respect to $\mathbb{P}$ we have

$$F_k(x|\xi_0 = \xi) = \mathbb{P}(X_k \leq x|\xi_0 = \xi) = \mathbb{P}(X_{k+i} \leq x|\xi_i = \xi) \tag{2.2}$$

for all $i \in \mathbb{Z}$. In other words, $F_k(\cdot|\xi)$ is the cumulative distribution function of the random variable $g(\xi, \varepsilon_{i+1}, \ldots, \varepsilon_{i+k})$. Assume that for all $k \geq 1$ and almost every $\xi \in \cdots \times \mathbb{R} \times \mathbb{R}$ with respect to $\mathbb{P}$ that $F_k(x|\xi_0 = \xi)$ has a derivative $f_k(x|\xi_0 = \xi)$, which is the conditional density of X_k at x given $\xi_0 = \xi$. By (2.2), for any $i \in \mathbb{Z}$, $f_k(x|\xi_i)$ is the conditional density of X_{k+i} at x given ξ_i. Let the conditional characteristic function

$$\varphi_k(\theta|\xi_0 = \xi) = \mathbb{E}(e^{\sqrt{-1}\theta X_k}|\xi_0 = \xi) = \int_{\mathbb{R}} e^{\sqrt{-1}\theta t} f_k(t|\xi_0 = \xi) dt, \tag{2.3}$$

where $\sqrt{-1}$ is the imaginary unit. Our dependence condition is expressed in terms of the $\mathcal{L}^2$ norm $\|\varphi_k(\theta|\xi_0) - \varphi_k(\theta|\xi_0^*)\|$.

Theorem 2.1. *Assume that $b_n \to 0$, $\log n = O(nb_n)$ and that there exists a positive constant c_0 for which*

$$\sup_x f_1(x|\xi_0) \leq c_0 \tag{2.4}$$

holds almost surely. Further assume that

$$\sum_{k=1}^{\infty}\left[\int_{\mathbb{R}}(1+\theta^2)\|\varphi_k(\theta|\xi_0)-\varphi_k(\theta|\xi_0^*)\|^2 d\theta\right]^{1/2}<\infty. \tag{2.5}$$

Let

$$\iota(n)=(\log n)^{1/2}\log\log n. \tag{2.6}$$

Then

$$\Delta_n(b_n)=O_{\text{a.s.}}(\sqrt{b_n\log n})+o_{\text{a.s.}}[b_n\iota(n)]. \tag{2.7}$$

Roughly speaking, (2.5) is a short-range dependence condition. Recall that $f_k(x|\xi_0)$ is the conditional (predictive) density of X_k at x given ξ_0 and $\varphi_k(\theta|\xi_0)$ is the conditional characteristic function. So $\varphi_k(\theta|\xi_0)-\varphi_k(\theta|\xi_0^*)$ measures the degree of dependence of $\varphi_k(\cdot|\xi_0)$ on ε_0. Hence the summand in (2.5) quantifies a distance between the conditional distributions $[X_k|\xi_0]$ and $[X_k^*|\xi_0^*]$ and (2.5) means that the cumulative contribution of ε_0 in predicting future values is finite.

For the two terms in the bound (2.7), the first one $O_{\text{a.s.}}(\sqrt{b_n\log n})$ has the same order of magnitude as the one that one can obtain under independence. See Chapter 14 in [14] and Section 4. The second term $o_{\text{a.s.}}[b_n\iota(n)]$ is due to the dependence of the process (X_k). Clearly, if $b_n^{1/2}(\log\log n)=o(1)$, then the first term dominates the bound in (2.7). The latter condition holds under mild conditions on b_n, for example, if $b_n=O(n^{-\eta})$ for some $\eta>0$.

Let $k\geq 1$. Observe that $\varphi_k(\theta|\xi_0)=\mathbb{E}[\varphi_1(\theta|\xi_{k-1})|\xi_0]$ and

$$\mathbb{E}[\varphi_1(\theta|\xi_{k-1})|\xi_{-1}]=\mathbb{E}[\varphi_1(\theta|\xi_{k-1}^*)|\xi_{-1}]=\mathbb{E}[\varphi_1(\theta|\xi_{k-1}^*)|\xi_0]. \tag{2.8}$$

To see (2.8), write $h(\xi_{k-1})=\varphi_1(\theta|\xi_{k-1})$. Note that $\varepsilon_i,\varepsilon_0'$, $i\in\mathbb{Z}$, are iid and $k-1\geq 0$. Then we have $\mathbb{E}[h(\xi_{k-1})|\xi_{-1}]=\mathbb{E}[h(\xi_{k-1}^*)|\xi_{-1}]$ since ξ_{k-1}^* is a coupled version of ξ_{k-1} with ε_0 replaced by ε_0'. On the other hand, we have $\mathbb{E}[h(\xi_{k-1}^*)|\xi_{-1}]=\mathbb{E}[h(\xi_{k-1}^*)|\xi_0]$ since ε_0 is independent of ξ_{k-1}^*. So (2.8) follows. Define the projection operator $\mathcal{P}_k$ by

$$\mathcal{P}_kZ=\mathbb{E}(Z|\xi_k)-\mathbb{E}(Z|\xi_{k-1}),\quad Z\in\mathcal{L}^1.$$

By the Jensen and the triangle inequalities,

$$\begin{aligned}\|\varphi_k(\theta|\xi_0)-\varphi_k(\theta|\xi_0^*)\| &\leq \|\varphi_k(\theta|\xi_0)-\mathbb{E}[\varphi_k(\theta|\xi_0)|\xi_{-1}]\|\\ &\quad+\|\mathbb{E}[\varphi_k(\theta|\xi_0)|\xi_{-1}]-\varphi_k(\theta|\xi_0^*)\|\\ &=2\|\mathcal{P}_0\varphi_1(\theta|\xi_{k-1})\|\\ &\leq 2\|\varphi_1(\theta|\xi_{k-1})-\varphi_1(\theta|\xi_{k-1}^*)\|.\end{aligned}$$

Then a sufficient condition for (2.5) is

$$\sum_{k=0}^{\infty}\left[\int_{\mathbb{R}}(1+\theta^2)\|\varphi_1(\theta|\xi_k)-\varphi_1(\theta|\xi_k^*)\|^2 d\theta\right]^{1/2}<\infty. \tag{2.9}$$

In certain applications it is easier to work with (2.9). In Theorem 2.2 below we show that (2.9) holds for processes (X_k) with the structure

$$X_k=\varepsilon_k+Y_{k-1}, \tag{2.10}$$

where Y_{k-1} is $\xi_{k-1} = (\ldots, \varepsilon_{k-2}, \varepsilon_{k-1})$ measurable. It is also a large class. The widely used linear process $X_k = \sum_{i=0}^{\infty} a_i \varepsilon_{k-i}$ is of the form (2.10). Nonlinear processes of the form

$$X_k = m(X_{k-1}) + \varepsilon_k \tag{2.11}$$

also fall within the framework of (2.10) if (2.11) has a stationary solution. A prominent example of (2.11) is the threshold autoregressive model [19]

$$X_k = a \max(X_{k-1}, 0) + b \min(X_{k-1}, 0) + \varepsilon_k,$$

where a and b are real parameters. For processes of the form (2.10), condition (2.5) can be simplified. Let φ be the characteristic function of ε_1.

Theorem 2.2. *Let $0 < \alpha \le 2$. Assume* (2.10),

$$\int_{\mathbb{R}} |\varphi(t)|^2 (1 + t^2) |t|^\alpha dt < \infty \tag{2.12}$$

and

$$\sum_{k=0}^{\infty} \|X_k - X_k^*\|_\alpha^{\alpha/2} < \infty \tag{2.13}$$

Then (2.5) *is satisfied.*

Condition (2.12) does not appear to be overly restrictive. It is satisfied if $|\varphi(t)| = O(|t|^{-\eta})$ as $|t| \to \infty$, where $\eta > (3 + \alpha)/2$. It is also satisfied for symmetric-α-stable distributions, an important class of distributions with heavy tails. Let ε_k have standard symmetric α stable distribution with index $0 < \iota \le 2$. Then its characteristic function $\varphi(t) = \exp(-|t|^\iota)$ and (2.12) trivially holds.

We now discuss Condition (2.13). Recall $X_k^* = g(\xi_k^*)$. Note that X_k^* and X_k are identically distributed and X_k^* is a coupled version of X_k with ε_0 replaced by ε_0'. If we view (1.1) as a physical system with $\xi_k = (\ldots, \varepsilon_{k-1}, \varepsilon_k)$ being the input, g being a filter or tranform and X_i being the output, then the quantity $\|X_k - X_k^*\|_\alpha$ measures the degree of dependence of $g(\ldots, \varepsilon_{k-1}, \varepsilon_k)$ on ε_0. In [23] it is called the *physical* or *functional dependence measure.* With this input/output viewpoint, the condition (2.13) means that the cumulative impact of ε_0 is finite, and hence suggesting short-range dependence. In many applications it is easily verifiable since it is directly related to the data-generating mechanism and since the calculation of $\|X_k - X_k^*\|_\alpha$ is generally easy [23]. In the special case of linear process $X_k = \sum_{j=0}^{\infty} a_j \varepsilon_{k-j}$ with $\varepsilon_k \in \mathcal{L}^\alpha$ and $\alpha = 2$, then $\|X_k - X_k^*\|_\alpha = |a_k| \|\varepsilon_0 - \varepsilon_0'\|_\alpha$ and (2.13) is reduced to $\sum_{k=0}^{\infty} |a_k| < \infty$, which is a classical condition for linear processes to be short-range dependent. It is well-known that, if the latter condition is barely violated, then one enters the territory of long-range dependence. Consequently both the normalization and the bound in (2.7) will be different; see [8, 21].

For the nonlinear time series (2.11), assume that $\varepsilon_k \in \mathcal{L}^\alpha$ and $\rho = \sup_x |m'(x)| < 1$. Then (2.11) has a stationary distribution and $\|X_k - X_k^*\|_\alpha = O(\rho^k)$ (see [24]). Hence (2.13) holds.

3. Proofs

Lemma 3.1. *Let H be a differential function on $\mathbb{R}$. Then for any $\lambda > 0$,*

$$\sup_{x \in \mathbb{R}} H^2(x) \le \lambda \int_{\mathbb{R}} H^2(x) dx + \lambda^{-1} \int_{\mathbb{R}} [H'(x)]^2 dx. \tag{3.1}$$

Proof. By the arithmetic mean geometric inequality inequality, for all $x, y \in \mathbb{R}$,

$$\begin{aligned} H^2(x) &\le H^2(y) + \left| \int_x^y 2H(t)H'(t)dt \right| \\ &\le H^2(y) + \lambda \int_{\mathbb{R}} H^2(x)dx + \lambda^{-1} \int_{\mathbb{R}} [H'(x)]^2 dx. \end{aligned} \tag{3.2}$$

If $\inf_{y\in\mathbb{R}} |H(y)| > 0$, then $\int_{\mathbb{R}} H^2(x)dx = \infty$ and (3.1) holds. If on the other hand $\inf_{y\in\mathbb{R}} |H(y)| = 0$, let $(y_n)_{n\in\mathbb{N}}$ be a sequence such that $H(y_n) \to 0$. So (3.2) entails (3.1). □

Lemma 3.1 is a special case of the Kolmogorov-type inequalities [18]. The result in the latter paper asserts that $\sup_{x\in\mathbb{R}} H^4(x) \le \int_{\mathbb{R}} H^2(x)dx \times \int_{\mathbb{R}} H'(x)^2 dx$. For the sake of completeness, we decide to state Lemma 3.1 with a simple proof here.

Recall that $F_k(\cdot|\xi_0)$ is the conditional distribution function of X_k given ξ_0 (cf (2.1) and (2.2)). Introduce the conditional empirical distribution function

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\mathbf{1}_{X_i\le x}|\xi_{i-1}) = \frac{1}{n}\sum_{i=1}^n F_1(x|\xi_{i-1}).$$

Write

$$G_n(x) = G_n^\diamond(x) + G_n^*(x), \tag{3.3}$$

where

$$G_n^\diamond(x) = \sqrt{n}[F_n(x) - F_n^*(x)] \text{ and } G_n^*(x) = \sqrt{n}[F_n^*(x) - F(x)]. \tag{3.4}$$

Then

$$\sqrt{n}G_n^\diamond(x) = \sum_{i=1}^n d_i(x) \tag{3.5}$$

is a martingale with respect to the filtration $\sigma(\xi_n)$ and the increments $d_i(x) = \mathbf{1}_{X_i\le x} - \mathbb{E}(\mathbf{1}_{X_i\le x}|\xi_{i-1})$ are stationary, ergodic and bounded. On the other hand, if the conditional density $f_1(\cdot|\xi_i)$ exists, then G_n^* is differentiable. The latter differentiability property is quite useful.

Lemma 3.2. *Recall* (2.6) *for* $\iota(n)$. *Let* $g_n^*(x) = dG_n^*(x)/dx$. *Assume* (2.5). *Then*

$$\sup_x |g_n^*(x)| = o_{\text{a.s.}}[\iota(n)]. \tag{3.6}$$

Proof. Let $k \ge 1$. Recall that $f_1(x|\xi_{k-1})$ is the one-step-ahead conditional density of X_k at x given ξ_{k-1}. By (2.3), we have

$$\mathcal{P}_0\varphi_1(t|\xi_{k-1}) = \int_{\mathbb{R}} e^{\sqrt{-1}xt}\mathcal{P}_0 f_1(x|\xi_{k-1})dt.$$

By Parseval's identity, we have

$$\int_{\mathbb{R}} |\mathcal{P}_0\varphi_1(t|\xi_{k-1})|^2 dt = \frac{1}{2\pi}\int_{\mathbb{R}} |\mathcal{P}_0 f_1(x|\xi_{k-1})|^2 dx$$

and

$$\int_{\mathbb{R}} |\mathcal{P}_0\varphi_1(t|\xi_{k-1})|^2 t^2 dt = \frac{1}{2\pi}\int_{\mathbb{R}} |\mathcal{P}_0 f_1'(x|\xi_{k-1})|^2 dx.$$

Let

$$\alpha_k = \int_{\mathbb{R}} \|\mathcal{P}_0 f_1(x|\xi_{k-1})\|^2 dx$$

and

$$\beta_k = \int_{\mathbb{R}} \|\mathcal{P}_0 f_1'(x|\xi_{k-1})\|^2 dx.$$

By (2.8) and Jensen's inequality, $\|\mathcal{P}_0 \varphi_1(\theta|\xi_{k-1})\| \le \|\varphi_k(\theta|\xi_0) - \varphi_k(\theta|\xi_0^*)\|$. So (2.5) implies that

$$\sum_{k=1}^{\infty} \sqrt{\alpha_k + \beta_k} < \infty. \tag{3.7}$$

Let $\Lambda = \sum_{k=1}^{\infty} \sqrt{\alpha_k}$ and for $k \ge 0$

$$H_k(x) = \sum_{i=1}^{k} [f_1(x|\xi_{i-1}) - f(x)].$$

Then $H_k(x) = \sqrt{k} g_k(x)$. Note that for fixed $l \in \mathbb{N}$, $\mathcal{P}_{i-l} f_1(x|\xi_{i-1})$, $i = 1, 2, \ldots$, are stationary martingale differences with respect to the filtration $\sigma(\xi_{i-l})$. By the Cauchy-Schwarz inequality and Doob's maximal inequality, since $H_k(x) = \sum_{l=1}^{\infty} \sum_{i=1}^{k} \mathcal{P}_{i-l} f_1(x|\xi_{i-1})$,

$$\begin{aligned}
\mathbb{E} \int_{\mathbb{R}} \max_{k \le n} H_k^2(x) dx &\le \mathbb{E} \int_{\mathbb{R}} \sum_{l=1}^{\infty} \frac{\max_{k \le n} |\sum_{i=1}^{n} \mathcal{P}_{i-l} f_1(x|\xi_{i-1})|^2}{\sqrt{\alpha_l}} \Lambda dx \\
&\le \int_{\mathbb{R}} \sum_{l=1}^{\infty} \frac{4n \|\mathcal{P}_0 f_1(x|\xi_{l-1})\|^2}{\sqrt{\alpha_l}} \Lambda dx = 4n\Lambda^2 = O(n).
\end{aligned}$$

Similarly, since $\sum_{k=1}^{\infty} \sqrt{\beta_k} < \infty$,

$$\mathbb{E} \int_{\mathbb{R}} \max_{k \le n} |H_k'(x)|^2 dx = O(n).$$

By Lemma 3.1 with $\lambda = 1$, we have

$$\begin{aligned}
&\sum_{d=1}^{\infty} \frac{\mathbb{E}[\max_{k \le 2^d} \sup_{x \in \mathbb{R}} |H_k(x)|^2]}{2^d \iota^2(2^d)} \\
&\quad \le \sum_{d=1}^{\infty} \frac{\mathbb{E}[\max_{k \le 2^d} \int_{\mathbb{R}} |H_k(x)|^2 + |H_k'(x)|^2 dx]}{2^d \iota^2(2^d)} \\
&\quad \le \sum_{d=1}^{\infty} \frac{\mathbb{E} \int_{\mathbb{R}} \max_{k \le 2^d} |H_k(x)|^2 + \max_{k \le 2^d} |H_k'(x)|^2 dx}{2^d \iota^2(2^d)} \\
&\quad = \sum_{d=1}^{\infty} \frac{O(2^d)}{2^d \iota^2(2^d)} < \infty.
\end{aligned}$$

By the Borel-Cantelli lemma, $\sup_x \max_{k \le 2^d} |H_k(x)| = o_{\text{a.s.}}[2^{d/2} \iota(2^d)]$ as $d \to \infty$. For any $n \ge 2$ there is a $d \in \mathbb{N}$ such that $2^{d-1} < n \le 2^d$. Note that $\max_{k \le n} |H_k(x)| \le \max_{k \le 2^d} |H_k(x)|$ and $\iota(n)$ is slowly varying. So (3.6) follows. $\square$

Lemma 3.3. *Assume* $\log n = O(nb_n)$ *and* $X_k \in \mathcal{L}^\alpha$ *for some* $\alpha > 0$. *Then for any* $\tau > 2$, *there exists* $C = C_\tau > 0$ *such that*

$$\mathbb{P}\left[\sup_{|x-y|\le b_n} |G_n^\diamond(x) - G_n^\diamond(y)| > C\sqrt{b_n \log n}\right] = O(n^{-\tau}). \tag{3.8}$$

Proof. We can adopt the argument of Lemma 5 in [22]. Let $x_0 = n^{(3+\tau)/\alpha}$. Then for any $C > 0$, by Markov's inequality,

$$\begin{aligned}
&\mathbb{P}[n\{F_n(-x_0) + 1 - F_n(x_0)\} > c\sqrt{b_n \log n}] \\
&\quad \le \frac{n\mathbb{E}\{F_n(-x_0) + 1 - F_n(x_0)\}}{C\sqrt{b_n \log n}} \\
&\quad \le \frac{n x_0^{-\alpha}\mathbb{E}(|X_0|^\alpha)}{C\sqrt{b_n \log n}} = O(n^{-\tau}).
\end{aligned} \tag{3.9}$$

It is easily seen that the preceding inequality also holds if F_n is replaced by F_n^*.

Recall (3.5) for $G_n^\diamond(x)$ and $d_i(x) = \mathbf{1}_{X_i \le x} - \mathbb{E}(\mathbf{1}_{X_i \le x}|\xi_{i-1})$. Let $x \le y \le x + b_n$. By (2.4),

$$\sum_{i=1}^n \mathbb{E}[(d_i(y) - d_i(x))^2|\xi_{i-1}] \le \sum_{i=1}^n \mathbb{E}(\mathbf{1}_{x \le X_i \le y}|\xi_{i-1}) \le nb_n c_0.$$

By Freedman's inequality in [6], if $|x - y| \le b_n$, we have

$$\mathbb{P}\left[\sqrt{n}|G_n^\diamond(x) - G_n^\diamond(y)| > C\sqrt{nb_n \log n}\right] \le 2\exp\left[\frac{-C^2 nb_n \log n}{C\sqrt{nb_n \log n} + nb_n c_0}\right].$$

Let $\Theta_n = \{-x_0 + k/n^3 : k = 0, \ldots, \lfloor 2x_0 n^3 \rfloor\}$. Since $\log n = O(nb_n)$, it is easily seen that there exists a $C = C_\tau$ such that

$$\begin{aligned}
&\mathbb{P}\left[\sup_{x,y\in\Theta_n, |x-y|\le b_n} \sqrt{n}|G_n^\diamond(x) - G_n^\diamond(y)| > C\sqrt{nb_n \log n}\right] \\
&\quad = O(x_0^2 n^2)\exp\left[\frac{-C^2 nb_n \log n}{C\sqrt{nb_n \log n} + nb_n c_0}\right] = O(n^{-\tau}).
\end{aligned} \tag{3.10}$$

For every $x \in [-x_0, x_0]$, there exists a $\theta \in \Theta_n$ such that $\theta < x \le \theta + 1/n^3$. So (3.10) implies that

$$\begin{aligned}
&\mathbb{P}\left[\sup_{|x|\le x_0, |y|\le x_0, |x-y|\le b_n} \sqrt{n}|G_n^\diamond(x) - G_n^\diamond(y)| > (C+1)\sqrt{nb_n \log n}\right] \\
&\quad = O(n^{-\tau})
\end{aligned} \tag{3.11}$$

in view of the monotonicity of $F_n(\cdot)$ and the fact that, if $\theta \le \phi \le \theta + 1/n^3$,

$$\sum_{i=1}^n \mathbb{E}(\mathbf{1}_{\theta \le X_i \le \phi}|\xi_{i-1}) \le |\phi - \theta| nc_0 = c_0/n^2 = o(\sqrt{nb_n \log n}).$$

Combining (3.9) and (3.11), we have (3.8). □

Proof of Theorem 2.1. By (3.3), it easily follows from Lemmas 3.2 and 3.3. □

Proof of Theorem 2.2. By Lemma 3.1 and Parseval's identity, (2.12) implies

$$\sup_x f^2(x) \le \int_{\mathbb{R}} \{f^2(x) + [f'(x)]^2\} dx = \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi(t)|^2 (1+t^2) dt < \infty.$$

Since ε_k and Y_{k-1} are independent,

$$\mathbb{E}[e^{\sqrt{-1}tX_k} | \xi_{k-1}] = e^{\sqrt{-1}tY_{k-1}} \varphi(t).$$

Note that

$$\begin{aligned}
&\|e^{\sqrt{-1}tY_{k-1}}\varphi(t) - e^{\sqrt{-1}tY^*_{k-1}}\varphi(t)\|^2 \\
&\quad = |\varphi(t)|^2 \|e^{\sqrt{-1}tY_{k-1}} - e^{\sqrt{-1}tY^*_{k-1}}\|^2 \\
&\quad \le 4|\varphi(t)|^2 \|\min(1, |tY_{k-1} - tY^*_{k-1}|)\|^2 \\
&\quad \le 4|\varphi(t)|^2 \mathbb{E}(|tY_{k-1} - tY^*_{k-1}|^\alpha).
\end{aligned}$$

Hence (2.5) follows from (2.13). □

4. Conclusion and open problems

Let X_i be iid standard uniform random variables. Stute [16] obtained the following interesting result. Assume that $b_n \to 0$ is a sequence of positive numbers such that

$$\log n = o(nb_n) \text{ and } \log\log n = o(\log b_n^{-1}). \tag{4.1}$$

Then the convergence result holds:

$$\lim_{n\to\infty} \frac{\Delta_n(b_n)}{\sqrt{b_n \log b_n^{-1}}} = \sqrt{2} \text{ almost surely.} \tag{4.2}$$

If there exists $\eta > 0$ such that $b_n + (nb_n)^{-1} = O(n^{-\eta})$, then the bound in (2.7) becomes $\sqrt{b_n \log n}$, which has the same order of magnitude as $\sqrt{b_n \log b_n^{-1}}$, the bound asserted by (4.2). It is unclear whether there exists an almost sure limit for $\Delta_n(b_n)/\sqrt{b_n \log b_n^{-1}}$ if the dependence among observations is allowed. Mason et al [11] considered almost sure limit for $\Delta_n(b_n)/\sqrt{b_n \log b_n^{-1}}$ when (4.1) is violated. Deheuvels and Mason [3] (see also [2]) proved functional laws of the iterated logarithm for the increments of empirical processes. Local empirical processes in high dimensions have been studied in [5, 7, 17]. It is an open problem whether similar results hold for stationary causal processes. We expect that our decomposition (3.4) will be useful in establishing comparable results.

Acknowledgements

I am grateful for the very helpful suggestions from the reviewer and the Editor.

References

[1] Csörgő, M. and Révész, P., (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

[2] Deheuvels P. (1997). Strong laws for local quantile processes. *Ann. Probab.* **25** 2007–2054.

[3] Deheuvels P. and Mason, D. M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.* 20 1248–1287.

[4] Dehling, H., T. Mikosch and M. Sørensen (eds), (2002). *Empirical Process Techniques for Dependent Data*. Birkhäuser, Boston.

[5] Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** 1–37.

[6] Freedman, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118.

[7] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 907–921.

[8] Ho, H. C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024.

[9] Kallianpur, G. (1981). Some ramifications of Wieners ideas on nonlinear prediction. In *Norbert Wiener, Collected Works with Commentaries.* MIT Press, Cambridge, MA, pp. 402–424.

[10] Mason, D. M. (2004). A uniform functional law of the logarithm for the local empirical process *Ann. Probab.* **32** 1391–1418

[11] Mason, D. M., Shorack G. R. and J. A. Wellner (1983). Strong limit theorems for oscillation moduli of the uniform empirical process. *Z. Wahrsch. verw. Gebiete* **65** 83–97.

[12] Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis.* Academic Press, London.

[13] Rosenblatt, M. (1959). Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.* **8** 665–681.

[14] Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* John Wiley & Sons, New York.

[15] Stine, R. A. (1997) Nonlinear time series. In *Encyclopedia of Statistical Sciences*, Updated vol. 1, Edited by S. Kotz, C. B. Read and D. L. Banks. Wiley, pp. 430–437

[16] Stute, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

[17] Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case. *Ann. Probab.* **12** 361–379.

[18] Taĭkov, L. V. (1968). Inequalities of Kolmogorov type and best formulas for numerical differentiation (in Russian). *Mat. Zametki* **4** 233–238.

[19] Tong, H. (1990). *Nonlinear Time Series. A Dynamical System Approach.* Oxford University Press.

[20] Wiener, N. (1958). *Nonlinear Problems in Random Theory.* MIT Press, Cambridge, MA.

[21] Wu, W. B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9** 809–831

[22] Wu, W. B. (2005a). On the Bahadur representation of sample quantiles for stationary sequences. *Ann. Statist.* **33** 1934–1963

[23] Wu, W. B. (2005b). Nonlinear system theory: another look at dependence. *Proc. Natl Acad. Sci. USA* **102** 14150–14154.

[24] Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 62–76

DOI: 10.1214/074921706000000761

Karhunen-Loève expansions of mean-centered Wiener processes

Paul Deheuvels[1]

L.S.T.A., Université Pierre et Marie Curie (Paris 6)

Abstract: For $\gamma > -\frac{1}{2}$, we provide the Karhunen-Loève expansion of the weighted mean-centered Wiener process, defined by

$$W_\gamma(t) = \frac{1}{\sqrt{1+2\gamma}} \Big\{ W\Big(t^{1+2\gamma}\Big) - \int_0^1 W\Big(u^{1+2\gamma}\Big) du \Big\},$$

for $t \in (0,1]$. We show that the orthogonal functions in these expansions have simple expressions in term of Bessel functions. Moreover, we obtain that the $L^2[0,1]$ norm of W_γ is identical in distribution with the $L^2[0,1]$ norm of the weighted Brownian bridge $t^\gamma B(t)$.

1. Introduction and results

1.1. KL expansions of mean-centered Wiener processes

Let $\{W(t) : t \geq 0\}$ denote a standard Wiener process, and let

$$\{B(t) : t \geq 0\} \stackrel{\text{law}}{=} \{W(t) - tW(1) : t \geq 0\}, \tag{1.1}$$

denote a standard Brownian bridge, where "$\stackrel{\text{law}}{=}$" denotes equality in distribution. As a motivation to the present paper, we start by establishing, in Theorem 1.1 below, the Karhunen-Loève [KL] expansion of the *mean-centered Wiener process*

$$W_0(t) := W(t) - \int_0^1 W(u)du, \quad t \in [0,1]. \tag{1.2}$$

We recall the following basic facts about KL expansions (or representations). Let $d \geq 1$ be a positive integer. It is well-known (see, e.g., [1, 3, 9] and Ch.1 in [8]) that a centered Gaussian process $\{\zeta(\mathbf{t}) : \mathbf{t} \in [0,1]^d\}$, with covariance function $K_\zeta(\mathbf{s},\mathbf{t}) = E(\zeta(\mathbf{s})\zeta(\mathbf{t}))$ in $L^2\big([0,1]^d \times [0,1]^d\big)$, admits the (convergent in expected mean squares) KL expansion

$$\zeta(t) \stackrel{\text{law}}{=} \sum_{k=1}^\infty \omega_k \sqrt{\lambda_k}\, e_k(\mathbf{t}), \tag{1.3}$$

where $\{\omega_k : k \geq 1\}$ is a sequence of independent and identically distributed [iid] normal $N(0,1)$ random variables, and the $\{e_k : k \geq 1\}$ form an orthonormal sequence

[1]L.S.T.A., Université Paris 6, 7 avenue du Château, F-92340 Bourg-la-Reine, France, e-mail: pd@ccr.jussieu.fr

AMS 2000 subject classifications: primary 62G10; secondary 60F15, 60G15, 60H07, 62G30.

Keywords and phrases: Gaussian processes, Karhunen-Loève expansions, Wiener process, Brownian bridge, quadratic functionals.

in $L^2\big([0,1]^d\big)$, fulfilling

$$(1.4)\qquad \int_{[0,1]^d} e_k(\mathbf{t})e_\ell(\mathbf{t})d\mathbf{t} = \begin{cases} 1 & \text{if} \quad k=\ell, \\ 0 & \text{if} \quad k \neq \ell, \end{cases}$$

with $d\mathbf{t}$ denoting Lebesgue's measure. The $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ in (1.3) are the eigenvalues of the Fredholm operator $h \in L^2\big([0,1]^d\big) \to T_\zeta h \in L^2\big([0,1]^d\big)$, defined via

$$(1.5)\qquad T_\zeta h(\mathbf{t}) = \int_{[0,1]^d} K_\zeta(\mathbf{s},\mathbf{t})h(\mathbf{s})d\mathbf{s} \quad \text{for} \quad t \in [0,1]^d.$$

We have, namely, $T_\zeta e_k = \lambda_k e_k$ for each $k \geq 1$. A natural consequence of the KL representation (1.3) is the distributional identity

$$(1.6)\qquad \int_{[0,1]^d} \zeta^2(\mathbf{t})d\mathbf{t} \overset{\text{law}}{=} \sum_{k=1}^{\infty} \lambda_k \omega_k^2.$$

Given these preliminaries, we may now state our first theorem.

Theorem 1.1. *The KL expansion of* $\{W_0(t) : t \in [0,1]\}$ *is given by*

$$(1.7)\qquad W_0(t) = W(t) - \int_0^1 W(u)du \overset{\text{law}}{=} \sum_{k=1}^{\infty} \omega_k \frac{\sqrt{2}\ \cos(k\pi t)}{k\pi} \quad \textit{for} \quad t \in [0,1].$$

The proof of Theorem 1.1 is postponed until Section 2.1. In Remark 1.1 below, we will discuss some implications of this theorem in connection with well-known properties of the Brownian bridge, as defined in (1.1).

Remark 1.1. (1) We will provide, in the forthcoming Theorem 1.3 (in Section 1.4), a weighted version of the KL expansion of $W_0(\cdot)$, as given in (1.7). These KL expansions are new, up to our best knowledge.

(2) The well-known (see, e.g., [2] and [7]) KL expansion of the Brownian bridge (1.1) is very similar to (1.7), and given by

$$(1.8)\qquad B(t) = W(t) - tW(1) \overset{\text{law}}{=} \sum_{k=1}^{\infty} \omega_k \frac{\sqrt{2}\ \sin(k\pi t)}{k\pi} \quad \text{for} \quad 0 \leq t \leq 1.$$

As a direct consequence of (1.5)-(1.7)-(1.8), we obtain the distributional identities

$$(1.9)\qquad \int_0^1 \Big\{W(t) - \int_0^1 W(u)du\Big\}^2 dt \overset{\text{law}}{=} \int_0^1 B^2(t)dt \overset{\text{law}}{=} \sum_{k=1}^{\infty} \frac{\omega_k^2}{k^2\pi^2}.$$

The first identity in (1.9) is given, p.517 of [5], as a consequence of Fubini-Wiener arguments, in the spirit of the results of [6]. The second identity in (1.9) follows directly from (1.6), and is well-known (see, e.g., [4]).

The remainder of our paper is organized as follows. In Section 1.2 below, we mention, as a straightforward, but nevertheless useful, observation, that the knowledge of the distribution of the L^2 norm of a Gaussian process characterizes the eigenvalues of its KL expansion. In Section 1.3, we extend our study to Gaussian processes related to the Wiener sheet in $[0,1]^d$. The results of [5] will be instrumental in this case to establish a series of distributional identities of L^2 norms, between various Gaussian processes of interest. In Section 1.4, we provide the KL expansion of the *weighted mean-centered Wiener process* in dimension $d = 1$. The results of this section, are, as one could expect, closely related to [4], where related KL decompositions of weighted Wiener processes and Brownian bridges are established. In particular, in these results, we will make an instrumental use of Bessel functions. In the case of a general $d \geq 1$, we provide KL expansions for a general version of the mean-centered Wiener sheet. Finally, in Section 2, we will complete the proofs of the results given in Section 1.

1.2. The L^2 norm and KL eigenvalues

A question raised by (1.9) is as follows. Let $\zeta_1(\cdot)$ and $\zeta_2(\cdot)$ be two centered Gaussian processes on $[0,1]^d$, with covariance functions in $L^2\big([0,1]^d \times [0,1]^d\big)$, and KL expansions given by, for $\mathbf{t} \in [0,1]^d$,

$$\zeta_j(\mathbf{t}) \overset{\text{law}}{=} \sum_{k=1}^{\infty} \omega_k \sqrt{\lambda_{k,j}}\, e_{k,j}(\mathbf{t}) \quad \text{for} \quad j = 1, 2. \tag{1.10}$$

Making use of the notation of [5], we write $\zeta_1 \overset{\text{Quad}}{=} \zeta_2$, when the $L^2\big([0,1]^d\big)$ norms of these two processes are identical in distribution. We may therefore write the equivalence

$$\zeta_1 \overset{\text{Quad}}{=} \zeta_2 \quad \Leftrightarrow \quad \int_{[0,1]^d} \zeta_1^2(\mathbf{t})d\mathbf{t} \overset{\text{law}}{=} \int_{[0,1]^d} \zeta_2^2(\mathbf{t})d\mathbf{t}. \tag{1.11}$$

What can be said of the eigenvalue sequences $\{\lambda_{k,j} : k \geq 1\}$, $j = 1, 2$ when (1.11) is fulfilled? The answer to this question is given below.

Theorem 1.2. *The condition $\zeta_1 \overset{\text{Quad}}{=} \zeta_2$ is equivalent to the identity*

$$\lambda_{k,1} = \lambda_{k,2} \quad \textit{for all} \quad k \geq 1. \tag{1.12}$$

Proof. The fact that (1.12) implies $\zeta_1 \overset{\text{Quad}}{=} \zeta_2$ is trivial, via (1.6). A simple proof of the converse implication follows from the expression of the moment-generating functions [mgf] (see, e.g., pp. 60–61 in [4]), for $j = 1, 2$,

$$E\Big(\exp\Big\{z\int_{[0,1]^d} \zeta_j^2(\mathbf{t})d\mathbf{t}\Big\}\Big) = \prod_{k=1}^{\infty}\Big\{\frac{1}{1 - 2z\lambda_{k,j}}\Big\}^{1/2} \quad \text{for} \quad \mathrm{Re}(z) < \frac{1}{2\lambda_{1,j}}. \tag{1.13}$$

A variety of methods are available to show that the equality of the mgf's in (1.13), for $j = 1, 2$, implies (1.12). A simple argument, suggested by D. M. Mason (personal communication), uses the fact that the large deviations of the $L^2[0,1]$ norms of the $\zeta_j(\cdot)$ are governed by $\lambda_{1,j}$, $j = 1, 2$ (see, e.g., Lemma 1.1 in [4]). This allows to show that $\lambda_{1,1} = \lambda_{1,2}$. The proof is completed by a straightforward induction, after substracting from $\zeta_j(\cdot)$, $j = 1, 2$. the first components of their respective KL expansions. □

1.3. Multivariate Wiener sheets and Brownian Bridges

Inspired by Theorems 1.1-1.2 and Remark 1.1, we will devote the remainder of our paper to derive of a series of new KL expansions of Gaussian processes, in the spirit of (1.7). To give an additional motivation to the forthcoming results, we introduce the following notation and definitions. For any integer $d \geq 1$, the *d-variate Wiener process* (or *Brownian sheet*) is a centered Gaussian process $\mathbf{W}(\mathbf{t})$, defining a continuous function of $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{R}_+^d$, with covariance function given by, for $\mathbf{s} = (s_1, \ldots, s_d) \in \mathbb{R}_+^d$ and $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{R}_+^d$,

$$E\big(\mathbf{W}(\mathbf{s})\mathbf{W}(\mathbf{t})\big) = \prod_{i=1}^{d} \big(s_i \wedge t_i\big). \tag{1.14}$$

For each $i = 1, \ldots, d$, denote, respectively, by Δ_i, Σ_i and Θ_i, the operators which map a $L^1\big([0,1]^d\big)$ function $f\big(\mathbf{t}\big)$ of $\mathbf{t} = (t_1, \ldots, t_d) \in [0,1]^d$, into, respectively,

$$\Delta_i f\big(\mathbf{t}\big) = f\big(\mathbf{t}\big) - t_i f\big(t_1, \ldots, t_{i-1}, 1, t_{i+1}, \ldots, t_d\big), \tag{1.15}$$

$$\Sigma_i f\big(\mathbf{t}\big) = f\big(\mathbf{t}\big) - \int_0^1 f\big(t_1, \ldots, t_i, \ldots, t_d\big) dt_i, \tag{1.16}$$

$$\Theta_i f\big(\mathbf{t}\big) = f\big(\mathbf{t}\big) - f\big(t_1, \ldots, t_{i-1}, 1, t_{i+1}, \ldots, t_d\big). \tag{1.17}$$

The *d-variate tied-down Brownian bridge* is a centered Gaussian process $\mathbf{B}_*\big(\mathbf{t}\big)$, which is a continuous function of $\mathbf{t} = (t_1, \ldots, t_d) \in [0,1]^d$, defined, in terms of the d-variate Wiener process $\mathbf{W}(\mathbf{t})$, via the distributional identity

$$\mathbf{B}_*\big(\mathbf{t}\big) \stackrel{\text{law}}{=} \Delta_1 \circ \ldots \circ \Delta_d \mathbf{W}\big(\mathbf{t}\big), \tag{1.18}$$

where "$\circ$" denotes the composition of applications. We define likewise the *d-variate mean-centered Wiener sheet* by setting

$$\mathbf{W}_M\big(\mathbf{t}\big) \stackrel{\text{law}}{=} \Sigma_1 \circ \ldots \circ \Sigma_d \mathbf{W}\big(\mathbf{t}\big). \tag{1.19}$$

In (1.18), we use the notation $\mathbf{B}_*$ to distinguish the tied-down Brownian bridge $\mathbf{B}_*$ from the usual *d-variate standard Brownian bridge* $\mathbf{B}$, the latter being classically defined via the distributional identity

$$\mathbf{B}\big(\mathbf{t}\big) \stackrel{\text{law}}{=} \mathbf{W}\big(\mathbf{t}\big) - \Big\{\prod_{i=1}^{d} t_i\Big\}\mathbf{W}\big(1, \ldots, 1\big). \tag{1.20}$$

When $d = 1$ the processes $\mathbf{W}$ and $\mathbf{B}_*, \mathbf{B}$, reduce, respectively, to the standard Wiener process W, and to the standard Brownian bridge B, via the distributional identities

$$\mathbf{W} \stackrel{\text{law}}{=} W \quad \text{and} \quad \mathbf{B}_* \stackrel{\text{law}}{=} \mathbf{B} \stackrel{\text{law}}{=} B. \tag{1.21}$$

On the other hand, for $d \geq 2$, the processes $\mathbf{B}_*$ and $\mathbf{B}$ have different distributions. In particular, their covariance functions are defined, respectively, for $\mathbf{s} = (s_1, \ldots, s_d) \in \mathbb{R}_+^d$ and $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{R}_+^d$, by

$$E\big(\mathbf{B}_*(\mathbf{s})\mathbf{B}_*(\mathbf{t})\big) = \prod_{i=1}^{d} \big\{s_i \wedge t_i - s_i t_i\big\}, \tag{1.22}$$

$$E\big(\mathbf{B}(\mathbf{s})\mathbf{B}(\mathbf{t})\big) = \prod_{i=1}^{d} \big\{s_i \wedge t_i\big\} - \prod_{i=1}^{d} \big\{s_i t_i\big\}. \tag{1.23}$$

In the sequel, we will use the boldface notation $\mathbf{W}$, $\mathbf{B}_*$ and $\mathbf{B}$ in the general d-variate framework when $d \geq 1$ is arbitrary, and limit the formerly used notation, W and B, to the univariate case, when $d = 1$. Let now $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d) \in \mathbb{R}^d \in \big(-1, \infty \big)$ be a vector of constants. We define, respectively, the *weighted Wiener sheet* $\mathbf{W}^{(\boldsymbol{\gamma})}$, and the *weighted tied-down Brownian bridge* $\mathbf{B}_*^{(\gamma)}$, by

$$\mathbf{W}^{(\boldsymbol{\gamma})}\big(\mathbf{t}\big) = t_1^{\gamma_1} \ldots t_d^{\gamma_d} \mathbf{W}\big(\mathbf{t}\big) \quad \text{and} \quad \mathbf{B}_*^{(\boldsymbol{\gamma})}\big(\mathbf{t}\big) = t_1^{\gamma_1} \ldots t_d^{\gamma_d} \mathbf{B}\big(\mathbf{t}\big), \tag{1.24}$$

when $\mathbf{t} = (t_1, \ldots, t_d) \in (0,1]^d$. For convenience, we set $\mathbf{W}^{(\boldsymbol{\gamma})}\big(\mathbf{t}\big) = \mathbf{B}_*^{(\boldsymbol{\gamma})}\big(\mathbf{t}\big) = 0$, when $t_i = 0$ for some $i = 1, \ldots, d$. Introduce now the *upper-tail Wiener sheet* defined, for $\mathbf{t} = (t_1, \ldots, t_d) \in [0,1]^d$, by

$$\widetilde{\mathbf{W}}^{(\boldsymbol{\gamma})}(\mathbf{t}) = \int_{[t_1,1]\times\cdots\times[t_d,1]} u_1^{\gamma_1} \ldots u_d^{\gamma_d} \mathbf{W}(du_1, \ldots, du_d). \tag{1.25}$$

It is readily checked that, whenever $\gamma_i > -\frac{1}{2}$ for $i = 1, \ldots, d$, we have, for all $\mathbf{s} = (s_1, \ldots, s_d) \in [0,1]^d$ and $\mathbf{t} = (t_1, \ldots, t_d) \in [0,1]^d$,

$$\begin{aligned} E\big(\widetilde{\mathbf{W}}^{(\boldsymbol{\gamma})}(\mathbf{s})\widetilde{\mathbf{W}}^{(\boldsymbol{\gamma})}(\mathbf{t})\big) &= \prod_{i=1}^d \int_{s_i \vee t_i}^1 u_i^{2\gamma_i} du_i \\ &= \prod_{i=1}^d \Big\{ \frac{(1 - s_i^{2\gamma_i+1}) \wedge (1 - t_i^{2\gamma_i+1})}{1 + 2\gamma_i} \Big\}, \end{aligned}$$

so that we have the distributional identity (see, e.g. (3.11), p.505 in [5])

$$\widetilde{\mathbf{W}}^{(\boldsymbol{\gamma})}(\mathbf{t}) \stackrel{\text{law}}{=} \Big\{ \prod_{i=1}^d \frac{1}{1 + 2\gamma_i} \Big\} \mathbf{W}\big(1 - t_1^{1+2\gamma_1}, \ldots, 1 - t_d^{1+2\gamma_d}\big). \tag{1.26}$$

The following additional distributional identities will be useful, in view of the definitions (1.16) and (1.17) of Σ_i and Θ_i, for $i = 1, \ldots, d$. We set, for convenience, $\mathbf{0} = (0, \ldots, 0)$.

We define the *mean-centered weighted upper-tail Wiener sheet* by setting, for $\mathbf{t} = (t_1, \ldots, t_d) \in [0,1]^d$.

$$\widetilde{\mathbf{W}}_M^{(\boldsymbol{\gamma})}\big(\mathbf{t}\big) = \Sigma_1 \circ \cdots \circ \Sigma_d \widetilde{\mathbf{W}}^{(\gamma)}\big(\mathbf{t}\big) \tag{1.27}$$

Lemma 1.1. *For each* $\mathbf{t} \in [0,1]^d$*, we have*

$$\widetilde{\mathbf{W}}^{(\mathbf{0})}(\mathbf{t}) = \mathbf{W}\big(1 - t_1, \ldots, 1 - t_d\big) \stackrel{\text{law}}{=} \Theta_1 \circ \cdots \circ \Theta_d \mathbf{W}\big(\mathbf{t}\big), \tag{1.28}$$

and

$$\mathbf{W}_M \stackrel{\text{law}}{=} \Sigma_1 \circ \cdots \circ \Sigma_d \mathbf{W}\big(\mathbf{t}\big) \stackrel{\text{law}}{=} \Sigma_1 \circ \cdots \circ \Sigma_d \widetilde{\mathbf{W}}_M^{(\mathbf{0})} \stackrel{\text{law}}{=} \widetilde{\mathbf{W}}^{(\mathbf{0})}\big(\mathbf{t}\big). \tag{1.29}$$

Proof. The proofs of (1.28) and (1.29) are readily obtained by induction on d. For $d = 1$, we see that (1.28) is equivalent to the identity $W(1-t) \stackrel{\text{law}}{=} W(t) - W(1)$, which is obvious. Likewise, (1.29) reduces to the formula

$$W(t) - W(1) - \int_0^1 \{W(u) - W(1)\} du = W(t) - \int_0^1 W(u) du.$$

We now assume that (1.28) holds at rank $d-1$, so that

$$\mathbf{W}(1-t_1,\ldots,1-t_{d-1},t_d) \overset{\text{law}}{=} \Theta_1 \circ \ldots \circ \Theta_{d-1}\mathbf{W}(t_1,\ldots,t_d).$$

We now make use of the observation that

$$\mathbf{W}(1-t_1,\ldots,1-t_{d-1},1-t_d) \overset{\text{law}}{=} \Theta_d\mathbf{W}(1-t_1,\ldots,1-t_{d-1},t_d).$$

This, in combination with the easily verified fact that the operators $\Theta_1,\ldots,\Theta_d$ commute, readily establishes (1.28) at rank d. The completion of the proof of (1.29) is very similar and omitted. □

Under the above notation, Theorem 3.1 in [5] establishes the following fact.

Fact 1.1. *For $d=2$ and $\boldsymbol{\gamma} \in (-\frac{1}{2},\infty)^d$, we have*

$$\mathbf{B}_*^{(\boldsymbol{\gamma})} \overset{\text{Quad}}{=} \widetilde{\mathbf{W}}_M^{(\boldsymbol{\gamma})}. \tag{1.30}$$

Remark 1.2. Theorem 3.1 in [5] establishes (1.30), and, likewise, Corollary 3.1 of [5], establishes the forthcoming (1.31), only for $d=2$. It is natural to extend the validity of (1.30) to the case of $d=1$. In the forthcoming Section 1.5, we will make use of KL expansions to establish a result of the kind.

Remark 1.3. For $d=2$, we may rewrite (1.30) into the distributional identity, for $\gamma > -\frac{1}{2}$ and $\delta > -\frac{1}{2}$,

$$\begin{aligned}
&\int_0^1\int_0^1 s^{2\gamma}t^{2\delta}\Big\{\mathbf{W}(s,t)-s\mathbf{W}(1,t)-t\mathbf{W}(s,1)+st\mathbf{W}(1,1)\Big\}^2 \\
&\overset{\text{law}}{=} \int_0^1\int_0^1 \Big\{s^\gamma t^\delta\widetilde{\mathbf{W}}(s,t)-\int_0^1 u^\gamma t^\delta\widetilde{\mathbf{W}}(u,t)du \\
&\qquad -\int_0^1 s^\gamma v^\delta\widetilde{\mathbf{W}}(s,v)dv+\int_0^1\int_0^1 u^\gamma v^\delta\widetilde{\mathbf{W}}(u,v)dudv\Big\}^2 dsdt.
\end{aligned} \tag{1.31}$$

By combining (1.29), with (1.31), taken with $\gamma=\delta=0$, we obtain the distributional identity

$$\begin{aligned}
&\int_0^1\int_0^1 \Big\{\mathbf{W}(s,t)-s\mathbf{W}(1,t)-t\mathbf{W}(s,1)+st\mathbf{W}(1,1)\Big\}^2 \\
&\overset{\text{law}}{=} \int_0^1\int_0^1 \Big\{\mathbf{W}(s,t)-\int_0^1 \mathbf{W}(u,t)du \\
&\qquad -\int_0^1 \mathbf{W}(s,v)dv+\int_0^1\int_0^1 \mathbf{W}(u,v)dudv\Big\}^2 dsdt.
\end{aligned} \tag{1.32}$$

1.4. Weighted mean-centered Wiener processes ($d=1$)

In this section, we establish the KL expansion of the univariate *mean-centered weighted Wiener process*, defined, for $\gamma > -\frac{1}{2}$ and $t \in [0,1]$, by

$$W_\gamma(t) = \frac{1}{\sqrt{1+2\gamma}}\Big\{W\big(t^{1+2\gamma}\big)-\int_0^1 W\big(u^{1+2\gamma}\big)du\Big\}. \tag{1.33}$$

In view of (1.24), when $\gamma = 0$, we have $W_\gamma \overset{\text{Quad}}{=} B^{(\gamma)}$, therefore, by Theorem 1.2, the eigenvalues of the KL expansions of W_γ and $B^{(\gamma)}$ must coincide in this case. We will largely extend this result by giving, in Theorem 1.3 below, the complete KL expansion of $W_\gamma(\cdot)$ for an arbitrary $\gamma > -\frac{1}{2}$. We need first to recall a few basic facts about Bessel functions (we refer to [12], and to Section 2 in [4] for additional details and references).

For each $\nu \in \mathbb{R}$, the *Bessel function of the first kind*, of index ν, is defined by

$$J_\nu(x) = \left(\tfrac{1}{2}x\right)^\nu \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{4}x^2\right)^k}{\Gamma(\nu+k+1)\Gamma(k+1)} . \tag{1.34}$$

To render this definition meaningful when $\nu \in \{-1,-2,\ldots\}$ is a negative integer, we use the convention that $a/\infty = 0$ for $a \neq 0$. Since, when $\nu = -n$ is a negative integer, $\Gamma(\nu+k+1) = \Gamma(n+k+1) = \infty$ for $k = 0,\ldots,n-1$, the corresponding terms in the series (1.34) vanish, allowing us to write

$$J_{-n}(x) = (-1)^n J_n(x). \tag{1.35}$$

One of the most important properties of Bessel functions is related to the second order homogeneous differential equation

$$x^2y'' + xy' + (x^2 - \nu^2)y = 0. \tag{1.36}$$

When $\nu \in \mathbb{R}$ is noninteger, the Bessel functions J_ν and $J_{-\nu}$ provide a pair of linearly independent solutions of (1.36) on $(0,\infty)$. On the other hand, when $\nu = n$ is integer, $J_n(x)$ and $J_{-n}(x)$ are linearly dependent, via (1.35). To cover both of these cases, it is useful to introduce the *Bessel function of the second kind* of index ν. Whenever ν is noninteger, this function is defined by

$$Y_\nu(x) = \frac{J_\nu(x)\cos\nu\pi - J_{-\nu}(x)}{\sin\nu\pi} , \tag{1.37}$$

and whenever $\nu = n$ is integer, we set

$$Y_n(x) = \lim_{\nu\to n} \frac{J_\nu(x)\cos\nu\pi - J_{-\nu}(x)}{\sin\nu\pi}. \tag{1.38}$$

In view of the definitions (1.37)–(1.38), we see that, for an arbitrary $\nu \in \mathbb{R}$, J_ν and Y_ν provide a pair of linearly independent solutions of (1.36) on $(0,\infty)$. The behavior of the Bessel functions of first and second kind largely differ at 0. In particular, when $\nu > 0$, we have, as $x \downarrow 0$ (see, e.g., p.82 in [4]),

$$J_\nu(x) = (1+o(1))\frac{(\frac{1}{2})^\nu x^\nu}{\Gamma(\nu+1)}, \tag{1.39}$$

and

$$Y_\nu(x) = (1+o(1))\frac{\Gamma(\nu)}{\pi}\left(\tfrac{1}{2}x\right)^{-\nu}. \tag{1.40}$$

When $\nu > -1$, the positive roots (or zeros) of J_ν are isolated (see, e.g., Fact 2.1 in [4]) and form an increasing sequence

$$0 < z_{\nu,1} < z_{\nu,2} < \cdots , \tag{1.41}$$

such that, as $k \to \infty$,

$$z_{\nu,k} = \left\{k + \tfrac{1}{2}(\nu - \tfrac{1}{2})\right\}\pi + o(1). \tag{1.42}$$

The next fact is Theorem 1.4 of [4].

Fact 1.2. *For any $\gamma > -1$, or, equivalently, for each $\nu = 1/(2(1+\gamma)) > 0$, the KL expansion of $B^{(\gamma)}(t) = t^\gamma B(t)$ on $(0,1]$ is given by*

$$t^\gamma B(t) \overset{\text{law}}{=} \sum_{k=1}^{\infty} \omega_k \sqrt{\lambda_k}\, e_k(t), \tag{1.43}$$

where the $\{\omega_k : k \geq 1\}$ are i.i.d. random variables, and, for $k \geq 1$ and $t \in (0,1]$,

$$\lambda_k = \Big\{\frac{2\nu}{z_{\nu,k}}\Big\}^2 \quad \text{and} \quad e_k(t) = t^{\frac{1}{2\nu}-\frac{1}{2}} \Big\{\frac{J_\nu\big(z_{\nu,k} t^{\frac{1}{2\nu}}\big)}{\sqrt{\nu}\, J_{\nu-1}\big(z_{\nu,k}\big)}\Big\}. \tag{1.44}$$

Theorem 1.3. *Let $\gamma > -\frac{1}{2}$, or, equivalently, let $0 < \nu = 1/(2(1+\gamma)) < 1$. Then, the KL expansion of W_γ is given by*

$$W_\gamma(t) = \frac{1}{\sqrt{1+2\gamma}} \Big\{W\big(t^{1+2\gamma}\big) - \int_0^1 W\big(u^{1+2\gamma}\big) du\Big\} \overset{\text{law}}{=} \sum_{k=1}^{\infty} \omega_k \sqrt{\lambda_k}\, e_k(t), \tag{1.45}$$

where the $\{\omega_k : k \geq 1\}$ are i.i.d. random variables, and, for $k \geq 1$ and $t \in (0,1]$,

$$\lambda_k = \Big\{\frac{2\nu}{z_{\nu,k}}\Big\}^2 \quad \text{and} \quad e_k(t) = a t^{\frac{1}{2\nu}-\frac{1}{2}} \frac{J_{\nu-1}\big(z_{\nu,k}\, t^{\frac{1}{2\nu}}\big)}{\sqrt{\nu} J_{\nu-1}(z_{\nu,k})}. \tag{1.46}$$

The proof of Theorem 1.3 is given in Section 2.2. As an immediate consequence of this theorem, in combination with Fact 1.2, we obtain that the relation

$$W_\gamma \overset{\text{Quad}}{=} B^{(\gamma)}, \tag{1.47}$$

holds for each $\gamma > -\frac{1}{2}$. A simple change of variables transforms this formula into

$$\begin{aligned} &\frac{1}{(1+2\gamma)^2} \int_0^1 t^{\frac{-2\gamma}{1+2\gamma}} \Big\{W(t) - \frac{1}{2\gamma+1} \int_0^1 s^{\frac{-2\gamma}{1+2\gamma}} W(s) ds\Big\}^2 dt \\ &\qquad \overset{\text{law}}{=} \int_0^1 t^{2\gamma} \Big\{W(t) - tW(1)\Big\}^2 dt. \end{aligned} \tag{1.48}$$

1.5. Mean-centered Wiener processes ($d \geq 1$)

In this section, we establish the KL expansion of the (unweighted) multivariate *mean-centered Wiener process*, defined, for $\mathbf{t} \in [0,1]^d$, by

$$\mathbf{W}_M(t) \overset{\text{law}}{=} \Sigma_1 \circ \cdots \circ \Sigma_d \mathbf{W}(\mathbf{t}). \tag{1.49}$$

We obtain the following theorem.

Theorem 1.4. *The KL expansion of $\mathbf{W}_M$ is given by*

$$\mathbf{W}_M(\mathbf{t}) \overset{\text{law}}{=} \sum_{k_1=1}^{\infty} \cdots \sum_{k_d=1}^{\infty} \omega_{k_1,\ldots,k_d} \Big\{\prod_{i=1}^{d} \Big(\frac{\sqrt{2}\, \cos(k_i \pi t_i)}{k_i \pi}\Big)\Big\}, \tag{1.50}$$

where $\{\omega_{k_1,\ldots,k_d} : k_1 \geq 0, \ldots, k_d \geq 0\}$ denotes an array of i.i.d. normal $N(0,1)$ random variables.

The proof of Theorem 1.4 is postponed until Section 2.1.

2. Proofs

2.1. Proof of Theorems 1.1 and 1.4

In spite of the fact that Theorem 1.1 is a particular case of Theorem 1.3, it is useful to give details about its proof, to introduce the arguments which will be used later on, in the much more complex setup of weighted processes. We start with the following easy lemma. Below, we let $W_0(t)$ be as in (1.2).

Lemma 2.1. *The covariance function of* $W_0(\cdot)$ *is given, for* $0 \le s,t \le 1$, *by*

$$K_{W_0}(s,t) = E(W_0(s)W_0(t)) = s \wedge t - s - t + \tfrac{1}{2}s^2 + \tfrac{1}{2}t^2 + \tfrac{1}{3}. \tag{2.1}$$

Proof. In view of (1.2), we have the chain of equalities, for $0 \le s,t \le 1$,

$$\begin{aligned} E(W_0(s)W_0(t)) &= E\Big(W(s)W(t) - W(s)\int_0^1 W(u)du \\ &\quad - W(t)\int_0^1 W(u)du + \int_0^1\int_0^1 W(u)W(v)dudv\Big) \\ &= s\wedge t - \int_0^1 (s\wedge u)du - \int_0^1 (t\wedge u)du + \int_0^1\int_0^1 (u\wedge v)dudv, \end{aligned}$$

from where (2.1) is obtained by elementary calculations. □

Recalling the definition (1.5) of T_ζ, we set below $\zeta(\cdot) = W_0(\cdot)$.

Lemma 2.2. *Let* $\{y(t) : 0 \le t \le 1\}$ *denote an eigenfunction of the Fredholm transformation* T_{W_0}, *pertaining to the eigenvalue* $\lambda > 0$. *Then,* $y(\cdot)$ *is infinitely differentiable on* $[0,1]$, *and a solution of the differential equation*

$$\lambda y''(t) + y(t) = \int_0^1 y(u)du, \tag{2.2}$$

subject to the boundary conditions

$$y'(0) = y'(1) = 0. \tag{2.3}$$

Proof. By (2.1), we have, for each $t \in [0,1]$,

$$\begin{aligned} \lambda y(t) &= \int_0^t sy(s)ds + t\int_t^1 y(s)ds + (\tfrac{1}{2}t^2 - t)\int_0^1 y(s)ds \\ &\quad + \int_0^1 (\tfrac{1}{2}s^2 - s + \tfrac{1}{3})y(s)ds. \end{aligned} \tag{2.4}$$

It is readily checked that the RHS of (2.4) is a continuous function of $t \in [0,1]$. This, together with the condition that $\lambda > 0$ entails, in turn, that $y(\cdot)$ is continuous on $[0,1]$. By repeating this argument, a straightforward induction implies that $y(\cdot)$ is infinitely differentiable on $[0,1]$. This allows us to derivate both sides of (2.4) with respect to t. We so obtain that, for $t \in [0,1]$,

$$\lambda y'(t) = (t-1)\int_0^1 y(s)ds + \int_t^1 y(s)ds. \tag{2.5}$$

By setting, successively, $t = 0$ and $t = 1$ in (2.5), we get (2.3). Finally,(2.2) is obtained by derivating both sides of (2.5) with respect to t. □

Proof of Theorem 1.1. It is readily checked that the general solution of the differential equation (2.2) is of the form

$$y(t) = a\cos\Big(\frac{t}{\sqrt{\lambda}} + b\Big) + c, \tag{2.6}$$

for arbitrary choices of the constants $a, b, c \in \mathbb{R}$. The limit conditions (2.3) imply that, in (2.6), we must restrict b and λ to fulfill $b = 0$ and $\lambda \in \{1/(k^2\pi^2) : k \geq 1\}$. We now check that the function $y(t) = a\cos(k\pi t)$ is a solution of the equation (2.4), taken with $\lambda = 1/(k^2\pi^2)$. Towards this aim, we first notice that $y(t) = a\cos(k\pi t)$ fulfills

$$\int_0^1 y(s)ds = \frac{a}{k\pi}\Big[\sin(k\pi t)\Big]_{t=0}^{t=1} = 0. \tag{2.7}$$

This, in turn, readily entails that $y(t) = a\cos(k\pi t)$ satisfies the relation (2.5). By integrating both sides of this relation with respect to t, we obtain, in turn, that $y(t) = a\cos(k\pi t)$ fulfills

$$\begin{aligned}\lambda y(t) = \int_0^t sy(s)ds + t\int_t^1 y(s)ds + (\tfrac{1}{2}t^2 - t)\int_0^1 y(s)ds \\ + \int_0^1 (\tfrac{1}{2}s^2 - s + \tfrac{1}{3})y(s)ds + C,\end{aligned}$$

for some constant $C \in \mathbb{R}$. All we need is to check that $C = 0$ for some particular value of t. If we set $t = 1$, and make use of (2.7), this reduces to show, by integration by parts, that

$$\begin{aligned}\frac{\cos(k\pi)}{k^2\pi^2} &= \frac{1}{2}\int_0^1 s^2 y(s)ds = \frac{1}{2k\pi}\Big[t^2\sin(k\pi t)\Big]_{t=0}^{t=1} - \frac{1}{k\pi}\int_0^1 s\sin(k\pi s)ds \\ &= \frac{1}{(k\pi)^2}\Big[t\cos(k\pi t)\Big]_{t=0}^{t=1} - \frac{1}{(k\pi)^2}\int_0^1 \cos(k\pi s)ds = \frac{\cos(k\pi)}{k^2\pi^2}.\end{aligned}$$

The just-proved fact that $y(t) = a\cos(k\pi t)$ satisfies the relation (2.5) implies, in turn, that this same equation, taken with $y(t) = a\cos(k\pi t) + c$, reduces to

$$\frac{c}{k^2\pi^2} = c\int_0^1 (\tfrac{1}{2}s^2 - s + \tfrac{1}{3})ds = 0.$$

Since this relation is only possible for $c = 0$, we conclude that the eigenfunctions of T_{W_0} are of the form $y(t) = a\cos(k\pi t)$, for an arbitrary $a \in \mathbb{R}$. Given this fact, the remainder of the proof of Theorem 1.1 is straightforward. □

To establish Theorem 1.4, we will make use of the following lemma. We let $\mathbf{W}_M$ be defined as in (1.19), and recall (2.1).

Lemma 2.3. *We have, for* $\mathbf{s} = (s_1, \ldots, s_d) \in [0,1]^d$ *and* $\mathbf{t} = (t_1, \ldots, t_d) \in [0,1]^d$,

$$E\big(\mathbf{W}_M(\mathbf{s})\mathbf{W}_M(\mathbf{s})\big) = \prod_{i=1}^d K_{W_0}(s_i, t_i). \tag{2.8}$$

Proof. It follows from the following simple argument. We will show that, for an arbitrary $1 \leq j \leq d$,

$$\begin{aligned} &E\big(\{\Sigma_1 \circ \ldots \circ \Sigma_j \mathbf{W}(\mathbf{s})\}\{\Sigma_1 \circ \ldots \circ \Sigma_j \mathbf{W}(\mathbf{t})\}\big) \\ &= \Big\{\prod_{i=1}^{j} K_{W_0}(s_i, t_i)\Big\} \prod_{i=j+1}^{d} \{s_i \wedge t_i\}. \end{aligned} \tag{2.9}$$

Since $\Sigma_1, \ldots, \Sigma_d$ are linear mappings, the proof of (2.9) is readily obtained by induction on $j = 1, \ldots, d$. This, in turn, implies (2.8) for $j = d$. □

Proof of Theorem 1.4. In view of (2.8), the proof of the theorm follows readily from a repeated use of Lemma 4.1 in [5], which we state below for convenience. Let $\zeta_1(\mathbf{s})$, $\zeta_2(\mathbf{t})$, and $\zeta_3(\mathbf{s}, \mathbf{t})$ be three centered Gaussian processes, functions of $\mathbf{s} \in [0,1]^p$ and $\mathbf{t} \in [0,1]^q$. We assume that

$$E\big(\zeta_3(\mathbf{s}', \mathbf{t}')\big)\zeta_3(\mathbf{s}'', \mathbf{t}'')\big) = E\big(\zeta_1(\mathbf{s}')\big)\zeta_1(\mathbf{s}'')\big)E\big(\zeta_2(\mathbf{t}')\big)\zeta_2(\mathbf{t}'')\big).$$

Then, if the KL expansions of ζ_1 and ζ_2 are given by

$$\zeta_1(\mathbf{s}) \stackrel{\text{law}}{=} \sum_{k=1}^{\infty} \omega_k \sqrt{\lambda_{k,1}}\; e_{k,1}(\mathbf{s}) \quad \text{and} \quad \zeta_2(\mathbf{t}) \stackrel{\text{law}}{=} \sum_{k=1}^{\infty} \omega_k \sqrt{\lambda_{k,2}}\; e_{k,2}(\mathbf{t}),$$

the KL expansion of ζ_3 is given by

$$\zeta_3(\mathbf{s}, \mathbf{t}) \stackrel{\text{law}}{=} \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} \omega_{k,\ell} \sqrt{\lambda_{k,1}\lambda_{k,2}}\; e_{k,1}(\mathbf{s}) e_{k,2}(\mathbf{t}),$$

where $\{\omega_{k,\ell} : k \geq 1, \ell \geq 1\}$ denotes an array of iid normal $N(0,1)$ r.v.'s. □

2.2. Proof of Theorem 1.3

Recall that

$$W_\gamma(t) = \frac{1}{\sqrt{1+2\gamma}}\Big\{W(t^{1+2\gamma}) - \int_0^1 W(u^{1+2\gamma})du\Big\}. \tag{2.10}$$

Lemma 2.4. *The covariance function of $W_\gamma(\cdot)$ is given, for $0 \leq s, t \leq 1$, by*

$$\begin{aligned} K_{W_\gamma}(s,t) &= E(W_\gamma(s)W_\gamma(t)) \\ &= \frac{1}{2\gamma+1}\Big\{(s \wedge t)^{1+2\gamma} - s^{1+2\gamma} - t^{1+2\gamma} \\ &\quad + \frac{1+2\gamma}{2+2\gamma}s^{2+2\gamma} + \frac{1+2\gamma}{2+2\gamma}t^{2+2\gamma} + \frac{2}{(2+2\gamma)(3+2\gamma)}\Big\}. \end{aligned} \tag{2.11}$$

Proof. The proof of (2.11) being very similar to the above given proof of (2.1), we omit details. □

Lemma 2.5. *Fix any $\gamma > -\frac{1}{2}$. Let $e(t)$ be an eigenfunction of T_{W_γ} pertaining to a positive eigenvalue, $\lambda > 0$, of this operator. Then, the function $y(x) = e(x^{1/(1+2\gamma)})$ is solution on $[0,1]$ of the differential equation*

$$(1+2\gamma)^2 \lambda x^{\frac{2\gamma}{1+2\gamma}}\; y''(x) + y(x) = \frac{1}{1+2\gamma}\int_0^1 u^{-\frac{2\gamma}{1+2\gamma}}\; y(u)du, \tag{2.12}$$

with boundary conditions

$$y'(0) = y'(1) = 0. \tag{2.13}$$

Proof. Let $e(t)$ fulfill, for some $\lambda > 0$,

$$\lambda e(t) = \int_0^1 K_{W_\gamma}(s,t)e(s)ds.$$

By (2.11), we may rewrite this identity into

$$\begin{aligned}(2\gamma+1)\lambda e(t) &= \int_0^t s^{1+2\gamma}e(s)ds + t^{1+2\gamma}\int_t^1 e(s)ds \\ &- \int_0^1 s^{1+2\gamma}e(s)ds - t^{1+2\gamma}\int_0^1 e(s)ds + \Big\{\frac{1+2\gamma}{2+2\gamma}\Big\}\int_0^1 s^{2+2\gamma}e(s)ds \\ &+\Big\{\frac{1+2\gamma}{2+2\gamma}\Big\}t^{2+2\gamma}\int_0^1 e(s)ds + \frac{1}{(1+\gamma)(3+2\gamma)}\int_0^1 e(s)ds.\end{aligned} \tag{2.14}$$

A straightforward induction shows readily that any function $e(\cdot)$ fulfilling (2.14) is infinitely differentiable on $[0,1]$. This, in turn, allows us to derivate both sides of (2.14), as to obtain the equation

$$\lambda e'(t) = -t^{2\gamma}\int_0^t e(s)ds + t^{1+2\gamma}\int_0^1 e(s)ds. \tag{2.15}$$

Set now $e(t) = y(t^{1+2\gamma})$ in (2.15). By changing variables in the LHS of (2.15), we obtain the equation

$$(1+2\gamma)\lambda y'(t^{1+2\gamma}) = -\int_0^t e(s)ds + t\int_0^1 e(s)ds. \tag{2.16}$$

The relation (2.13) follows readily from (2.16), taken, successively, for $t = 0$ and $t = 1$. By derivating both sides of (2.16), and after setting $e(s) = y(s^{1+2\gamma}))$, we get

$$(1+2\gamma)^2\lambda y''(t^{1+2\gamma})t^{2\gamma} + y(t^{1+2\gamma}) = \int_0^1 y(s^{1+2\gamma})ds.$$

After making the change of variables $t = x^{1/(1+2\gamma)}$ and $s = u^{1/(1+2\gamma)}$, we may rewrite this last equation into

$$(1+2\gamma)^2\lambda x^{\frac{2\gamma}{1+2\gamma}}\, y''(x) + y(x) = \frac{1}{1+2\gamma}\int_0^1 u^{-\frac{2\gamma}{1+2\gamma}}\, y(u)du,$$

which is (2.16). The proof of Lemma 2.5 is now completed. □

Recall the definitions (1.34) and (1.37)–(1.38). In view of (2.12) the following fact will be instrumental for our needs (refer to Fact 2.3 in [4], and p.666 in [10]).

Fact 2.1. *Let $\lambda > 0$ and $\beta > -1$ be real constants. Then, the differential equation*

$$\lambda y''(x) + x^{2\beta}y(x) = 0, \tag{2.17}$$

has fundamental solutions on $(0,\infty)$ given by

$$x^{1/2}J_{\frac{1}{2(\beta+1)}}\Big(\frac{x^{\beta+1}}{(\beta+1)\sqrt{\lambda}}\Big) \quad \text{and} \quad x^{1/2}Y_{\frac{1}{2(\beta+1)}}\Big(\frac{x^{\beta+1}}{(\beta+1)\sqrt{\lambda}}\Big), \tag{2.18}$$

when $1/(2(\beta+1)) \in \mathbb{R}$, *and*

$$x^{1/2} J_{\frac{1}{2(\beta+1)}}\Big(\frac{x^{\beta+1}}{(\beta+1)\sqrt{\lambda}}\Big) \quad \text{and} \quad x^{1/2} J_{-\frac{1}{2(\beta+1)}}\Big(\frac{x^{\beta+1}}{(\beta+1)\sqrt{\lambda}}\Big), \tag{2.19}$$

when $1/(2(\beta+1))$ *is noninteger.*

Lemma 2.6. *Assume that* $\gamma > -\frac{1}{2}$. *Then, the solutions on* $[0,1]$ *of the differential equation*

$$(1+2\gamma)^2 \lambda x^{\frac{2\gamma}{1+2\gamma}} \, y''(x) + y(x) = 0, \tag{2.20}$$

with boundary conditions

$$y'(0) = y'(1) = 0, \tag{2.21}$$

are of the form

$$y(x) = a x^{1/2} J_{-\frac{1+2\gamma}{2(1+\gamma)}}\Big(\frac{x^{\frac{1+\gamma}{1+2\gamma}}}{(1+\gamma)\sqrt{\lambda}}\Big), \tag{2.22}$$

for some arbitrary constant $a \in \mathbb{R}$.

Proof. Set $\beta = -\gamma/(1+2\gamma)$ in Fact 2.1, and observe that the assumption that $\gamma > -\frac{1}{2}$ implies that $\beta > -1$, and that $1/(2(\beta+1))$ is noninteger. By Fact 2.1, the general solution on $(0,1]$ of the homogeneous differential equation (2.20) is of the form

$$y(x) = b x^{1/2} J_{\frac{1+2\gamma}{2(1+\gamma)}}\Big(\frac{x^{\frac{1+\gamma}{1+2\gamma}}}{(1+\gamma)\sqrt{\lambda}}\Big) + a x^{1/2} J_{-\frac{1+2\gamma}{2(1+\gamma)}}\Big(\frac{x^{\frac{1+\gamma}{1+2\gamma}}}{(1+\gamma)\sqrt{\lambda}}\Big), \tag{2.23}$$

where a and b are arbitrary constants. It is straightforward, given (1.39) and (1.40), that, for some constant $\rho_1 > 0$, as $x \downarrow 0$,

$$x^{1/2} J_{\frac{1+2\gamma}{2(1+\gamma)}}\Big(\frac{x^{\frac{1+\gamma}{1+2\gamma}}}{(1+\gamma)\sqrt{\lambda}}\Big) = (1+o(1))\rho_1 x^{\frac{1}{2}+\left\{\frac{1+2\gamma}{2(1+\gamma)}\right\}\frac{1+\gamma}{1+2\gamma}} \sim \rho_1 x, \tag{2.24}$$

whereas, for some constant $\rho_2 > 0$,

$$x^{1/2} Y_{-\frac{1+2\gamma}{2(1+\gamma)}}\Big(\frac{x^{\frac{1+\gamma}{1+2\gamma}}}{(1+\gamma)\sqrt{\lambda}}\Big) = (1+o(1))\rho_2 x^{\frac{1}{2}-\left\{\frac{1+2\gamma}{2(1+\gamma)}\right\}\frac{1+\gamma}{1+2\gamma}} \to \rho_2. \tag{2.25}$$

In view of (2.21), these relations imply that $b = 0$, so that (2.22) now follows from (2.23). □

Proof of Theorem 1.3. Set $\nu = 1/(2(1+\gamma))$. Obviously, the condition that $\gamma > -\frac{1}{2}$ is equivalent to $0 < \nu < 1$. We will show that the eigenvalues of T_{W_γ} are given by

$$\lambda_k = \Big\{\frac{2\nu}{z_{\nu,k}}\Big\} \quad \text{for} \quad k = 1, 2 \ldots . \tag{2.26}$$

Given this notation, it is readily checked that

$$\frac{1+2\gamma}{2(1+\gamma)} = 1 - \nu \quad \text{and} \quad 1 + 2\gamma = \frac{1}{\nu} - 1.$$

Therefore, letting $y(x)$ be as in (2.22), and setting $x = t^{1+2\gamma}$, we see that the eigenfunction $e_k(t)$ of T_{W_γ} pertaining to λ_k is of the form

$$e_k(t) = e(t) = y(t^{1+2\gamma}) = at^{\frac{1}{2\nu}-\frac{1}{2}} J_{\nu-1}\Big(z_{\nu,k}\, t^{\frac{1}{2\nu}}\Big), \tag{2.27}$$

for some $a \in \mathbb{R}$. Here, we may use Formula 50:10:2, p.529 in [11], namely, the relation

$$\frac{d}{dx}\Big\{x^\rho J_\rho(x)\Big\} = x^\rho J_{\rho+1}(x),$$

to show that, for some appropriate constant C,

$$\frac{d}{dx}\Big\{x^{1-\nu} J_{\nu-1}\big(z_{\nu,k}\, x\big)\Big\} = Cx^{1-\nu} J_\nu\big(z_{\nu,k}\, x\big) = 0, \tag{2.28}$$

for $x = 1$ and $x = 0$. This, in turn, allows to check that the function $y(x)$ in (2.27) fulfills (2.21). Now, we infer from (2.27), in combination with Fact 2.2, p.84 in [4], that

$$\begin{aligned} 1 &= \int_0^1 e_k^2(t)dt = 2\nu a^2 \int_0^1 t^{\frac{1}{2\nu}} J_{\nu-1}^2\Big(z_{\nu,k}\, t^{\frac{1}{2\nu}}\Big)dt^{\frac{1}{2\nu}} \\ &= 2\nu a^2 \int_0^1 u J_{\nu-1}^2\big(z_{\nu,k}\, u\big)du = \nu a^2 J_{\nu-1}^2\big(z_{\nu,k}\big). \end{aligned}$$

Given this last result, the completion of the proof of Theorem 1.3 is straightforward. □

References

[1] Adler, R. J. (1990). *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes.* IMS Lecture Notes-Monograph Series **12**. Institute of Mathematical Statistics, Hayward, California.

[2] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

[3] del Barrio, E., Cuesta-Albertos, J. A. and Matran, C. (2000). Contributions fo empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *Test.* **9** 1–96.

[4] Deheuvels, P. and Martynov, G. (2003). Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. In: *Progress in Probability* **55**. Birkhäuser, Basel., pp. 57–93.

[5] Deheuvels, P., Peccati, G. and Yor, M. (2006). On quadratic functionals of the Brownian sheet and related processes. *Stochastic Processes Appl.* **116** 493–538.

[6] Donati-Martin, C. and Yor, M. (1991). Fubini's theorem for double Wiener integrals and the variance of the Brownian path. *Ann. Inst. Henri Poincaré* **27** 181–200.

[7] Durbin, J. (1973). *Distribution Theory for Tests based on the Sample Distribution Function.* Regional Conference Series in Applied Mathematics **9**. SIAM, Philadelphia.

[8] Kac, M. (1980). *Integration on Function Spaces and some of its Applications.* Lezioni Fermiani della Scuola Normale Superiore, Pisa.

[9] Kac, M. and Siegert, A. J. F. (1947). An explicit representation of a stationary Gaussian process. *Ann. Math. Statist.* **18** 438–442.

[10] Kamke, E. (1959). *Differentialgleichungen Lösungsmethoden un Lösungen.* Chelsea, New York.

[11] Spanier, J. and Oldham, K. B. (1987). *An Atlas of Functions.* Hemisphere Publishing Corporation, Washington.

[12] Watson, G. N. (1952). *A Treatise on the Theory of Bessel Functions.* Cambridge University Press, Cambridge.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 77–95

DOI: 10.1214/074921706000000770

Fractional Brownian fields, duality, and martingales

Vladimir Dobrić[1] **and Francisco M. Ojeda**[2]

Lehigh University and Universidad Simón Bolívar and Lehigh University

Abstract: In this paper the whole family of fractional Brownian motions is constructed as a single Gaussian field indexed by time and the Hurst index simultaneously. The field has a simple covariance structure and it is related to two generalizations of fractional Brownian motion known as multifractional Brownian motions. A mistake common to the existing literature regarding multifractional Brownian motions is pointed out and corrected. The Gaussian field, due to inherited "duality", reveals a new way of constructing martingales associated with the odd and even part of a fractional Brownian motion and therefore of the fractional Brownian motion. The existence of those martingales and their stochastic representations is the first step to the study of natural wavelet expansions associated to those processes in the spirit of our earlier work on a construction of natural wavelets associated to Gaussian-Markov processes.

1. Introduction

The basic quantities describing movements of a viscous fluid are related through the Reynolds number. When the Reynolds number exceeds a critical value, which depends on the fluid, the average velocity of the region and its geometry, the flow becomes unstable and random. In his work on the theory of turbulent flow Kolmogorov proposed a model [10] which assumes that the kinetic energy in large scale motions of a turbulent flow is transferred to smaller scale turbulent motions. At smaller scales the Reynolds number associated with that region is reduced. When the Reynolds number of a region falls below the critical value of the region, turbulent motion stops and the remaining kinetic energy is dissipated as heat. Kolmogorov assumed that smaller scale turbulent motions can be described by a random field. Kolmogorov modeling assumptions produce the random field B_H that is self-similar with stationary increments and the second moment of its increments is of the form $\mathbb{E}\left|B_H(t+s)-B_H(s)\right|^2 = c\left|t\right|^{2H}$. Analyzing the properties of those fields [8, 9] Kolmogorov obtained a spectral representation for the fields with stationary increments.

In 1968 Mandelbrot and Van Ness interpreted the nonanticipating representation of B_H,

$$B_H(t)-B_H(s)=c\int_R\left((t-x)_+^{H-\frac{1}{2}}-(s-x)_+^{H-\frac{1}{2}}\right)db(x),$$

with respect to an orthogonal white noise db, obtained earlier by Pinsker and Yaglom, as a fractional integral, and called B_H the fractional Brownian motion in

[1]Department of Mathematics, 14 Packer Av., Lehigh University, Bethlehem, PA 18015, e-mail: vd00@lehigh.edu

[2]Departamento de Matemáticas, Universidad Simón Bolívar, Apartado 89000, Caracas 1080-A, Venezuela,e-mail: fojeda@usb.ve

AMS 2000 subject classifications: primary 60G15, 60H10; secondary 60G44.

Keywords and phrases: fractional Brownian motion, fractional Brownian fields, fundamental martingales, duality for fractional Brownian motions.

the case when b is a Brownian motion. Most of the above information is quoted from Molchan [14]. His paper is an excellent short review of the history of fractional Brownian motion before 1970. The index H is sometimes called the Hurst exponent, after the British hydrologist H.E. Hurst, who studied the annual flows of the Nile.

Fractional Brownian motion has applications in financial mathematics, telecommunication networks, hydrology, physical and biological sciences, just to mention a few.

An appropriate representation of fractional Brownian motion is important for analyzing its properties and for computational purposes. Meyer, Sellan and Taqqu developed a method for constructing wavelet expansions of fractional Brownian motion [12]. Their construction was based on fractional integration and differentiation of wavelet expansions of a Gaussian white noise and encompasses a large class of wavelets.

We have found an iterative method for obtaining an orthogonal expansion of a Gaussian Markov process [5] directly from its covariance function. It turns out that our method produces a wavelet expansion if time is measured by the natural measure associated to the process. It is therefore natural to ask if it is possible to construct a "natural" wavelet expansion associated with a fractional Brownian motion (fBm in short) in the spirit of our work in [5]. The main properties used in our construction of wavelets associated with the Gaussian Markov processes were the invariance of the processes on pinning (stays Markov), independence of associated pinned processes, and the existence of associated martingales.

The first step toward finding natural wavelets for fBm is to investigate invariances for the entire class of fractional Brownian motions. That means considering the family of processes $(B_H)_{0<H<1}$ where $B_H = (B_H(t))_{t\in\mathbb{R}}$ is a fBm. Let us point out that it is not straight forward to prove that, for the fixed Hurst index H, the covariance function,

$$\mathbb{E}B_H(t)\,B_H(s) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \tag{1}$$

of fBm is indeed a positive definite form . It uses tricks and the fact that the characteristic function φ of an $2H$ stable random variable is given by $\varphi(t) = e^{-a|t|^{2H}}$ for some constant a. In the search for invariances relevant to achieving our goal, that is to obtaining natural wavelet expansions for fBm, we have constructed a Gaussian field, indexed by $\mathbb{R}\times(0,1)$, that encompasses at once all fractional Brownian motions for all H. More precisely, this field has the property that, when the second parameter, H, is fixed, the resulting process is an H-fBm. We started our construction by using already mentioned nonanticipating representation of fBm, which, for $0 < H < 1$, can be written as

$$B_H(t) = \frac{\sqrt{\Gamma(2H+1)\sin(\pi H)}}{\Gamma\left(H+\frac{1}{2}\right)} \int_{-\infty}^{\infty} (t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}}\, dW(x) \tag{2}$$

for $t \in R$, where $(W(t))_{t\in R}$ is a Brownian motion. When we started the computations our hope was that the covariance of our Gaussian field would resemble the covariance of fractional Brownian motion (equation (1)), with $2H$ substituted by $H+H^{'}$ modulo a function of H and $H^{'}$. This does not turn out to be the case. The resulting covariance was more complex. Its form revealed that the nonanticipating representation may not be the most "efficient" representation of fBm. Section 3 of this paper contains all computations relevant to finding that covariance.

Although the Gaussian field developed in Section 3 did not meet our objective, its structure revealed what intrinsic dependences of the nonanticipating representation of fBm are, and what should be done to construct a new field having, for our purposes, the "right" covariance, namely the covariance

$$\mathbb{E}B_H(t)\,B_{H'}(s) = \alpha_{H,H'}(|t|^{H+H'} + |s|^{H+H'} - |t-s|^{H+H'}). \tag{3}$$

We have succeeded to obtain such a field. Furthermore we have found a nonanticipating representation of $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$, which, when H is fixed, equals in distribution to the standard nonanticipating representation of fBm. Those results are presented in Section 4.

Our new field has the property that when $H + H' = 1$ the right hand side of (3) becomes the covariance of a Brownian motion. In that case we call $B_H, B_{H'}$ a dual pair. A particular property of a dual pair is that it generates two martingales, one driving B_H and the other $B_{H'}$. In [4] we have obtained stochastic integral representations for those martingales. It turns out that our representation coincides with the fundamental martingale of a fBm discovered by Molchan and Golosov [13, 15]. Our work on that subject is in Section 5.

Lévy Véhel and Peltier [23] and Benassi et al. [2] have independently introduced a multifractional Brownian motion. In Section 6 we clarify the relationship between multifractional Brownian motion and the fractional Brownian fields introduced in Section 3 and 4. Furthermore we point out and correct an error in the covariance of the multifractional Brownian motion obtained from the nonanticipating moving average of fBm and show that in fact the processes of Lévy Véhel and Peltier [23] and Benassi et al. [2] are not the same, as it has been claimed in Cohen [3].

For readers convenience some basic facts about fractional Brownian motions and the notation that is used in the paper is included in Section 2.

2. Notation and preliminaries

This section is a brief overview of some properties of the fractional Brownian motion. The material presented here can be found in Taqqu [22], Molchan [14], chapter 7 in Samorodnitsky and Taqqu [21] and Dzhaparidze and Van Zanten [6]. All the processes and fields considered will be real valued.

The following proposition summarizes some properties of H-self-similar processes with stationary increments and finite second moments.

Proposition 1. *If $Z = \{Z_t\}_{t\in\mathbb{R}}$ is H-self-similar with stationary increments and finite second moments then*

1. $Z_0 = 0$ *with probability one.*
2. *If* $H \neq 1$, *then* $\mathbb{E}Z_t = 0$ *for all* $t \in \mathbb{R}$.
3. $H \leq 1$.
4. *The covariance of* Z *is given by*

$$\mathbb{E}(Z_s Z_t) = \frac{\mathbb{E}\left(Z_1^2\right)}{2}\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right). \tag{4}$$

As already mentioned in the introduction Kolmogorov described an H-self-similar process with stationary increments and its covariance (4) in [9]. A centered Gaussian process $\{Z_t\}_{t\in\mathbb{R}}$ with covariance (4) is called a fractional Brownian motion (fBm) with index H. If $\mathbb{E}\left(Z_1^2\right) = 1$, the process is a standard fBm. Since the distribution of a Gaussian process is entirely determined by its covariance, (4) implies that fBm is H-self-similar and has stationary increments. Conversely, an

H-self-similar Gaussian process with stationary increments is a fBm. For $H = \frac{1}{2}$ a standard fractional Brownian motion is a standard Brownian motion on the real line, that is,

$$B_{\frac{1}{2}}(t) = \begin{cases} B^1(-t) & t < 0 \\ B^2(t) & t \geq 0 \end{cases},$$

where $\{B_t^1\}_{t\geq 0}$, $\{B_t^2\}_{t\geq 0}$ are two independent standard Brownian motions.

Unless otherwise stated all Brownian motions and fractional Brownian motions appearing on this work will be assumed to be standard.

Using Kolmogorov-Centsov criterion for Hölder continuity of a process, for every γ in $(0, H)$, an H-fBm has a continuous modification that is locally Hölder continuous with exponent γ. Throughout the paper we will assume that fBm is such a modification.

Let $\{B_t\}_{t\in R}$ and $\{W_t\}_{t\in R}$ be Brownian motions on the real line. Then the process defined by the moving average stochastic integral

$$B_H(t) = \frac{1}{c_H}\int_{-\infty}^{\infty} (t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}}\, dB_x \text{ for } t \in \mathbb{R} \tag{5}$$

is a fBm, where

$$c_H = \frac{\Gamma\left(H+\frac{1}{2}\right)}{\sqrt{\Gamma(2H+1)\sin(\pi H)}} \tag{6}$$

and $x_+^\alpha = 0$ if $x \leq 0$ and $x_+^\alpha = x^\alpha$ if $x > 0$. This representation is called nonanticipating, since for $t \geq 0$, it involves only integration over $(-\infty, t]$. Another moving average representation of fBm for $H \neq \frac{1}{2}$ is given by

$$W_H(t) = \frac{1}{d_H}\int_{-\infty}^{\infty} |t-x|^{H-\frac{1}{2}} - |x|^{H-\frac{1}{2}}\, dW_x \text{ for } t \in \mathbb{R} \tag{7}$$

where

$$d_H = \frac{\Gamma\left(H+\frac{1}{2}\right)}{\sqrt{\Gamma(2H+1)}}\sqrt{2\left(\frac{1-\sin(\pi H)}{\sin(\pi H)}\right)}, \tag{8}$$

and when $H = \frac{1}{2}$ by

$$W_{\frac{1}{2}}(t) = \frac{1}{d_{\frac{1}{2}}}\int_{-\infty}^{\infty} \log\left(\frac{1}{|t-x|}\right) - \log\left(\frac{1}{|x|}\right) dW_x \text{ for } t \in \mathbb{R} \tag{9}$$

where $d_{\frac{1}{2}} = \pi$. This representation is called well-balanced.

The odd and even part of fBm play an important role in this paper. If $\{Z_H(t)\}_{t\in\mathbb{R}}$ is a H-fBm, then the odd and the even process of fBm are defined by

$$Z_H^o(t) = \frac{1}{2}\left(Z_H(t) - Z_H(-t)\right), \text{ and } Z_H^e(t) = \frac{1}{2}\left(Z_H(t) + Z_H(-t)\right),\ t \in \mathbb{R}$$

respectively. It is straightforward to check that the processes Z_H^o and Z_H^e are independent and that their covariances are given by

$$\mathbb{E}Z_H^o(s)\, Z_H^o(t) = \frac{1}{4}\left(|s+t|^{2H} - |s-t|^{2H}\right) \tag{10}$$

and

$$\mathbb{E}Z_H^e(s)\,Z_H^e(t) = \frac{|s|^{2H} + |t|^{2H}}{2} - \frac{|s+t|^{2H} + |s-t|^{2H}}{4} \tag{11}$$

for s, t in $\mathbb{R}$. Both Z_H^e and Z_H^o are H-self-similar processes and by assumption have continuous sample paths with probability 1. Clearly, each path of the process Z_H^o (Z_H^e) is an odd (even) function on a set of probability 1, and so it suffices to consider these processes for $t \geq 0$. For $H = \frac{1}{2}$ the odd and even part of fBm are Brownian motions (up to a constant multiple).

Moving average representations for the odd and even part of fBm can be found in Dzhaparidze and Van Zanten [6], Nuzman and Poor [17].

3. Dependent fractional Brownian field

In this section we will assume that each element of the family $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ is represented by a nonanticipating moving average stochastic integral (5) respect to the same $\{B_t\}_{t\in\mathbb{R}}$. For $0 < H, H^{'} < 1$ set

$$c_{H,H'} = \frac{\sqrt{\Gamma(2H+1)\sin(\pi H)}\sqrt{\Gamma(2H'+1)\sin(\pi H')}}{\pi}.$$

Theorem 2. *Let K be the covariance $\mathbb{E}B_H(t)\,B_{H'}(s)$. If $H+ H^{'} \neq 1$ then*

$$\begin{aligned} K = c_{H,H'}\Gamma\left(-\left(H+H^{'}\right)\right)&\left\{\cos\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)\cos\left(\left(H+H^{'}\right)\frac{\pi}{2}\right)\right.\\ &\times\left(|t-s|^{H+H^{'}} - |t|^{H+H^{'}} - |s|^{H+H^{'}}\right)\\ &-\sin\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)\sin\left(\left(H+H^{'}\right)\frac{\pi}{2}\right)\\ &\left.\times\left(sgn(t)\,|t|^{H+H^{'}} - sgn(s)\,|s|^{H+H^{'}} - sgn(t-s)\,|t-s|^{H+H^{'}}\right)\right\}. \end{aligned} \tag{12}$$

If $H + H^{'} = 1$ and $t \neq s$, $t \neq 0$, $s \neq 0$, then

$$\begin{aligned} K = c_{H,H'}&\left\{\cos\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)\frac{\pi}{2}\left(|t|+|s|-|t-s|\right)\right.\\ &-\sin\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)\\ &\left.\times\left(t\log|t| - s\log|s| - (t-s)\log|t-s|\right)\right\}. \end{aligned} \tag{13}$$

If $H + H^{'} = 1$ and $t = s$, $t \neq 0$, then

$$K = c_{H,H'}\cos\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)\frac{\pi}{2}\left\{|t|+|s|\right\}. \tag{14}$$

If $H + H^{'} = 1$ and $t = 0$ or $s = 0$, then

$$K = 0. \tag{15}$$

Fourier transform is the main tool in proving the theorem. Before proving it we will establish some technical results.

Let $\widehat{f}(\xi) = \int_{-\infty}^{\infty} e^{i\xi x} f(x)\,dx$ be the Fourier transform of a function that belong to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ With this convention $\sqrt{2\pi}\,\|f\|_2 = \left\|\widehat{f}\right\|_2$ for all $f \in L^2(\mathbb{R})$.

Lemma 3. *For $0 < H < 1$ and $t \in \mathbb{R}$*

$$\left(\widehat{\frac{(t-\cdot)_{+}^{H-\frac{1}{2}} - (-\cdot)_{+}^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)}}\right)(\xi) = \frac{e^{it\xi}-1}{i\xi}\,(i\xi)^{-\left(H-\frac{1}{2}\right)}, \tag{16}$$

$$\left(\widehat{\frac{(t-\cdot)_{-}^{H-\frac{1}{2}} - (-\cdot)_{-}^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)}}\right)(\xi) = -\frac{e^{it\xi}-1}{i\xi}\,(-i\xi)^{-\left(H-\frac{1}{2}\right)}, \tag{17}$$

where

$$(i\xi)^{-\left(H-\frac{1}{2}\right)} = \begin{cases} |\xi|^{-\left(H-\frac{1}{2}\right)} e^{i\frac{\pi}{2}\left(H-\frac{1}{2}\right)} & \xi < 0 \\ |\xi|^{-\left(H-\frac{1}{2}\right)} e^{-i\frac{\pi}{2}\left(H-\frac{1}{2}\right)} & \xi > 0 \end{cases}. \tag{18}$$

Proof. Equation (16) follows from the theory of fractional integration and differentiation, see for example Lemma 1 and Lemma 3 in [20]. Let

$$f_t(x) = \frac{(t-x)_{+}^{H-\frac{1}{2}} - (-x)_{+}^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)},$$

then the identity

$$\frac{(t-x)_{-}^{H-\frac{1}{2}} - (-x)_{-}^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} = f_{-t}(-x)$$

is a consequence of the relation $x_{+}^{\alpha} = (-x)_{-}^{\alpha}$. Equation (17) now follows from (16). $\square$

Lemma 4. *For $s, t \in \mathbb{R}$ and $0 < H, H^{'} < 1$ let*

$$I_1 \stackrel{def}{=} \int_0^\infty \frac{\sin^2\left(\frac{t\xi}{2}\right) + \sin^2\left(\frac{s\xi}{2}\right) - \sin^2\left(\frac{\xi(t-s)}{2}\right)}{\xi^{1+H+H'}}\,d\xi. \tag{19}$$

If $H + H^{'} \neq 1$ then

$$\begin{aligned} I_1 = {} & \frac{\Gamma\left(-\left(H+H^{'}\right)\right)\cos\left(\left(H+H^{'}\right)\frac{\pi}{2}\right)}{2} \\ & \times \left(|t-s|^{H+H^{'}} - |t|^{H+H^{'}} - |s|^{H+H^{'}}\right) \end{aligned} \tag{20}$$

and if $H + H^{'} = 1$ then

$$I_1 = \frac{\pi}{4}\left(|t| + |s| - |t-s|\right).$$

Proof. The proof is trivial if either $s = 0$ or $t = 0$, so assume that $s \neq 0$, $t \neq 0$. Formula 3.823 in Gradshteyn and Ryzhik [7] states

$$\int_0^\infty x^{\mu-1}\sin^2(ax)\,dx = -\frac{\Gamma(\mu)\cos\left(\frac{\mu\pi}{2}\right)}{2^{\mu+1}a^{\mu}} \text{ for } a > 0 \text{ and } -2 < \operatorname{Re}(\mu) < 0. \tag{21}$$

When $H+H^{'}\neq 1$ set $\mu=-\left(H+H^{'}\right)$. Observe that $\sin^2(x)=\sin^2(|x|)$. Applying the identity (21) to the right hand side of (19) when $s=t$ gives

$$I_1=-\Gamma\left(-H-H^{'}\right)\cos\left(\left(H+H^{'}\right)\frac{\pi}{2}\right)|t|^{H+H^{'}},$$

and when $s\neq t$

$$\begin{aligned}I_1=&\frac{\Gamma\left(-H-H^{'}\right)}{2}\cos\left(\left(H+H^{'}\right)\frac{\pi}{2}\right)\beta(H+H^{'})\\&\times\left\{|t-s|^{H+H^{'}}-|t|^{H+H^{'}}-|s|^{H+H^{'}}\right\}.\end{aligned}$$

In the case of $H+H^{'}=1$ we will use formula 3.821 (9) from Gradshteyn and Ryzhik [7],

$$\int_0^\infty\frac{\sin^2(ax)}{x^2}dx=\frac{a\pi}{2}\text{ for }a>0.$$

to conclude that when $s\neq t$ then $I_1=\pi/4\left(|t|+|s|-|t-s|\right)$, and when $s=t$ then $I_1=1/2\,|t|\,\pi$. □

Lemma 5. *For $s,t\in\mathbb{R}$ and $0<H,H^{'}<1$ let*

$$I_2\overset{def}{=}\int_0^\infty\frac{\sin\left(\xi\left(t-s\right)\right)+\sin\left(s\xi\right)-\sin\left(t\xi\right)}{\xi^{1+H+H^{'}}}d\xi. \tag{22}$$

If $s=t$ or $s=0$ or $t=0$ then $I_2=0$. Otherwise, if $H+H^{'}\neq 1$ then

$$\begin{aligned}I_2=&\Gamma\left(-\left(H+H^{'}\right)\right)\sin\left(\left(H+H^{'}\right)\frac{\pi}{2}\right)\\&\times\left\{sgn\left(t\right)|t|^{H+H^{'}}-sgn\left(s\right)|s|^{H+H^{'}}-sgn\left(t-s\right)|t-s|^{H+H^{'}}\right\},\end{aligned} \tag{23}$$

and if $H+H^{'}=1$ then

$$I_2=t\log|t|-s\log|s|-(t-s)\log|t-s|. \tag{24}$$

Proof. The proof is trivial if either $s=0$ or $t=0$ or $s=t$, so assume that $s\neq 0$, $t\neq 0$ and $s\neq t$. Formula 3.761 (4) in Gradshteyn and Ryzhik [7] reads:

$$\int_0^\infty x^{\mu-1}\sin(ax)\,dx=\frac{\Gamma(\mu)}{a^\mu}\sin\left(\frac{\mu\pi}{2}\right)\text{ for }a>0\text{ and }0<|\mathrm{Re}\,(\mu)|<1. \tag{25}$$

Set $\mu=-\left(H+H^{'}\right)$, and observe that $\sin(\xi x)=\mathrm{sgn}(x)\sin(\xi\,|x|)$ Applying (25) to the right hand side of (22) in the case when $0<H+H^{'}<1$ yields (23). In the case when $1 < H + H^{'} < 2$, by the dominated convergence theorem, it follows that

$$I_2 = \lim_{\substack{a \to 0^+ \\ b \to \infty}} \int_a^b \frac{sgn\,(t-s)\sin\left(\xi\,|t-s|\right) + sgn\,(s)\sin\left(|s|\,\xi\right) - sgn\,(t)\sin\left(|t|\,\xi|\right)}{\xi^{1+H+H'}} d\xi$$

$$= \lim_{\substack{a \to 0^+ \\ b \to \infty}} \frac{1}{H+H'} \int_a^b \left(sgn\,(t-s)\,|t-s|\cos\left(\xi\,|t-s|\right)\right.$$

$$\left. + sgn\,(s)\,|s|\cos\left(|s|\,\xi\right) - sgn\,(t)\,|t|\cos\left(|t|\,\xi\right)\right)\xi^{-H-H^{'}} d\xi$$

$$= \frac{1}{(H+H')\,(H+H'-1)} \int_0^\infty \left(-sgn\,(t-s)\,|t-s|^2\sin\left(\xi\,|t-s|\right)\right.$$

$$\left. -sgn\,(s)\,|s|^2\sin\left(|s|\,\xi\right) + sgn\,(t)\,|t|^2\sin\left(|t|\,\xi\right)\right)\xi^{1-H-H^{'}} d\xi,$$

where the last two equalities are result of integration by parts and the fact that the boundary terms converge to zero as $b \to \infty$, and $a \to 0^+$. Applying (25) with $\mu = 2 - \left(H + H^{'}\right)$ and using $x\Gamma\,(x) = \Gamma\,(x+1)$ the equation (23) now follows readily.

Let us turn to the case $H + H^{'} = 1$. For $x \in [1, 1.5]$ set

$$f(x) = \int_0^\infty \frac{sgn\,(t-s)\sin\left(\xi\,|t-s|\right) + sgn\,(s)\sin\left(|s|\,\xi\right) - sgn\,(t)\sin\left(|t|\,\xi\right)}{\xi^{1+x}} d\xi.$$

When $x = H + H^{'}$ the function f equals to the right hand side of (23), The integrand that defines f is bounded by

$$g\,(\xi) = \begin{cases} \frac{|sgn(t-s)\sin(\xi|t-s|) + sgn(s)\sin(|s|\xi) - sgn(t)\sin(|t|\xi)|}{\xi^{1+1.5}} & \xi \in (0,1) \\ \frac{3}{\xi^{1+1}} & \xi \in [1,\infty) \end{cases}$$

which is an integrable function. By the dominated convergence theorem f is continuous on $[1, 1.5]$. Finally $\lim_{x \downarrow 1} f\,(x)$ establishes (24), where we used (23) and the gamma function property $y\Gamma\,(y) = \Gamma\,(y+1)$ to rewrite $f\,(x)$ for $x \in (1, 1.5]$ as

$$f\,(x) = \frac{\Gamma\,(2-x)}{(-x)\,(1-x)} \sin\left(\frac{x\pi}{2}\right)\left\{sgn\,(t)\,|t|^x - sgn\,(s)\,|s|^x - sgn\,(t-s)\,|t-s|^x\right\}$$

and L'Hospital rule to compute the limit

$$\lim_{x \downarrow 1} \frac{sgn\,(t)\,|t|^x - sgn\,(s)\,|s|^x - sgn\,(t-s)\,|t-s|^x}{(1-x)}.$$

□

We have prepared the groundwork to prove Theorem 2.

Proof. By the Ito isometry

$$K = \frac{1}{c_H c_{H'}} \int_{-\infty}^\infty \left((t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}}\right)\left((s-x)_+^{H^{'}-\frac{1}{2}} - (-x)_+^{H^{'}-\frac{1}{2}}\right) dx,$$

and by Plancherel identity and Lemma 3

$$\begin{aligned}K &= \frac{\Gamma\left(H+\frac{1}{2}\right)\Gamma\left(H^{'}+\frac{1}{2}\right)}{2\pi c_H c_{H'}}\int_{-\infty}^{\infty}\frac{e^{it\xi}-1}{i\xi}(i\xi)^{-\left(H-\frac{1}{2}\right)}\overline{\frac{e^{is\xi}-1}{i\xi}(i\xi)^{-\left(H^{'}-\frac{1}{2}\right)}}d\xi \\ &= \frac{\Gamma\left(H+\frac{1}{2}\right)\Gamma\left(H^{'}+\frac{1}{2}\right)}{2\pi c_H c_{H'}}\left\{\int_{-\infty}^{0}\frac{\left(e^{it\xi}-1\right)\left(e^{-is\xi}-1\right)}{|\xi|^{1+H+H'}}e^{-i\left(H^{'}-H\right)\frac{\pi}{2}}d\xi\right. \\ &\quad\left.+\int_{0}^{\infty}\frac{\left(e^{it\xi}-1\right)\left(e^{-is\xi}-1\right)}{|\xi|^{1+H+H'}}e^{i\left(H^{'}-H\right)\frac{\pi}{2}}d\xi\right\}.\end{aligned}$$

Substituting $\xi^{'}=-\xi$ in the first integral of the last equality above and then combining the two integrals yields

$$K=\frac{\Gamma\left(H+\frac{1}{2}\right)\Gamma\left(H^{'}+\frac{1}{2}\right)}{\pi c_H c_{H'}}\operatorname{Re}\left(e^{i\left(H^{'}-H\right)\frac{\pi}{2}}\int_{0}^{\infty}\frac{\left(e^{it\xi}-1\right)\left(e^{-is\xi}-1\right)}{|\xi|^{1+H+H'}}d\xi\right). \tag{26}$$

Using Euler's formula on $e^{i\left(H^{'}-H\right)\frac{\pi}{2}}$ and $\left(e^{it\xi}-1\right)\left(e^{-is\xi}-1\right)$ and the identity $\sin^2 x=\frac{1-\cos 2x}{2}$ we obtain

$$\begin{aligned}&\operatorname{Re}\left(e^{i\left(H^{'}-H\right)\frac{\pi}{2}}\int_{0}^{\infty}\frac{\left(e^{it\xi}-1\right)\left(e^{-is\xi}-1\right)}{|\xi|^{1+H+H'}}d\xi\right) \\ &=\cos\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)2\int_{0}^{\infty}\frac{\sin^2\left(\frac{t\xi}{2}\right)+\sin^2\left(\frac{s\xi}{2}\right)-\sin^2\left(\frac{\xi(t-s)}{2}\right)}{\xi^{1+H+H'}}d\xi \\ &\quad-\sin\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)\int_{0}^{\infty}\frac{\sin\left(\xi\left(t-s\right)\right)+\sin\left(s\xi\right)-\sin\left(t\xi\right)}{\xi^{1+H+H'}}d\xi.\end{aligned} \tag{27}$$

Observing that

$$\frac{\Gamma\left(H+\frac{1}{2}\right)\Gamma\left(H^{'}+\frac{1}{2}\right)}{c_H c_{H'}\pi}=\frac{\sqrt{\Gamma\left(2H+1\right)\sin\left(\pi H\right)}\sqrt{\Gamma\left(2H'+1\right)\sin\left(\pi H'\right)}}{\pi}, \tag{28}$$

the expressions for $\mathbb{E}B_H(t)B_{H'}(s)$ now follow from equations (26), (27), (28) and Lemmas 4, 5 □

We will call a centered Gaussian field $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ with the covariance given by Theorem 2 a dependent fractional Brownian field and refer to it as dfBf. The rest of the section elaborates on a property of the field that justifies that name.

Let $\{B_H^o(t)\}_{t\in[0,\infty),H\in(0,1)}$ and $\{B_H^e(t)\}_{t\in[0,\infty),H\in(0,1)}$, be the odd and even part of the dfBf $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$, that is

$$B_H^o(t)=\frac{B_H(t)-B_H(-t)}{2}\text{ and }B_H^e(t)=\frac{B_H(t)+B_H(-t)}{2},\ t\geq 0.$$

For $H+H^{'}\neq 1$ set

$$\begin{aligned}a_{H,H'}&=-2\frac{\sqrt{\Gamma\left(2H+1\right)\sin\left(\pi H\right)}\sqrt{\Gamma\left(2H'+1\right)\sin\left(\pi H'\right)}}{\pi} \\ &\quad\times\Gamma\left(-\left(H+H^{'}\right)\right)\cos\left(\left(H^{'}-H\right)\frac{\pi}{2}\right)\cos\left(\left(H+H^{'}\right)\frac{\pi}{2}\right)\end{aligned} \tag{29}$$

and for $H + H^{'} = 1$

$$a_{H,H'} = \sqrt{\Gamma(2H+1)\Gamma(3-2H)}\sin^2(\pi H). \tag{30}$$

Theorem 6. *The covariance of the odd part and the even part of dfBf are*

$$\mathbb{E}B_H^o(t)\,B_{H'}^o(s) = a_{H,H'}\frac{|t+s|^{H+H^{'}} - |t-s|^{H+H^{'}}}{4} \tag{31}$$

and

$$\begin{aligned}\mathbb{E}B_H^e(t)\,B_{H'}^e(s) \\ = a_{H,H'}\left(\frac{|t|^{H+H^{'}} + |s|^{H+H^{'}}}{2} - \frac{|t-s|^{H+H^{'}} + |t+s|^{H+H^{'}}}{4}\right)\end{aligned} \tag{32}$$

respectively.

Proof. The covariance of the dfBf (Theorem 2), is of the form

$$\begin{aligned}\mathbb{E}B_H(t)\,B_{H'}(s) = f\left(H,H^{'}\right)\left(|t|^{H+H^{'}} + |s|^{H+H^{'}} - |t-s|^{H+H^{'}}\right) \\ + g\left(s,t,H,H^{'}\right).\end{aligned} \tag{33}$$

It is a matter of straightforward computation to check that in both cases

$$\mathbb{E}B_H(t)\,B_{H'}(s) + \mathbb{E}B_H(-t)\,B_{H'}(-s) \text{ and } \mathbb{E}B_H(t)\,B_{H'}(-s) + \mathbb{E}B_H(-t)\,B_{H'}(s)$$

the g function cancels. The result of the theorem follows by simple algebraic manipulation of the first part of the right hand side of (33) only. □

The Gaussian fields $\{B_H^o(t)\}_{t\in[0,\infty),H\in(0,1)}$ and $\{B_H^e(t)\}_{t\in[0,\infty),H\in(0,1)}$ with covariances given by Theorem 6, will be called the odd and the even fractional Brownian field respectively. It is very simple to check that for every $a > 0$

$$\{B_H^o(at)\}_{t\in[0,\infty),H\in(0,1)} \overset{f.d.d.}{=} \{a^H B_H^o(at)\}_{t\in[0,\infty),H\in(0,1)}$$

and

$$\{B_H^e(at)\}_{t\in[0,\infty),H\in(0,1)} \overset{f.d.d.}{=} \{a^H B_H^e(at)\}_{t\in[0,\infty),H\in(0,1)},$$

where $\overset{f.d.d.}{=}$ indicates the equality of finite dimensional distributions.

Given a fBm its odd and even part are independent processes (indexed by t). However, this is not the case with the dfBf $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$. The fields $\{B_H^o(t)\}_{t\in[0,\infty),H\in(0,1)}$ and $\{B_H^e(t)\}_{t\in[0,\infty),H\in(0,1)}$ are not independent. For example if $H + H^{'} = 1$ then

$$\begin{aligned}\mathbb{E}B_H^e(t)\,B_{H'}^o(s) &= \frac{1}{\pi}\sqrt{\Gamma(2H+1)\sin(\pi H)} \\ &\times\sqrt{\Gamma(2H'+1)\sin(\pi H')}\sin\left(\left(H^{'}-H\right)\frac{\pi}{2}\right) \\ &\times\left\{s\log|s| - \frac{(t+s)\log|t+s| - (t-s)\log|t-s|}{2}\right\}.\end{aligned}$$

which is clearly not equal to 0. That is the reason for calling that field the dependent fractional Brownian field.

Another glance at the computation of the covariance $\mathbb{E}B_H^i(t)\,B_{H'}^j(s)\,,\ i,j \in \{o,e\}$, reveals that when $i=j$ the g part of (33) cancels out while in the case when $i \neq j$ the first part of (33) cancels out leaving the g part. Therefore the existence the g part in (33) is the reason for dependence between $\{B_H^o(t)\}_{t\in\mathbb{R},H\in(0,1)}$ and $\{B_H^e(t)\}_{t\in\mathbb{R},H\in(0,1)}$. So it is natural to search for a method of creating a fractional Brownian field that would be of the form (33) with $g=0$. One way of attacking that problem is a direct verification of positive definiteness of such a form, a very unattractive task. In the next section we present a straightforward construction of the field with the desired covariance.

4. Fractional Brownian field

The last remark of the previous points the direction for constructing a fractional Brownian field $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ with the covariance of the form of (33) with $g=0$. The new Gaussian field contains all fractional Brownian motions too.

Theorem 7. *Let $B=\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ and $W=\{W_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ be two dfBf generated by two independent Brownian motions $\{B_t\}_{t\in\mathbb{R}}$ and $\{W_t\}_{t\in\mathbb{R}}$ respectively. Let $\left\{B_H^i(t)\right\}_{t\in[0,\infty),H\in(0,1)}, i=o$ be the odd and $i=e$ be the even part of B, and let $\left\{W_H^i(t)\right\}_{t\in[0,\infty),H\in(0,1)}$, $i=o$ be the odd and $i=e$ the even part of W. Then the fractional Brownian field $\{Z_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ defined by*

$$Z_H(t)=\begin{cases} B_H^e(t)+W_H^o(t) & \text{for } t\ge 0\\ B_H^e(-t)-W_H^o(-t) & \text{for } t<0\end{cases}\tag{34}$$

has the covariance

$$\mathbb{E}Z_H(t)\,Z_{H'}(s)=a_{H,H'}\left\{\frac{|t|^{H+H'}+|s|^{H+H'}-|t-s|^{H+H'}}{2}\right\},\tag{35}$$

where $a_{H,H'}$ is given by equations (29) *and* (30).

Proof. The proof follows from (32), (31) and independence of $\{B_t\}_{t\in\mathbb{R}}$ and $\{W_t\}_{t\in\mathbb{R}}$. □

We will call the process $\{Z_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ fractional Brownian field (fBf in short). Note that for any $t\in\mathbb{R}$,

$$Z_H(t)=\frac{1}{2}\left(B_H(t)+W_H(t)+B_H(-t)-W_H(-t)\right),$$

and that $\left\{\frac{B_H(t)+W_H(t)}{\sqrt{2}}\right\}_{t\in\mathbb{R},}$ and $\left\{\frac{B_H(t)-W_H(t)}{\sqrt{2}}\right\}_{t\in\mathbb{R},H\in(0,1)}$ are two independent dfBf. Consequently

$$Z_H(t)=\frac{X_H(t)+Y_H(-t)}{\sqrt{2}}\tag{36}$$

where $\{X_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ and $\{Y_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ are two independent fractional Brownian fields, that is $\{Z_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ is a properly symmetrized dfBf.

Proposition 8. *Let $\{B_t\}_{t\in\mathbb{R}}$ and $\{W_t\}_{t\in\mathbb{R}}$ be two independent Brownian motion processes on the real line and let $\{Z_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ be defined by* (34)*. Then*

$$Z_H(t) = \frac{1}{\sqrt{2}c_H}\int_{-\infty}^{\infty}(t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}}\,dB_x$$
$$+\frac{1}{\sqrt{2}c_H}\int_{-\infty}^{\infty}(-t-x)_+^{H-\frac{1}{2}} - (-x)_+^{H-\frac{1}{2}}\,dW_x.$$

Proof. Follows directly from (36) and (5). □

A fBf has the same self-similarity property in the time variable as the odd and even fractional Browian field, namely for $a>0$

$$\{Z_H(at)\}_{t\in\mathbb{R},H\in(0,1)} \overset{f.d.d.}{=} \left\{a^H Z_H(at)\right\}_{t\in\mathbb{R},H\in(0,1)}.$$

Moreover, the stationary in the time variable of increments of the fBf easily follows from Theorem 7, that is

$$\{W_H(t)\}_{t\in\mathbb{R},H\in(0,1)} \overset{f.d.d.}{=} \{W_H(t+\delta) - W_H(\delta)\}_{t\in\mathbb{R},H\in(0,1)}$$

for any δ.

An immediate consequence of the covariance structure of a fractional Brownian field is that when $H+H'=1$ then

$$\mathbb{E}\left(Z_H(t)\,Z_{H'}(t)\right) = a_{H,H'}\begin{cases}|s|\wedge|t| & \text{for } 0\le s,t \text{ or } 0\ge s,t\\ 0 & \text{otherwise}\end{cases}. \tag{37}$$

That property leads to a construction of martingales associated to fractional Brownian motions. The methodology of the construction is the subject of the next section.

5. Duality and fundamental martingales

In what follows it is assumed that $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ is an fBf. Whenever $H+H'=1$ we will call $B_H, B_{H'}$ (or $B_H^o, B_{H'}^o$ or $B_H^e, B_{H'}^e$) a dual pair. Dual pairs have unique properties. They generate martingales associated in a natural way to fractional Brownian motions B_H and $B_{H'}$. The construction and explanation of the nature of those martingales is the subject of this section.

Every fBf is a sum of an even and an odd part of two independent dfBf's. For that reason it suffices to construct martingales, M_H^o and M_H^e, adapted to the filtrations of the odd and even part of $\{B_H(t)\}_{t\in\mathbb{R}}$ respectively. The filtration generated by M_H^o (M_H^e) coincides with the filtration of the odd (even) part of fBm, and for that reason, following the terminology used in Norros et al. [16] to describe a martingale for the fBm originally discovered by Molchan and Golosov (see Molchan [14]), we call M_H^o (M_H^e) a fundamental martingale for the odd (even) part of fBm. Furthermore we derive a stochastic integral representation for those martingales. In a similar fashion this was done in Pipiras and Taqqu [18] and Pipiras and Taqqu [19] for the fractional Brownian motion.

For $i\in\{o,e\}$ set

$$\mathcal{F}_t^{H,i} = \sigma\left(B_H^i(s): 0\le s\le t\right) \text{ and } G_t^{H,i} = \overline{span}\left(B_H^i(s): 0\le s\le t\right),$$

where $\{B_H^o(t)\}_{t\ge 0}$ and $\{B_H^e(t)\}_{t\ge 0}$ are the odd and even part of $\{B_H(t)\}_{t\in\mathbb{R}}$.

For $t \geq 0$, $i \in \{o, e\}$ and $H + H^{'} = 1$ define

$$M_H^i(t) = \mathbb{E}\left(B_{H'}^i(t) \mid \mathcal{F}_t^{H,i}\right). \tag{38}$$

Theorem 9. *$\{M_H^o(t)\}_{t\geq 0}$ and $\{M_H^e(t)\}_{t\geq 0}$ are H-self-similar Gaussian martingales adapted to the filtration $\{\mathcal{F}_t^{H,o}\}_{t\geq 0}$ and $\{\mathcal{F}_t^{H,e}\}_{t\geq 0}$respectively.*

Proof. It is enough to verify the statement for M_H^o only, because the verification for M_H^e is similar. By construction M_H^o is a Gaussian process. It follows from (31) that for $s \leq t$,

$$B_{H'}^o(t) - B_{H'}^o(s) \perp G_s^{H,o},$$

which implies

$$\begin{aligned}
\mathbb{E}\left(M_H^o(t) \mid \mathcal{F}_s^{H,o}\right) &= \mathbb{E}\left(\mathbb{E}\left(B_{H'}^o(t) \mid \mathcal{F}_t^{H,o}\right) \mid \mathcal{F}_s^{H,o}\right) = \mathbb{E}\left(B_{H'}^o(t) \mid \mathcal{F}_s^{H,o}\right) \\
&= \mathbb{E}\left(B_{H'}^o(t) - B_{H'}^o(s) \mid \mathcal{F}_s^{H,o}\right) + \mathbb{E}\left(B_{H'}^o(s) \mid \mathcal{F}_s^{H,o}\right) \\
&= 0 + M_H^o(s) = M_H^o(s).
\end{aligned}$$

By H-self-similarity property of the odd part $\{B_H^o(t)\}_{t\in[0,\infty),H\in(0,1)}$ of the fBf $\{B_H(t)\}_{t\in\mathbb{R},H\in(0,1)}$ the field $\{Z_H^o(t)\}_{t\in[0,\infty),H\in(0,1)}$ defined by

$$Z_H^o(t) = a^{-H} B_H^o(at)$$

is an odd fBf and, therefore for $H + H^{'} = 1$,

$$\left\{\mathbb{E}\left(Z_{H'}^o(t) \mid \sigma\left(Z_H^o(r) : 0 \leq r \leq t\right)\right)\right\}_{t\geq 0} \overset{f.d.d.}{=} \{M_H^o(t)\}_{t\geq 0}.$$

Furthermore

$$\begin{aligned}
\sigma\left(Z_H^o(r) : 0 \leq r \leq t\right) &= \sigma\left(a^{-H} B_H^o(ar) : 0 \leq r \leq t\right) = \sigma\left(B_H^o(ar) : 0 \leq r \leq t\right) \\
&= \sigma\left(B_H^o(s) : 0 \leq s \leq at\right) = \mathcal{F}_{at}^{H,o}.
\end{aligned}$$

Therefore

$$\mathbb{E}\left(Z_{H'}^o(t) \mid \sigma\left(Z_H^o(r) : 0 \leq r \leq t\right)\right) = \mathbb{E}\left(a^{-H} B_{H'}^o(at) \mid \mathcal{F}_{at}^{H,o}\right),$$

which concludes the proof. □

So far we have shown that M_H^o and M_H^e are H-self-similar Gaussian martingales. By construction, for $i \in \{o, e\}$, $M_H^i(t)$ is an element of $G_t^{H,i}$, and therefore it may be possible to express it as a stochastic integral, up to time t, of B_H^i. In the case $H = \frac{1}{2}$ this is trivial, since then $H^{'} = \frac{1}{2}$ and therefore $B_H^i = B_{H'}^i$ is a constant multiple of Brownian motion and $M_H^i(t) = B_H^i(t)$. The case $H \neq \frac{1}{2}$ has been solved in our paper [4]. We state the result below without proof. The supporting materials are too long for the present paper. It should also be mentioned that the natural filtration of the martingale M_H^i coincides with the natural filtration of the process B_H^i [4].

Theorem 10. *Let $H \in (0,1) \setminus \{\frac{1}{2}\}$. If $t \geq 0$ then*

$$M_H^o(t) = \frac{\sqrt{\pi}\alpha_H}{\Gamma(1-H)} \int_0^t \left(t^2 - s^2\right)_+^{\frac{1}{2}-H} dB_H^o(s)$$

and

$$M_H^e(t) = -\frac{\alpha_H}{\Gamma\left(\frac{3}{2}-H\right)} \int_0^t \frac{d}{ds}\left(\int_s^t \left(x^2-s^2\right)^{\frac{1}{2}-H} dx\right) dB_H^e(s),$$

where

$$\alpha_H = \frac{2^{2H-1}\sqrt{\Gamma(3-2H)}\sin(\pi H)}{\Gamma\left(\frac{3}{2}-H\right)\sqrt{\Gamma(2H+1)}}.$$

6. Remarks on multifractional Brownian motions

Lévy Véhel and Peltier [23] and Benassi et al. [2] have introduced independently, multifractional Brownian motion. In this section we will clarify the relationship between multifractional Brownian motion and the fractional Brownian fields introduced in Sections 3 and 4. Additionally we will point out an error in the covariance of multifractional Brownian motion obtained from the nonanticipating moving average representation of fBm which shows that in fact the processes of Lévy Véhel [23] and Peltier and Benassi et al. [2] are not the same, as it has been claimed in Cohen [3].

Let $\{W_s\}_{t\in\mathbb{R}}$ be a Brownian motion. For $t \geq 0$ Lévy Véhel and Peltier [23] called

$$X_t = \frac{1}{\left(H(t)+\frac{1}{2}\right)} \int_{\mathbb{R}} \left((t-s)_+^{H(t)-\frac{1}{2}} - (-s)_+^{H(t)-\frac{1}{2}}\right) dW_s, \tag{39}$$

where $H : [0,\infty) \to (0,1)$ is a deterministic Hölder function with exponent $\beta > 0$, a multifractional Brownian motion. This process is introduced as a generalization of fBm that has different regularity at each t, more precisely, if $0 < H(t) < \min(1,\beta)$ then at each t_0 the multifractional Brownian motion has Hölder exponent $H(t_0)$ with probability 1. It is clear that if $H(t) \equiv H$ for some $0 < H < 1$, then $\{X_t\}_{t\geq 0}$ is a (nonstandard) H-fBm.

Benassi et al. [2] have introduced the process

$$Y_t = \int_{\mathbb{R}} \frac{e^{it\xi}-1}{|\xi|^{\frac{1}{2}+H(t)}} d\widehat{\widehat{W}}_\xi, \tag{40}$$

where in "some sense" the random measure $d\widehat{\widehat{W}}$ is the Fourier transform of dW and for $g, h \in L^2(\mathbb{R})$ it satisfies

$$\mathbb{E}\left(\int_{-\infty}^{\infty} g(\xi)\, d\widehat{\widehat{W}}_\xi \overline{\int_{-\infty}^{\infty} h(\xi)\, d\widehat{\widehat{W}}_\xi}\right) = \int_{-\infty}^{\infty} g(\xi)\, \overline{h(\xi)} d\xi$$

(see section 7.2.2 in Samorodnitsky and Taqqu [21]). If $H(t) \equiv H$, for some $0 < H < 1$, the process $\{Y_t\}_{t\geq 0}$ is an H-fBm, because the right-hand-side of (40) reduces to the well-known harmonizable representation of fBm. The result concerning the Hölder exponent for $\{X_t\}_{t\geq 0}$ holds for the process $\{Y_t\}_{t\geq 0}$ too. Although the process $\{Y_t\}_{t\geq 0}$ is called multifractional Brownian motion, we will refer to it as a harmonizable multifractional Brownian motion to emphasize the differences between the two processes.

In [3] Cohen states that both multifractional Brownian motions $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ if normalized appropriately are versions of the same process. More precisely the following is stated in [3] (as Theorem 1):

The harmonizable representation of the multifractional Brownian motion:

$$\int_{\mathbb{R}} \frac{e^{it\xi}-1}{|\xi|^{\frac{1}{2}+H(t)}} d\widehat{\widehat{W}}_{\xi}, \tag{41}$$

is almost surely equal up to a multiplicative deterministic function to the well balanced moving average

$$\int_{\mathbb{R}} \left(|t-s|^{H(t)-\frac{1}{2}} - |s|^{H(t)-\frac{1}{2}} \right) dW_s.$$

When $H(t) = \frac{1}{2}$, $\left(|t-s|^{H(t)-\frac{1}{2}} - |s|^{H(t)-\frac{1}{2}} \right)$ *is ambiguous, hence the conventional meaning*

$$|t-s|^0 - |s|^0 = \log\left(\frac{1}{|t-s|} \right) - \log\left(\frac{1}{|s|} \right)$$

is to be used. Conversely, one can show that the non anticipating moving average

$$\int_{\mathbb{R}} \left((t-s)_+^{H(t)-\frac{1}{2}} - (s)_+^{H(t)-\frac{1}{2}} \right) dW_s \tag{42}$$

is equal up to a multiplicative deterministic function to the harmonizable representation

$$\int_{\mathbb{R}} \frac{e^{it\xi}-1}{i\xi\, |\xi|^{H(t)-\frac{1}{2}}} d\widehat{\widehat{W}}_{\xi}.$$

Hence the mfBm given by the non anticipating moving average (42) *has the same law as the mfBm given by the harmonizable representation* (41) *up to a multiplicative deterministic function.*

The arguments used in the proof of the of Theorem 1 in Cohen [3] are based on the fact that the Fourier transform of $x \to |t-x|^{H(t)-\frac{1}{2}} - |x|^{H(t)-\frac{1}{2}}$ for $H(t) \neq \frac{1}{2}$ and of $\log(\frac{1}{|t-x|}) - \log(\frac{1}{|x|})$ for $H(t) = \frac{1}{2}$ equal, up to a multiplicative constant, to $\xi \to \frac{e^{it\xi}-1}{|\xi|^{\frac{1}{2}+H(t)}}$; and an incorrect statement that the Fourier transform of $x \to (t-x)_+^{H(t)-\frac{1}{2}} - (-x)_+^{H(t)-\frac{1}{2}}$ equals, up to a multiplicative constant, to $\xi \to \frac{e^{it\xi}-1}{i\xi|\xi|^{H(t)-\frac{1}{2}}}$. The equations (16) and (18) are the correct expression for that Fourier transform. Consequently the last two statements of the above Theorem 1 in Cohen are incorrect. In order to see why, consider two multifractional Brownian motions

$$X_t = \frac{1}{c_{H(t)}} \int_{-\infty}^{\infty} (t-x)_+^{H(t)-\frac{1}{2}} - (-x)_+^{H(t)-\frac{1}{2}} \, dB_x$$

and

$$Y_t = \begin{cases} \frac{1}{d_{H(t)}} \int_{-\infty}^{\infty} |t-x|^{H(t)-\frac{1}{2}} - |x|^{H(t)-\frac{1}{2}} \, dW_x \text{ for } t\, H(t) \neq \frac{1}{2} \\ \frac{1}{d_{H(t)}} \int_{-\infty}^{\infty} \log\left(\frac{1}{|t-x|}\right) - \log\left(\frac{1}{|x|}\right) dW_x \qquad H(t) = \frac{1}{2} \end{cases},$$

where $t \geq 0$, $c_{H(t)}$ and $d_{H(t)}$ are defined by (6), (8) respectively, $d_{\frac{1}{2}} = \pi$, and $\{B_t\}_{t\in\mathbb{R}}$ and $\{W_t\}_{t\in\mathbb{R}}$ are Brownian motions. According to the last statement of Theorem 1 in Cohen [3] there is a deterministic function f_t such that the processes $\{X_t\}_{t\geq 0}$ and $\{f_t Y_t\}_{t\geq 0}$ have the same law. The chosen normalization assures that $\mathbb{E}\left(X_t^2\right) = \mathbb{E}\left(Y_t^2\right)$ for all $t \geq 0$, implying that $|f_t| = 1$ for all t such that $t > 0$. It follows that $|\mathbb{E}(X_t X_s)| = |\mathbb{E}(Y_t Y_s)|$ for all s, t. It is clear that the process $\{X_t\}_{t\geq 0}$

can be obtained from a dependent fractional Brownian field $\{B_H(t)\}_{t\geq 0, H\in(0,1)}$ as $\{B_{H(t)}(t)\}_{t\geq 0}$. Similarly, the Gaussian field $\{W_H(t)\}_{t\geq 0, H\in(0,1)}$ defined for $H \neq \frac{1}{2}$ by equation (7) and for $H = \frac{1}{2}$ by (9) gives $\{Y_t\}_{t\geq 0}$ via $\{W_{H(t)}(t)\}_{t\geq 0}$. Since the last statement of Theorem 1 in [3] is supposed to hold for every Hölder function $t \to H(t)$, that statement holds if and only if the Gaussian fields $\{B_H(t)\}_{t\in[0,\infty), H\in(0,1)}$ and $\{W_H(t)\}_{t\in[0,\infty), H\in(0,1)}$ have the same covariance in absolute value. The covariance of $\{B_H(t)\}_{t\in[0,\infty), H\in(0,1)}$ is given by Theorem 2. Proposition 11 right below gives the covariance of $\{W_H(t)\}_{t\in[0,\infty), H\in(0,1)}$. The proposition shows that if $H \neq \frac{1}{2}$ and $H + H' = 1$ the covariance $\mathbb{E}W_H(t)W_{H'}(s)$ is a multiple of $s\wedge t$. From Theorem 2 we see that this is not the case for $\mathbb{E}B_H(t)B_{H'}(s)$.

Proposition 11. *The covariance of the Gaussian field $\{W_H(t)\}_{t\in\mathbb{R}, H\in(0,1)}$, defined for $H \neq \frac{1}{2}$ by equation (7) and for $H = \frac{1}{2}$ by (9) is*

$$\mathbb{E}W_H(t)W_{H'}(s) = \frac{k_H k_{H'}}{d_H d_{H'}} \frac{d^2_{\frac{H+H'}{2}}}{k^2_{\frac{H+H'}{2}}} \left(\frac{|t|^{H+H'} + |s|^{H+H'} - |t-s|^{H+H'}}{2} \right),$$

where d_H is defined by (8) and

$$k_H = -2\Gamma\left(H + \frac{1}{2}\right)\sin\left(\left(H - \frac{1}{2}\right)\frac{\pi}{2}\right) \text{ for } H \neq \frac{1}{2} \text{ and } k_{\frac{1}{2}} = \pi. \tag{43}$$

Before proving Proposition 11, a few technical results are needed.

Lemma 12. *Let $f_{t,\frac{1}{2}}(x) = \log\frac{1}{|t-x|} - \log\frac{1}{|x|}$ for $x, t \in \mathbb{R}$. Then*

$$\widehat{f_{t,\frac{1}{2}}}(\xi) = \pi\frac{e^{it\xi} - 1}{\xi}$$

Proof. Suppose that $\xi \neq 0$. Since $f_{t,\frac{1}{2}} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ the Fourier transform of $f_{t,\frac{1}{2}}$ can be computed as

$$\widehat{f_{t,\frac{1}{2}}}(\xi) = \int_{-\infty}^{\infty}\left(\log\frac{1}{|t-x|} - \log\frac{1}{|x|}\right)e^{ix\xi}dx = \int_{-\infty}^{\infty}\log\left(\frac{|x|}{|x-t|}\right)e^{ix\xi}dx.$$

Substituting $u = x - \frac{t}{2}$ yields

$$\widehat{f_{t,\frac{1}{2}}}(\xi) = e^{i\frac{t}{2}\xi}\int_{-\infty}^{\infty}\log\left(\frac{\left|u + \frac{t}{2}\right|}{\left|u - \frac{t}{2}\right|}\right)e^{iu\xi}du,$$

and after the substitution $u = -v$ on $(-\infty, 0)$ to

$$\widehat{f_{t,\frac{1}{2}}}(\xi) = e^{i\frac{t}{2}\xi}2i\int_0^{\infty}\log\left(\frac{\left|u + \frac{t}{2}\right|}{\left|u - \frac{t}{2}\right|}\right)\sin(u\xi)\,du.$$

Formula 4.382 (1) in [7] reads:

$$\int_0^{\infty}\log\left(\frac{|u+a|}{|u-a|}\right)\sin(bx)\,dx = \frac{\pi}{b}\sin(ab) \text{ for } a, b > 0.$$

If $t > 0$ set $a = \frac{t}{2}$ and $b = |\xi|$ and use that $\sin(z\xi) = sgn(\xi)\sin(z|\xi|)$ for $z \geq 0$ to get

$$\widehat{f_{t,\frac{1}{2}}}(\xi) = e^{i\frac{t}{2}\xi}2i\frac{\pi}{|\xi|}\sin\left(\frac{t\xi}{2}\right)$$

Standard trigonometric identities $\sin(2a) = 2\sin(a)\cos(a)$ and $\sin^2(a) = \frac{1-\cos(2a)}{2}$ complete the proof for $t > 0$. The proof for $t < 0$ is similar. $\square$

Lemma 13. *Let* $f_{t,H}(x) = |t-x|^{H-\frac{1}{2}} - |x|^{H-\frac{1}{2}}$ *for* $x, t \in \mathbb{R}$ *and* $H \in (0,1) \setminus \left\{\frac{1}{2}\right\}$. *Then*

$$\widehat{f_{t,H}}(\xi) = -2\Gamma\left(H+\frac{1}{2}\right)\sin\left(\left(H-\frac{1}{2}\right)\frac{\pi}{2}\right)\frac{e^{it\xi}-1}{|\xi|^{H+\frac{1}{2}}}.$$

Proof. Rewriting $f_{t,H}$ as

$$f_{t,H}(x) = (t-x)_+^{H-\frac{1}{2}} - (x)_+^{H-\frac{1}{2}} + (t-x)_-^{H-\frac{1}{2}} - (x)_-^{H-\frac{1}{2}},$$

and applying equations (16) and (17) it follows that

$$\widehat{f_{t,H}}(\xi) = \Gamma\left(H+\frac{1}{2}\right)\frac{e^{it\xi}-1}{i\xi}(i\xi)^{-\left(H-\frac{1}{2}\right)} - \Gamma\left(H+\frac{1}{2}\right)\frac{e^{it\xi}-1}{i\xi}(-i\xi)^{-\left(H-\frac{1}{2}\right)}.$$

and from (18) that

$$(i\xi)^{-\left(H-\frac{1}{2}\right)} - (-i\xi)^{-\left(H-\frac{1}{2}\right)} = 2i|\xi|^{-\left(H-\frac{1}{2}\right)} sgn(-\xi)\sin\left(\left(H-\frac{1}{2}\right)\frac{\pi}{2}\right).$$

Therefore

$$\begin{aligned}\widehat{f_{t,H}}(\xi) &= 2\Gamma\left(H+\frac{1}{2}\right)\frac{e^{it\xi}-1}{i\xi} i|\xi|^{-\left(H-\frac{1}{2}\right)} sgn(-\xi)\sin\left(\left(H-\frac{1}{2}\right)\frac{\pi}{2}\right)\\ &= -2\Gamma\left(H+\frac{1}{2}\right)\sin\left(\left(H-\frac{1}{2}\right)\frac{\pi}{2}\right)\frac{e^{it\xi}-1}{|\xi|^{H+\frac{1}{2}}}.\end{aligned}$$

$\square$

We are now ready to prove Proposition 11. The idea is to use the fact that up to a multiplicative constant $\widehat{f_{t,H}}(\xi)\overline{\widehat{f_{s,H'}}(\xi)}$ equals $\widehat{f_{t,\frac{H+H'}{2}}}(\xi)\overline{\widehat{f_{s,\frac{H+H'}{2}}}(\xi)}$ and that up to a multiplicative constant the integral over $\mathbb{R}$ of the later is the covariance of an $\frac{H+H'}{2}$-fBm. This argument is used in Ayache et al. [1] to compute the covariance of a multifractional Brownian motion given by (40). In Ayache et al. [1] it is also erroneously claimed that the covariance of the multifractional Brownian motion given by (39) is the same, if properly normalized, as the covariance of the harmonizable multifractional Brownian motion given by (40). Their proof is based on the last statement of Theorem 1 in Cohen [3]. In section 3 of Lim and Muniandy [11], the authors give another incorrect argument about the equivalence (up to a deterministic multiplicative function) between the harmonizable multifractional Brownian motion (40) and the nonanticipative multifractional Brownian motion (39).

Proof of Proposition 11. By the Plancherel identity

$$\begin{aligned}\mathbb{E}W_H(t)W_{H'}(s) &= \frac{1}{d_H d_{H'}}\int_{-\infty}^{\infty} f_{t,H}(x) f_{s,H'}(x)\,dx\\ &= \frac{1}{2\pi}\frac{1}{d_H d_{H'}}\int_{-\infty}^{\infty}\widehat{f_{t,H}}(\xi)\overline{\widehat{f_{s,H'}}(\xi)}d\xi,\end{aligned} \tag{44}$$

where $f_{t,H}$, when $H = \frac{1}{2}$, is the same as in Lemma 12, and, when $H \neq \frac{1}{2}$, as in and Lemma 13. From Lemma 12 and Lemma 13 it follows

$$\int_{-\infty}^{\infty} \widehat{f_{t,H}}(\xi)\, \overline{\widehat{f_{s,H'}}(\xi)} d\xi = k_H k_{H'} \int \frac{\left(e^{it\xi} - 1\right)}{|\xi|^{H+\frac{1}{2}}} \frac{\overline{\left(e^{is\xi} - 1\right)}}{|\xi|^{H'+\frac{1}{2}}} d\xi$$

Let $H_0 = \frac{H+H'}{2}$. Observe that $H_0 \in (0,1)$ and so

$$\begin{aligned}
\int_{-\infty}^{\infty} \widehat{f_{t,H}}(\xi)\, \overline{\widehat{f_{s,H'}}(\xi)} d\xi &= k_H k_{H'} \frac{d_{H_0}^2}{k_{H_0}^2} \frac{k_{H_0}^2}{d_{H_0}^2} \int \frac{\left(e^{it\xi} - 1\right)}{|\xi|^{H_0+\frac{1}{2}}} \frac{\overline{\left(e^{is\xi} - 1\right)}}{|\xi|^{H_0+\frac{1}{2}}} d\xi \\
&= k_H k_{H'} \frac{d_{H_0}^2}{k_{H_0}^2} \frac{2\pi}{d_{H_0}^2} \int_{-\infty}^{\infty} f_{t,H_0}(x)\, f_{s,H_0}(x)\, dx,
\end{aligned}$$

where the last equality follows from Lemma 12, Lemma 13 and Plancherel identity. Hence

$$\begin{aligned}
\int_{-\infty}^{\infty} \widehat{f_{t,H}}(\xi)\, \overline{\widehat{f_{s,H'}}(\xi)} d\xi &= k_H k_{H'} \frac{d_{H_0}^2}{k_{H_0}^2} 2\pi \mathbb{E} W_{H_0}(t)\, W_{H_0}(s) \\
&= k_H k_{H'} \frac{d_{H_0}^2}{k_{H_0}^2} 2\pi \left(\frac{|t|^{2H_0} + |s|^{2H_0} - |t-s|^{2H_0}}{2} \right).
\end{aligned}$$

Hence (44) becomes

$$\mathbb{E} W_H(t)\, W_{H'}(s) = \frac{k_H k_{H'}}{d_H d_{H'}} \frac{d_{H_0}^2}{k_{H_0}^2} \left(\frac{|t|^{2H_0} + |s|^{2H_0} - |t-s|^{2H_0}}{2} \right).$$

This finishes the proof. □

References

[1] Ayache, A., Cohen, S. and Lévy Véhel, J. (2000). The covariance structure of multifractional Brownian motion, with application to long range dependence. In *IEEE International Conference on Acoustics, Speech, and Signal Processing 2000*, volume 6, pp. 3810–3813.

[2] Benassi, A., Jaffard, S. and Roux, D. (1997). Elliptic Gaussian random processes. *Rev. Mat. Iberoamericana* **13** (1) 19–90. MR1462329

[3] Cohen, S. (1999). From self-similarity to local self-similarity: the estimation problem. In *Fractals: Theory and Applications in Engineering.* Springer, London, pp. 3–16. MR1726364

[4] Dobrić, V. and Ojeda, F. M. (2005). Even and odd fractional Brownian motions and its martingales. Manuscript.

[5] Dobrić, V. and Ojeda, F. M. (2005). Natural wavelet expansions for Gaussian Markov processes. Manuscript.

[6] Dzhaparidze, K. and van Zanten, H. (2004). A series expansion of fractional Brownian motion. *Probab. Theory Related Fields* **130** (1) 39–55. MR2092872

[7] Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products.* Academic Press, San Diego, CA. MR1773820

[8] Kolmogorov, A. N. (1940). Kurven im Hilbertschen Raum, die gegenüber einer einparametrigen Gruppe von Bewegungen invariant sind. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **26** 6–9. MR0003440

[9] Kolmogorov, A. N. (1940). Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **26** 115–118. MR0003441

[10] Kolmogorov, A. N. (1991). The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Proc. Roy. Soc. London Ser. A* **434** (1890) 9–13. MR1124922

[11] Lim, S. C. and Muniandy, S. V. (2000). On some possible generalizations of fractional Brownian motion. *Phys. Lett. A* **266** (2–3) 140–145. MR1741314

[12] Meyer, Y., Sellan, F. and Taqqu, M. S. (1999). Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion. *J. Fourier Anal. Appl.* **5** (5) 465–494. MR1755100

[13] Molchan, G. M. (1969). Gaussian processes with asymptotic power spectrum. Summaries of papers presented at the sessions of the probability and statistic section of the Moscow Mathematical Society (February–December 1968). *Theory of Probability and its Applications* **14** (3) 556–559.

[14] Molchan, G. M. (2003). Historical comments related to fractional Brownian motion. In *Theory and Applications of Long-Range Dependence.* Birkhäuser Boston, Boston, MA, pp. 39–42. MR1956043

[15] Molchan, G. M. and Golosov, J. I. (1969). Gaussian stationary processes with asymptotically a power spectrum. *Dokl. Akad. Nauk SSSR* **184** 546–549. MR0242247

[16] Norros, I., Valkeila, E. and Virtamo, J. (1999). An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. *Bernoulli* **5** (4) 571–587. MR1704556

[17] Nuzman, C. J. and Poor, H. V. (2001). Reproducing kernel Hilbert space methods for wide-sense self-similar processes. *Ann. Appl. Probab.* **11** (4) 1199–1219. MR1878295

[18] Pipiras, V. and Taqqu, M. S. (2000). Integration questions related to fractional Brownian motion. *Probab. Theory Related Fields* **118** (2) 251–291. MR1790083

[19] Pipiras, V. and Taqqu, M. S. (2001). Are classes of deterministic integrands for fractional Brownian motion on an interval complete? *Bernoulli* **7** (6) 873–897. MR1873833

[20] Pipiras, V. and Taqqu, M. S. (2002). Deconvolution of fractional Brownian motion. *J. Time Ser. Anal.* **23** (4) 487–501. MR1910894

[21] Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes. Stochastic Modeling.* Chapman & Hall, New York. MR1280932

[22] Taqqu, M. S. (2003). Fractional Brownian motion and long-range dependence. In *Theory and Applications of Long-Range Dependence.* Birkhäuser Boston, Boston, MA, pp. 5–38. MR1956042

[23] Lévy Véhel, J. and Peltier, R. F. (1995). Multifractional Brownian motion: definition and preliminary results. Technical Report 2645, Institut National de Recherche en Informatique et an Automatique, INRIA, Le Chesnay, France.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 96–116

DOI: 10.1214/074921706000000789

Risk bounds for the non-parametric estimation of Lévy processes

José E. Figueroa-López[1] **and Christian Houdré**[2]

Purdue University and Georgia Institute of Technology

Abstract: Estimation methods for the Lévy density of a Lévy process are developed under mild qualitative assumptions. A classical model selection approach made up of two steps is studied. The first step consists in the selection of a good estimator, from an approximating (finite-dimensional) linear model $\mathcal{S}$ for the true Lévy density. The second is a data-driven selection of a linear model $\mathcal{S}$, among a given collection $\{\mathcal{S}_m\}_{m\in\mathcal{M}}$, that approximately realizes the best trade-off between the error of estimation within $\mathcal{S}$ and the error incurred when approximating the true Lévy density by the linear model $\mathcal{S}$. Using recent concentration inequalities for functionals of Poisson integrals, a bound for the risk of estimation is obtained. As a byproduct, oracle inequalities and long-run asymptotics for spline estimators are derived. Even though the resulting underlying statistics are based on continuous time observations of the process, approximations based on high-frequency discrete-data can be easily devised.

1. Introduction

Lévy processes are central to the classical theory of stochastic processes, not only as discontinuous generalizations of Brownian motion, but also as prototypical Markov processes and semimartingales (see [27] and [5] for monographs on these topics). In recent years, continuous-time models driven by Lévy processes have received a great deal of attention mainly because of their applications in the area of mathematical finance (see e.g. [14] and references therein). The scope of these models goes from simple exponential Lévy models (e.g. [2, 10, 12] and [16]), where the underlying source of randomness in the Black-Scholes model is replaced by a Lévy process, to exponential time-changed Lévy processes (e.g. [11]-[13]) and to stochastic differential equations driven by multivariate Lévy processes (e.g. [3, 29]). Exponential Lévy models have proved successful to account for several empirical features observed in time series of financial returns such as heavy tails, high-kurtosis, and asymmetry (see, for example, [10] and [16]). Lévy processes, as models capturing the most basic features of returns and as "first-order approximations" to other more accurate models, should be considered first in developing and testing a successful statistical methodology. However, even in such parsimonious models, there are several issues in performing statistical inference by standard likelihood-based methods.

Lévy processes are determined by three "parameters": a non-negative real σ^2, a real μ, and a measure ν on $\mathbb{R}\backslash\{0\}$. These three parameters characterize a Lévy process $\{X(t)\}_{t\geq 0}$ as the superposition of a Brownian motion with drift, $\sigma B(t)+\mu t$,

[1]Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA, e-mail: jfiguero@math.purdue.edu

[2]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA, e-mail: houdre@math.gatech.edu

AMS 2000 subject classifications: primary 62G05, 60G51; secondary 62P05, 60E07.

Keywords and phrases: estimation of Lévy processes, estimation of Poisson processes, projection estimation, model selection, oracle inequalities, concentration inequalities.

and an independent pure-jump Lévy process, whose jump behavior is specified by the measure ν in that for any $A \in \mathcal{B}(\mathbb{R})$, whose indicator χ_A vanishes in a neighborhood of the origin,

$$\nu(A) = \frac{1}{t}\mathbb{E}\left[\sum_{s\leq t} \chi_{A}\left(\Delta X(s)\right)\right],$$

for any $t > 0$ (see Section 19 of [27]). Here, $\Delta X(t) \equiv X(t) - X(t^{-})$ denotes the jump of X at time t. Thus, $\nu(A)$ gives the average number of jumps (per unit time) whose magnitudes fall in the set A. A common assumption in Lévy-based financial models is that ν is determined by a function $p : \mathbb{R}\backslash\{0\} \to [0,\infty)$, called the *Lévy density*, as follows

$$\nu(A) = \int_A p(x)dx, \quad \forall A \in \mathcal{B}(\mathbb{R}\backslash\{0\}).$$

Intuitively, the value of p at x_0 provides information on the frequency of jumps with sizes "close" to x_0.

Estimating the Lévy density poses a nontrivial problem, even when p takes simple parametric forms. Parsimonious Lévy densities usually produce not only intractable marginal densities, but sometimes marginal densities which are not even expressible in a closed form. The current practice of estimation relies on numerical approximations of the density function of $X(t)$ using *inversion formulas* combined with maximum likelihood estimation (see for instance [10]). Such approximations make the estimation computationally expensive and particularly susceptible to numerical errors and mis-specifications. Even in the case of closed form marginal densities, maximum-likelihood based methods present serious numerical problems. For instance, analyzing generalized hyperbolic Lévy processes, the author of [24] notices that the likelihood function is highly flat for a wide range of parameters and good starting values as well as convergence are critical. Also, the separation of parameters and identification between different subclasses is difficult. These issues worsen when dealing with "high-frequency" data. Other calibration methods include methods based on moments, simulation based methods, and multinomial log likelihoods (see e.g. [29] and [6] and references therein). However, our goal in the present paper is not to match the precision of some of these parametric methods, but rather gain in robustness and efficiency using *non-parametric methods*. That is to say, assuming only qualitative information on the Lévy density, we develop estimation schemes for the Lévy density p that provide fairly general function estimators $\hat{p}$.

We follow the so-called *model selection methodology* developed in the context of density estimation in [8], and recently extended to the estimation of intensity functions for Poisson processes in [25]. The essence of this approach is to approximate an infinite-dimensional, nonparametric model by a sequence of finite-dimensional models. This strategy has its origins in *Grenander's method of sieves* (see [17]). Concretely, the procedure addresses two problems. First, the selection of a good estimator $\hat{p}_{\mathcal{S}}$, called the *projection estimator*, out of an approximating (finite-dimensional) linear model $\mathcal{S}$ for the Lévy density. Second, the selection of a linear model $\mathcal{S}_{\hat{m}}$, among a given collection of linear models $\{\mathcal{S}_m\}_m$, that approximately realizes the best trade-off between the error of estimation from the first step, and the error incurred when approximating the unknown Lévy density by the linear model $\mathcal{S}$. The technique used in the second step has the general flavor of cross-validation via a penalization term, leading to *penalized projection estimators* $\tilde{p}$ (p.p.e.).

Comparing our approach to other non-parametric methods for non-homogeneous Poisson processes (see e.g. [20, 21] and [25]), we will see that the main difficulty here is the fact that the jump process associated with a Lévy process has potentially infinitely many small jumps. To overcome this problem, we introduce a reference measure and estimate instead the Lévy density with respect to this measure. In contrast to [25], our treatment does not rely on the finiteness of such a reference measure. Our main objective here is to estimate the order of magnitude of the mean-square error, $\mathbb{E}\,\|p-\tilde{p}\|^2$, between the true Lévy density and the p.p.e. To accomplish this, we apply concentration inequalities for functionals of Poisson point processes such as functions of stochastic Poisson integrals (see e.g. [9, 18, 25]). This important statistical application of concentration inequalities is well-known in other contexts such as regression and density estimation (see [8] and references therein). The bound for the risk of estimation leads in turn to *oracle inequalities* implying that the p.p.e. is at least as good (in terms of the long term rate of convergence) as the best projection estimator (see Section 4 for details). Also, combining the bound with results on the approximation of smooth functions by *sieves*, one can determine the long-term rate of convergence of the p.p.e. on certain well-known approximating spaces of functions such as splines.

The statistics underlying our estimators are expressed in terms of deterministic functions of the jumps of the process, and thus, they are intrinsically based on a continuous-time observation of the process during some time period $[0,T]$. Even though this observation scheme has an obvious drawback, statistical analysis under it presents a lot of interest for two reasons. First, very powerful theoretical results can be obtained, thus providing benchmarks of what can be achieved by discrete-data-based statistical methods. Second, since the path of the process can in principle be approximated by high-frequency sampling, it is possible to construct feasible estimators by approximating the continuous-time based statistics using discrete-observations. We use this last idea to obtain estimators by replacing the jumps by increments, based on equally spaced observations of the process.

Let us describe the outline of the paper. We develop the model selection approach in Sections 2 and 3. We proceed to obtain in Section 4 bounds for the risk of estimation, and consequently prove *oracle inequalities*. In Section 5 the rate of convergence of the p.p.e. on *regular splines*, when the Lévy density belongs to some *Lipschitz spaces* or *Besov spaces* of smooth functions, are derived. In Section 6, implementation of the method using discrete-time sampling of the process is briefly discussed. We finish with proofs of the main results.

2. A non-parametric estimation method

Consider a real Lévy process $X=\{X(t)\}_{t\geq 0}$ with Lévy density $p:\mathbb{R}_0\to\mathbb{R}_+$, where $\mathbb{R}_0\equiv\mathbb{R}\backslash\{0\}$. Then, X is a càdlàg process with independent and stationary increments such that the characteristic function of its marginals is given by

$$\mathbb{E}\left[e^{iuX(t)}\right]=\exp\left\{t\left(iub-\frac{u^2\sigma^2}{2}+\int_{\mathbb{R}_0}\left\{e^{iux}-1-iux1_{[|x|\leq 1]}\right\}p(x)dx\right)\right\}, \tag{2.1}$$

with $p:\mathbb{R}_0\to\mathbb{R}_+$ such that

$$\int_{\mathbb{R}_0}(1\wedge x^2)p(x)dx<\infty. \tag{2.2}$$

Since X is a càdlàg process, the set of its jump times

$$\left\{t>0:\Delta X(t)\equiv X(t)-X(t^-)\neq 0\right\}$$

is countable. Moreover, for Borel subsets B of $[0,\infty)\times\mathbb{R}_0$,

$$\mathcal{J}(B) \equiv \#\left\{t>0 : (t, X(t)-X(t^-)) \in B\right\}, \tag{2.3}$$

is a well-defined random measure on $[0,\infty)\times\mathbb{R}_0$, where $\#$ denotes cardinality. The Lévy-Itô decomposition of the sample paths (see Theorem 19.2 of [27]) implies that $\mathcal{J}$ is a Poisson process on the Borel sets $\mathcal{B}([0,\infty)\times\mathbb{R}_0)$ with mean measure

$$\mu(B) = \iint_B p(x)\,dt\,dx. \tag{2.4}$$

Recall also that the stochastic integral of a deterministic function $f:\mathbb{R}_0\to\mathbb{R}$ with respect to $\mathcal{J}$ is defined by

$$I(f) \equiv \iint_{[0,T]\times\mathbb{R}_0} f(x)\mathcal{J}(dt,dx) = \sum_{t\le T} f(\Delta X(t)), \tag{2.5}$$

where this last expression is well defined if

$$\int_0^T\int_{\mathbb{R}_0} |f(x)|\mu(dt,dx) = T\int_{\mathbb{R}_0}|f(x)|p(x)dx < \infty;$$

see e.g. Chapter 10 in [19].

We consider the problem of estimating the Lévy density p on a Borel set $D\in\mathcal{B}(\mathbb{R}_0)$ using a *projection estimation approach.* According to this paradigm, p is estimated by estimating its best approximating function in a finite-dimensional linear space $\mathcal{S}$. The linear space $\mathcal{S}$ is taken so that it has good approximation properties for general classes of functions. Typical choices are piecewise polynomials or wavelets. Throughout, we make the following standing assumption.

Assumption 1. The Lévy measure $\nu(dx)\equiv p(x)dx$ is absolutely continuous with respect to a known measure η on $\mathcal{B}(D)$ so that the Radon-Nikodym derivative

$$\frac{d\nu}{d\eta}(x) = s(x), \quad x\in D, \tag{2.6}$$

is positive, bounded, and satisfies

$$\int_D s^2(x)\eta(dx) < \infty. \tag{2.7}$$

In that case, s is called the Lévy density, on D, of the process with respect to the reference measure η.

Remark 2.1. Under the previous assumption, the measure $\mathcal{J}$ of (2.3), when restricted to $\mathcal{B}([0,\infty)\times D)$, is a Poisson process with mean measure

$$\mu(B) = \iint_B s(x)\,dt\,\eta(dx), \quad B\in\mathcal{B}([0,\infty)\times D). \tag{2.8}$$

Our goal will be to estimate the Lévy density s, which itself could in turn be used to retrieve p on D via (2.6). To illustrate this strategy consider a continuous Lévy density p such that

$$p(x) = O\left(x^{-1}\right), \text{ as } x\to 0.$$

This type of densities satisfies the above assumption with respect to the measure $\eta(dx) = x^{-2}dx$ on domains of the form $D = \{x : 0 < |x| < b\}$. Clearly, an estimator $\hat{p}$ for the Lévy density p can be generated from an estimator $\hat{s}$ for s by fixing $\hat{p}(x) \equiv x^{-2}\hat{s}(x)$.

Let us now describe the main ingredients of our approach. Let $\mathcal{S}$ be a finite dimensional subspace of $L^2 \equiv L^2\left((D,\eta)\right)$ equipped with the standard norm

$$\|f\|_{\eta}^2 \equiv \int_D f^2(x)\,\eta(dx).$$

The space $\mathcal{S}$ plays the role of an approximating linear model for the Lévy density s. Of course, under the L^2 norm, the best approximation of s on $\mathcal{S}$ is the orthogonal projection defined by

$$s^{\perp}(x) \equiv \sum_{i=1}^{d} \left(\int_D \varphi_i(y) s(y) \eta(dy) \right) \varphi_i(x), \tag{2.9}$$

where $\{\varphi_1, \dots, \varphi_d\}$ is an arbitrary orthonormal basis of $\mathcal{S}$. The *projection estimator* of s on S is defined by

$$\hat{s}(x) \equiv \sum_{i=1}^{d} \hat{\beta}_i \varphi_i(x), \tag{2.10}$$

where we fix

$$\hat{\beta}_i \equiv \frac{1}{T} \iint\limits_{[0,T]\times D} \varphi_i(x) \mathcal{J}(dt, dx). \tag{2.11}$$

This is the most natural unbiased estimator for the *orthogonal projection* $s^{\perp}$. Notice also that $\hat{s}$ is independent of the specific orthonormal basis of $\mathcal{S}$. Indeed, the projection estimator is the unique solution to the minimization problem

$$\min_{f \in \mathcal{S}} \gamma_D(f),$$

where $\gamma_D : L^2\left((D,\eta)\right) \to \mathbb{R}$ is given by

$$\gamma_D(f) \equiv -\frac{2}{T} \iint\limits_{[0,T]\times D} f(x)\, \mathcal{J}(dt, dx) + \int\limits_D f^2(x)\, \eta(dx). \tag{2.12}$$

In the literature on model selection (see e.g. [7] and [25]), γ_D is the so-called *contrast function*. The previous characterization also provides a mechanism to numerically evaluate $\hat{s}$ when an orthonormal basis of $\mathcal{S}$ is not explicitly available.

The following proposition provides both the first-order and the second-order properties of $\hat{s}$. These follow directly from the well-known formulas for the mean and variance of Poisson integrals (see e.g. [19] Chapter 10).

Proposition 2.2. *Under Assumption 1, $\hat{s}$ is an unbiased estimator for $s^{\perp}$ and its "mean-square error", defined by*

$$\chi^2 \equiv \|\hat{s} - s^{\perp}\|_{\eta}^2 = \int \left(\hat{s}(x) - s^{\perp}(x)\right)^2 \eta(dx),$$

is such that

$$\mathbb{E}\left[\chi^2\right] = \frac{1}{T}\sum_{i=1}^{d}\int_D \varphi_i^2(x)s(x)\,\eta(dx). \tag{2.13}$$

The risk *of $\hat{s}$ admits the decomposition*

$$\mathbb{E}\left[\|s-\hat{s}\|_\eta^2\right] = \|s-s^\perp\|_\eta^2 + \mathbb{E}\left[\chi^2\right]. \tag{2.14}$$

The first term in (2.14), the *bias term*, accounts for the distance of the unknown function s to the model $\mathcal{S}$, while the second term, the *variance term*, measures the error of estimation within the linear model $\mathcal{S}$. Notice that (2.13) is finite because s is assumed bounded on D and thus,

$$\mathbb{E}\left[\chi^2\right] \le \|s\|_\infty \frac{d}{T}. \tag{2.15}$$

3. Model selection via penalized projection estimator

A crucial issue in the above approach is the selection of the approximating linear model $\mathcal{S}$. In principle, a "nice" density s can be approximated closely by general linear models such as splines or wavelet. However, a more robust model $\mathcal{S}'$ containing $\mathcal{S}$ will result in a better approximation of s, but with a larger variance. This raises the natural problem of selecting one model, out of a collection of linear models $\{\mathcal{S}_m, m \in \mathcal{M}\}$, that approximately realizes the best trade-off between the risk of estimation within the model and the distance of the unknown Lévy density to the approximating model.

Let $\hat{s}_m$ and $s_m^\perp$ be respectively the projection estimator and the orthogonal projection of s on $\mathcal{S}_m$. The following equation, readily derived from (2.14), gives insight on a sensible solution to the model selection problem:

$$\mathbb{E}\left[\|s-\hat{s}_m\|_\eta^2\right] = \|s\|_\eta^2 + \mathbb{E}\left[-\|\hat{s}_m\|_\eta^2 + \mathrm{pen}(m)\right]. \tag{3.1}$$

Here, $\mathrm{pen}(m)$ is defined in terms of an orthonormal basis $\{\varphi_{1,m}, \dots, \varphi_{d_m,m}\}$ of $\mathcal{S}_m$ by the equation:

$$\mathrm{pen}(m) = \frac{2}{T^2}\iint\limits_{[0,T]\times D}\left(\sum_{i=1}^{d_m}\varphi_{i,m}^2(x)\right)\mathcal{J}(dt,dx). \tag{3.2}$$

Equation (3.1) shows that the risk of $\hat{s}_m$ moves "parallel" to the expectation of the *observable statistics* $-\|\hat{s}_m\|_\eta^2 + \mathrm{pen}(m)$. This fact justifies to choose the model that minimizes such statistics. We will see later that other choices for $\mathrm{pen}(\cdot)$ also give good results. Therefore, given a penalization function $\mathrm{pen} : \mathcal{M} \to [0,\infty)$, we consider estimators of the form

$$\tilde{s} \equiv \hat{s}_{\hat{m}}, \tag{3.3}$$

where $\hat{s}_m$ is the projection estimator on $\mathcal{S}_m$ and

$$\hat{m} \equiv \mathrm{argmin}_{m\in\mathcal{M}}\left\{-\|\hat{s}_m\|_\eta^2 + \mathrm{pen}(m)\right\}.$$

An estimator $\tilde{s}$ as in (3.3) is called a **penalized projection estimator** (p.p.e.) on the collection of linear models $\{\mathcal{S}_m, m \in \mathcal{M}\}$.

Methods of estimation based on the minimization of penalty functions have a long history in the literature on regression and density estimation (for instance, [1, 22], and [28]). The general idea is to choose, among a given collection of parametric models, the model that minimizes a loss function plus a penalty term that controls the variance term, which will forcefully increase as the approximating linear models become more detailed. Such penalized estimation was promoted for nonparametric density estimation in [8], and in the context of non-homogeneous Poisson processes in [25].

4. Risk bound and oracle inequalities

The penalization idea of the previous section provides a sensible criterion to select an estimator $\tilde{s} \equiv \hat{s}_{\hat{m}}$ out of the projection estimators $\{\hat{s}_m : m \in \mathcal{M}\}$ induced by a given collection of approximating linear models $\{\mathcal{S}_m, m \in \mathcal{M}\}$. Ideally, one wishes to choose that projection estimator $\hat{s}_{m^*}$ that minimizes the risk; namely, such that

$$\mathbb{E}\left[\|s - \hat{s}_{m^*}\|_\eta^2\right] \leq \mathbb{E}\left[\|s - \hat{s}_m\|_\eta^2\right], \quad \text{for all } m \in \mathcal{M}. \tag{4.1}$$

Of course, to pick the best $\hat{s}_m$ is not feasible since s is not available to actually compute and compare the risks. But, how bad would the risk of $\tilde{s}$ be compared to the best possible risk that can be achieved by projection estimators? One can aspire to achieve the smallest possible risk up to a constant. In other words, it is desirable that our estimator $\tilde{s}$ comply with an inequality of the form

$$\mathbb{E}\left[\|s - \tilde{s}\|_\eta^2\right] \leq C \inf_{m \in \mathcal{M}} \mathbb{E}\left[\|s - \hat{s}_m\|_\eta^2\right], \tag{4.2}$$

for a constant C independent of the linear models. The model $\mathcal{S}_{m^*}$ that achieves the minimal risk (using projection estimation) is the *oracle model* and inequalities of the type (4.2) are called *oracle inequalities*. Approximate oracle inequalities were proved in [25] for the intensity function of a nonhomogeneous Poisson process. In this section we show that for certain penalization functions, the resulting penalized projection estimator $\tilde{s}$ defined by (3.3) satisfies the inequality

$$\mathbb{E}\left[\|s - \tilde{s}\|_\eta^2\right] \leq C \inf_{m \in \mathcal{M}} \mathbb{E}\left[\|s - \hat{s}_m\|_\eta^2\right] + \frac{C'}{T}, \tag{4.3}$$

for some "model free" constants C, C' (remember that the time period of observations is $[0, T]$). The main tool in obtaining oracle inequalities is an upper bound for the risk of the penalized projection estimator $\tilde{s}$. The proof of (4.3) follows essentially from the arguments in [25]; however, to overcome the possible lack of finiteness on the reference measure η (see Assumption 1), which is required in [25], and to avoid superfluous rough upper bounds, the dimension of the linear model is explicitly included in the penalization and the arguments are refined.

Let us introduce some notation. Below, d_m denotes the dimension of the linear model $\mathcal{S}_m$, and $\{\varphi_{1,m}, \ldots, \varphi_{d_m,m}\}$ is an arbitrary orthonormal basis of $\mathcal{S}_m$. Define

$$D_m = \sup\left\{\|f\|_\infty^2 : f \in S_m, \|f\|_\eta^2 = 1\right\}, \tag{4.4}$$

which is assumed to be finite and can be proved to be equal to $\|\sum_{i=1}^{d_m} \varphi_{i,m}^2\|_\infty$.

We make the following regularity condition, introduced in [25], that essentially controls the complexity of the linear models. This assumption is satisfied by splines and trigonometric polynomials, but not by wavelet bases.

Assumption 2. There exist constants $\Gamma > 0$ and $R \geq 0$ such that for every positive integer n,

$$\#\{m \in \mathcal{M} : d_m = n\} \leq \Gamma n^R.$$

We now present our main result.

Theorem 4.1. *Let $\{\mathcal{S}_m, m \in \mathcal{M}\}$ be a family of finite dimensional linear subspaces of $L^2((D,\eta))$ satisfying Assumption 2 and such that $D_m < \infty$. Let $\mathcal{M}_T \equiv \{m \in \mathcal{M} : D_m \leq T\}$. If $\hat{s}_m$ and $s_m^\perp$ are respectively the projection estimator and the orthogonal projection of the Lévy density s on $\mathcal{S}_m$ then, the penalized projection estimator $\tilde{s}_T$ on $\{\mathcal{S}_m\}_{m\in\mathcal{M}_T}$ defined by* (3.3) *is such that*

$$\mathbb{E}\left[\|s - \tilde{s}_T\|_\eta^2\right] \leq C \inf_{m\in\mathcal{M}_T} \left\{\|s - s_m^\perp\|_\eta^2 + \mathbb{E}\left[\mathrm{pen}(m)\right]\right\} + \frac{C'}{T}, \tag{4.5}$$

whenever $\mathrm{pen} : \mathcal{M} \to [0,\infty)$ *takes either one of the following forms for some fixed (but arbitrary) constants $c > 1$, $c' > 0$, and $c'' > 0$:*
(a) $\mathrm{pen}(m) \geq c\frac{D_m \mathcal{N}}{T^2} + c'\frac{d_m}{T}$*, where $\mathcal{N} \equiv \mathcal{J}([0,T] \times D)$ is the number of jumps prior to T with sizes in D and where it is assumed that $\rho \equiv \int_D s(x)\eta(dx) < \infty$;*
(b) $\mathrm{pen}(m) \geq c\frac{\hat{V}_m}{T}$*, where $\hat{V}_m$ is defined by*

$$\hat{V}_m \equiv \frac{1}{T} \iint\limits_{[0,T]\times D} \left(\sum_{i=1}^{d_m} \varphi_{i,m}^2(x)\right) \mathcal{J}(dt, dx), \tag{4.6}$$

and where it is assumed that $\beta \equiv \inf_{m\in\mathcal{M}} \frac{\mathbb{E}[\hat{V}_m]}{D_m} > 0$ and that $\phi \equiv \inf_{m\in\mathcal{M}} \frac{D_m}{d_m} > 0$;
(c) $\mathrm{pen}(m) \geq c\frac{\hat{V}_m}{T} + c'\frac{D_m}{T} + c''\frac{d_m}{T}$.

In (4.5)*, the constant C depends only on c, c' and c'', while C' varies with c, c', c'', Γ, R, $\|s\|_\eta$, $\|s\|_\infty$, ρ, β, and ϕ.*

Remark 4.2. It can be shown that if $c \geq 2$, then for arbitrary $\varepsilon > 0$, there is a constant $C'(\varepsilon)$ (increasing as $\varepsilon \downarrow 0$) such that

$$\mathbb{E}\|s - \tilde{s}\|_\eta^2 \leq (1+\varepsilon) \inf_{m\in\mathcal{M}} \left\{\|s - s_m^\perp\|_\eta^2 + \mathbb{E}\left[\mathrm{pen}(m)\right]\right\} + \frac{C'(\varepsilon)}{T}. \tag{4.7}$$

One important consequence of the risk bound (4.5) is the following oracle inequality:

Corollary 4.3. *In the setting of Theorem* 4.1(b)*, if the penalty function is of the form* $\mathrm{pen}(m) \equiv c\frac{\hat{V}_m}{T}$*, for every $m \in \mathcal{M}_T$, $\beta > 0$, and $\phi > 0$, then*

$$\mathbb{E}\left[\|s - \tilde{s}_T\|_\eta^2\right] \leq \widetilde{C} \inf_{m\in\mathcal{M}_T} \left\{\mathbb{E}\left[\|s - \hat{s}_m\|_\eta^2\right]\right\} + \frac{\widetilde{C}'}{T}, \tag{4.8}$$

for a constant $\widetilde{C}$ depending only on c, and a constant $\widetilde{C}'$ depending on c, Γ, R, $\|s\|_\eta$, $\|s\|_\infty$, β, and ϕ.

5. Rate of convergence for smooth Lévy densities

We use the risk bound of the previous section to study the "long run" ($T \to \infty$) rate of convergence of penalized projection estimators based on regular piecewise polynomials, when the Lévy density is "smooth". More precisely, on a window of estimation $D \equiv [a,b] \subset \mathbb{R}_0$, the Lévy density of the process with respect to the Lebesgue measure $\eta(dx) \equiv dx$, denoted by s, is assumed to belong to the *Besov space* (also called *Lipschitz* space) $\mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$ for some $p \in [2,\infty]$ and $\alpha > 0$ (see for instance [15] and references therein for background on these spaces). Concretely, $\mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$ consists of those functions $f \in L^p([a,b],dx)$ if $0 < p < \infty$ (or f continuous if $p = \infty$) such that

$$|f|_{\mathcal{B}^{\alpha}_{\infty}(L^p)} \equiv \sup_{\delta>0} \frac{1}{\delta^{\alpha}} \sup_{0<h\leq\delta} \|\Delta^r_h(f,\cdot)\|_{L^p([a,b],dx)} < \infty,$$

where $\Delta_h(f,x) \equiv f(x+h) - f(x)$ and $\Delta^r_h(f,x)$ is the r^{th}-order difference of f defined by

$$\Delta^r_h(f,x) \equiv \Delta_h(\Delta^{r-1}_h(f,\cdot),x),$$

for x such that $x + rh \in D$ and $r \in \mathbb{N}$. The following spaces are closely related. For $k \in \mathbb{N}$ and $\beta \in (0,1]$ such that $\alpha = k + \beta$, let $\mathrm{Lip}(\alpha, L^p([a,b]))$ be the class of functions f such that $f, \dots, f^{(k-1)}$ are absolutely continuous on $[a,b]$ with $f^{(k)} \in L^p((a,b))$ satisfying

$$\|\Delta_h(f^{(k)},\cdot)\|_{L^p([a,b],dx)} \leq M h^{\beta},$$

for some $M < \infty$. It is know that if $\alpha > 0$ is not an integer and $1 \leq p \leq \infty$, then $f \in \mathrm{Lip}(\alpha, L^p([a,b]))$ if and only if f is a.e. equal to a function in $\mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$. In general, $\mathrm{Lip}(\alpha, L^p([a,b])) \subset \mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$, for any $0 < p \leq \infty$ and $\alpha > 0$ (see e.g. [15]).

An important reason for the choice of the Besov class of smooth functions is the availability of estimates for the error of approximation by splines, trigonometric polynomials, and wavelets (see e.g. [15] and [4]). In particular, if $\mathcal{S}^k_m$ denotes the space of piecewise polynomials of degree at most k, based on the regular partition of $[a,b]$ with m intervals ($m \geq 1$), and $s \in \mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$ with $k > \alpha - 1$, then there exists a constant $C(s)$ such that

$$\mathrm{d}_p\left(s, \mathcal{S}^k_m\right) \leq C(s) m^{-\alpha}, \tag{5.1}$$

where d_p is the distance induced by the L^p-norm on $([a,b],dx)$ (see [15]). The following gives the rate of convergence of the p.p.e. on regular splines.

Corollary 5.1. *With the notation of Theorem 4.1, taking $D = [a,b]$ and $\eta(dx) = dx$, let $\tilde{s}_T$ be the penalized projection estimator on $\{\mathcal{S}^k_m\}_{m \in \mathcal{M}_T}$ with penalization*

$$\mathrm{pen}(m) \equiv c\frac{\hat{V}_m}{T} + c'\frac{D_m}{T} + c''\frac{d_m}{T},$$

for some fixed $c > 1$ and $c', c'' > 0$. Then, if the restriction to D of the Lévy density s belongs to $\mathcal{B}^{\alpha}_{\infty}(L^p([a,b]))$, with $2 \leq p \leq \infty$ and $0 < \alpha < k+1$, then

$$\limsup_{T\to\infty} T^{2\alpha/(2\alpha+1)} \mathbb{E}\left[\|s - \tilde{s}_T\|^2_{\eta}\right] < \infty.$$

Moreover, for any $R > 0$ and $L > 0$,

$$\limsup_{T\to\infty} T^{2\alpha/(2\alpha+1)} \sup_{s\in\Theta(R,L)} \mathbb{E}\left[\|s - \tilde{s}_T\|^2_{\eta}\right] < \infty, \tag{5.2}$$

where $\Theta(R, L)$ *consists of all the Lévy densities* f *such that* $\|f\|_{L^\infty([a,b],dx)} < R$, *and such that the restriction of* f *to* $[a, b]$ *is a member of* $\mathcal{B}^\alpha_\infty\left(L^p([a, b])\right)$ *with* $|f|_{\mathcal{B}^\alpha_\infty(L^p)} < L$.

The previous result implies that the p.p.e. on regular splines has a rate of convergence of order $T^{-2\alpha/(2\alpha+1)}$ for the class of Besov Lévy densities $\Theta(R, L)$.

6. Estimation based on discrete time data

Let us finish with some remarks on how to approximate the continuous-time statistics of our methods using only discrete-time observations. In practice, we can aspire to sample the process $X(t)$ at discrete times, but we are neither able to measure the size of the jumps $\Delta X(t) \equiv X(t) - X(t^-)$ nor the times of the jumps $\{t : \Delta X(t) > 0\}$. In general, Poisson integrals of the type

$$I(f) \equiv \iint\limits_{[0,T]\times\mathbb{R}_0} f(x)\mathcal{J}(dt, dx) = \sum_{t\leq T} f(\Delta X(t)), \tag{6.1}$$

are not accessible. Intuitively, the following statistic is the most natural approximation to (6.1):

$$I_n(f) \equiv \sum_{k=1}^{n} f(\Delta_k X), \tag{6.2}$$

where $\Delta_k X$ is the k^{th} increment of the process with time span $h_n \equiv T/n$; that is,

$$\Delta_k X \equiv X(kh_n) - X((k-1)h_n), \quad k = 1, \ldots, n.$$

How good is this approximation and in what sense? Under some conditions on f, we can readily prove the weak convergence of (6.2) to (6.1) using properties of the transition distributions of X in small time (see [5], Corollary 8.9 of [27], and [26]). The following theorem summarizes some known results on the small-time transition distribution.

Theorem 6.1. *Let* $X = \{X(t)\}_{t\geq 0}$ *be a Lévy process with Lévy measure* ν. *The following statements hold true.*

(1) *For each* $a > 0$,

$$\lim_{t\to 0} \frac{1}{t}\mathbb{P}(X(t) > a) = \nu([a, \infty)), \tag{6.3}$$

$$\lim_{t\to 0} \frac{1}{t}\mathbb{P}(X(t) \leq -a) = \nu((-\infty, -a]). \tag{6.4}$$

(2) *For any continuous bounded function* h *vanishing in a neighborhood of the origin,*

$$\lim_{t\to 0} \frac{1}{t}\mathbb{E}[h(X(t))] = \int_{\mathbb{R}_0} h(x)\nu(dx). \tag{6.5}$$

(3) *If* h *is continuous and bounded and if* $\lim_{|x|\to 0} h(x)|x|^{-2} = 0$, *then*

$$\lim_{t\to 0} \frac{1}{t}\mathbb{E}[h(X(t))] = \int_{\mathbb{R}_0} h(x)\nu(dx).$$

Moreover, if $\int_{\mathbb{R}_0}(|x| \wedge 1)\nu(dx) < \infty$, *it suffices to have* $h(x)(|x| \wedge 1)^{-1}$ *continuous and bounded.*

Convergence results like (6.5) are useful to establish the convergence in distribution of $\mathrm{I}_n(f)$ since

$$\mathbb{E}\left[e^{\mathrm{i}uI_n(f)}\right] = \left(\mathbb{E}\left[e^{\mathrm{i}uf\left(X\left(\frac{T}{n}\right)\right)}\right]\right)^n = \left(1 + \frac{a_n}{n}\right)^n,$$

where $a_n = n\mathbb{E}\left[h\left(X\left(\frac{T}{n}\right)\right)\right]$ with $h(x) = e^{\mathrm{i}uf(x)} - 1$. So, if f is such that

$$\lim_{t\to 0}\frac{1}{t}\mathbb{E}\left[e^{\mathrm{i}uf(X(t))} - 1\right] = \int_{\mathbb{R}_0}\left(e^{\mathrm{i}uf(x)} - 1\right)\nu(dx), \tag{6.6}$$

then a_n converges to $a \equiv T\int_{\mathbb{R}_0} h(x)\nu(dx)$, and thus

$$\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = \lim_{n\to\infty} e^{n\log\left(1+\frac{a_n}{n}\right)} = e^a.$$

We thus have the following result.

Proposition 6.2. *Let $X = \{X(t)\}_{t\geq 0}$ be a Lévy process with Lévy measure ν. Then,*

$$\lim_{n\to\infty}\mathbb{E}\left[e^{\mathrm{i}uI_n(f)}\right] = \exp\left\{T\int_{\mathbb{R}_0}\left(e^{\mathrm{i}uf(x)} - 1\right)\nu(dx)\right\},$$

if f satisfies either one of the following conditions:

(1) *$f(x) = \mathbf{1}_{(a,b]}(x)h(x)$ for an interval $[a,b] \subset \mathbb{R}_0$ and a continuous function h;*
(2) *f is continuous on $\mathbb{R}_0$ and $\lim_{|x|\to 0} f(x)|x|^{-2} = 0$.*

In particular, $I_n(f)$ converges in distribution to $I(f)$ under any one of the previous two conditions.

Remark 6.3. Notice that if (6.5) holds true when replacing h by f and f^2, then the mean and variance of $I_n(f)$ obey the asymptotics:

$$\lim_{n\to\infty}\mathbb{E}\left[I_n(f)\right] = T\int_{\mathbb{R}_0} f(x)\nu(dx);$$

$$\lim_{n\to\infty}\mathrm{Var}\left[I_n(f)\right] = T\int_{\mathbb{R}_0} f^2(x)\nu(dx).$$

Remark 6.4. Very recently, [23] proposed a procedure to disentangle the jumps from the diffusion part in the case of jump-diffusion models driven by finite-jump activity Lévy processes. It is proved there that for certain functions $r : \mathbb{R}_+ \to \mathbb{R}_+$, there exists $N(\omega)$ such that for $n \geq N(\omega)$, a jump occurs in the interval $((k-1)h_n, kh_n]$ if and only if $(\Delta_k X)^2 > r(h_n)$. Here, $h_n = T/n$ and $\Delta_k X$ is the k^{th} increment of the process. These results suggest to use statistics of the form

$$\sum_{k=1}^{n} f\left(\Delta_k X\right)\mathbf{1}\left[\left(\Delta_k X\right)^2 > r(h_n)\right]$$

instead of (6.2) to approximate the integral (6.1).

7. Proofs

7.1. Proof of the risk bound

We break the proof of Theorem 4.1 into several preliminary results.

Lemma 7.1. *For any penalty function* pen : $\mathcal{M} \to [0, \infty)$ *and any* $m \in \mathcal{M}$, *the penalized projection estimator* $\tilde{s}$ *satisfies*

$$\|s - \tilde{s}\|_\eta^2 \le \|s - s_m^\perp\|_\eta^2 + 2\chi_{\hat{m}}^2 + 2\nu_D\left(s_{\hat{m}}^\perp - s_m^\perp\right) + \text{pen}(m) - \text{pen}(\hat{m}), \tag{7.1}$$

where $\chi_m^2 \equiv \|s_m^\perp - \hat{s}_m\|_\eta^2$ *and where the functional* $\nu_D : L^2((D, \eta)) \to \mathbb{R}$ *is defined by*

$$\nu_D(f) \equiv \iint\limits_{[0,T]\times D} f(x)\, \frac{\mathcal{J}(dt, dx) - s(x)\, dt\, \eta(dx)}{T}. \tag{7.2}$$

The general idea in obtaining (4.5) is to bound the "inaccessible" terms on the right hand side of (7.1) (namely $\chi_{\hat{m}}^2$ and $\nu_D\left(s_{\hat{m}}^\perp - s_m^\perp\right)$) by observable statistics. In fact, the penalizations pen($\cdot$) given in Theorem 4.1 are chosen so that the right hand side in (7.1) does not involve $\hat{m}$. To carry out this plan, we use concentration inequalities for $\chi_{\hat{m}}^2$ and for the compensated Poisson integrals $\nu_D(f)$. The following result gives a concentration inequality for general compensated Poisson integrals.

Proposition 7.2. *Let N be a Poisson process on a measurable space* $(\mathrm{V}, \mathcal{V})$ *with mean measure μ and let* $f : \mathrm{V} \to \mathbb{R}$ *be an essentially bounded measurable function satisfying* $0 < \|f\|_\mu^2 \equiv \int_{\mathrm{V}} f^2(v)\mu(dv)$ *and* $\int_{\mathrm{V}} |f(v)|\mu(dv) < \infty$. *Then, for any* $u > 0$,

$$\mathbb{P}\left[\int_{\mathrm{V}} f(v)(N(dv) - \mu(dv)) \ge \|f\|_\mu \sqrt{2u} + \frac{1}{3}\|f\|_\infty u\right] \le e^{-u}. \tag{7.3}$$

In particular, if $f : \mathrm{V} \to [0, \infty)$ *then, for any* $\epsilon > 0$ *and* $u > 0$,

$$\mathbb{P}\left[(1+\varepsilon)\left(\int_{\mathrm{V}} f(v)N(dv) + \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right)\|f\|_\infty u\right) \ge \int_{\mathrm{V}} f(v)\mu(dv)\right] \ge 1 - e^{-u}. \tag{7.4}$$

For a proof of the inequality (7.3), see [25] (Proposition 7) or [18] (Corollary 5.1). Inequality (7.4) is a direct consequence of (7.3) (see Section 7.2 for a proof).

The next result allows us to bound the Poisson functional χ_m^2. This result is essentially Proposition 9 of [25].

Lemma 7.3. *Let N be a Poisson process on a measurable space* $(\mathrm{V}, \mathcal{V})$ *with mean measure* $\mu(dv) = p(v)\zeta(dv)$ *and intensity function* $p \in L^2(\mathrm{V}, \mathcal{V}, \zeta)$. *Let $\mathcal{S}$ be a finite dimensional subspace of* $L^2(\mathrm{V}, \mathcal{V}, \zeta)$ *with orthonormal basis* $\{\tilde{\varphi}_1, \ldots, \tilde{\varphi}_d\}$, *and let*

$$\hat{p}(v) \equiv \sum_{i=1}^d \left(\int_{\mathrm{V}} \tilde{\varphi}_i(w)N(dw)\right)\tilde{\varphi}_i(v) \tag{7.5}$$

$$p^\perp(v) \equiv \sum_{i=1}^d \left(\int_{\mathrm{V}} p(w)\tilde{\varphi}_i(w)\eta(dw)\right)\tilde{\varphi}_i(v). \tag{7.6}$$

Then, $\chi^2(\mathcal{S}) \equiv \|\hat{p} - p^\perp\|_\zeta^2$ *is such that for any* $u > 0$ *and* $\varepsilon > 0$

$$\mathbb{P}\left[\chi(\mathcal{S}) \ge (1+\varepsilon)\sqrt{\mathbb{E}\left[\chi^2(\mathcal{S})\right]} + \sqrt{2kM_{\mathcal{S}}u} + k(\varepsilon)B_{\mathcal{S}}u\right] \le e^{-u}, \tag{7.7}$$

where we can take $k = 6$, $k(\varepsilon) = 1.25 + 32/\varepsilon$, and where

$$M_{\mathcal{S}} \equiv \sup\left\{\int_{\mathrm{V}} f^2(v)p(v)\zeta(dv) : f \in \mathcal{S}, \|f\|_{\zeta} = 1\right\}, \tag{7.8}$$

$$B_{\mathcal{S}} \equiv \sup\left\{\|f\|_{\infty} : f \in \mathcal{S}, \|f\|_{\zeta} = 1\right\}. \tag{7.9}$$

Following the same strategy as in [25], the idea is to obtain from the previous lemmas a concentration inequality of the form

$$\mathbb{P}\left[\|s - \tilde{s}\|_{\eta}^2 \le C\left(\|s - s_m^{\perp}\|_{\eta}^2 + \mathrm{pen}(m)\right) + h(\xi)\right] \ge 1 - C'e^{-\xi},$$

for constants C and C', and a function $h(\xi)$ (all independent of m). This will prove to be enough in view of the following elementary result (see Section 7.2 for a proof).

Lemma 7.4. *Let $h : [0, \infty) \to \mathbb{R}_+$ be a strictly increasing function with continuous derivative such that $h(0) = 0$ and $\lim_{\xi\to\infty} e^{-\xi}h(\xi) = 0$. If Z is random variable satisfying*

$$\mathbb{P}\left[Z \ge h(\xi)\right] \le Ke^{-\xi},$$

for every $\xi > 0$, then

$$\mathbb{E}Z \le K\int_0^{\infty} e^{-u}h(u)du.$$

We are now in a position to prove Theorem 4.1. Throughout the proof, we will have to introduce various constants and inequalities that will hold with high probability. In order to clarify the role that the constants play in these inequalities, we shall make some convention and give to the letters x, y, f, a, b, ξ, $\mathcal{K}$, c, and C, with various sub- or superscripts, special meaning. The letters with x are reserved to denote positive constants that can be chosen arbitrarily. The letters with y denote arbitrary constants greater than 1. $f, f_1, f_2, \ldots$ denote quadratic polynomials of the variable ξ whose coefficients (denoted by $a's$ and $b's$) are determined by the values of the $x's$ and $y's$. The inequalities will be true with probabilities greater that $1 - \mathcal{K}e^{-\xi}$, where $\mathcal{K}$ is determined by the values of the $x's$ and the $y's$. Finally, $c's$ and $C's$ are used to denote constants constrained by the $x's$ and $y's$. *It is important to remember that the constants in a given inequality are meant only for that inequality.* The pair of equivalent inequalities below will be repeatedly invoked throughout the proof:

$$\begin{array}{l}\text{(i)}\ \ 2ab \le xa^2 + \frac{1}{x}b^2, \quad \text{and} \\ \text{(ii)}\ \ (a+b)^2 \le (1+x)\,a^2 + \left(1 + \frac{1}{x}\right)b^2, \quad (\text{for } x > 0).\end{array} \tag{7.10}$$

Also, for simplicity, we write below $\|\cdot\|$ to denote the L^2-norm with respect to the reference measure η.

Proof of Theorem 4.1. We consider successive improvements of the inequality (7.1):

Inequality 1. *For any positive constants x_1, x_2, x_3, and x_4, there exist a positive number $\mathcal{K}$ and an increasing quadratic function f (both independent of the family of linear models and of T) such that, with probability larger than $1 - \mathcal{K}e^{-\xi}$,*

$$\begin{aligned}\|s - \tilde{s}\|^2 \le\ & \|s - s_m^{\perp}\|^2 + 2\chi_{\hat{m}}^2 + 2x_1\|s_{\hat{m}}^{\perp} - s_m^{\perp}\|^2 \\ & + x_2\frac{D_{\hat{m}}}{T} + x_3\frac{D_m}{T} + x_4\frac{d_{\hat{m}}}{T} \\ & + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) + \frac{f(\xi)}{T}.\end{aligned} \tag{7.11}$$

Proof. Let us find an upper bound for $\nu_D(s_{m'}^\perp - s_m^\perp)$, $m', m \in \mathcal{M}$. Since the operator ν_D defined by (7.2) is just a compensated integral with respect to a Poisson process with mean measure $\mu(dtdx) = dt\eta(dx)$, we can apply Proposition 7.2 to obtain that, for any $x'_{m'} > 0$, and with probability larger than $1 - e^{-x'_{m'}}$

$$\nu_D\left(s_{m'}^\perp - s_m^\perp\right) \le \left\|\frac{s_{m'}^\perp - s_m^\perp}{T}\right\|_\mu \sqrt{2x'_{m'}} + \frac{\|s_{m'}^\perp - s_m^\perp\|_\infty x'_{m'}}{3T}. \tag{7.12}$$

In that case, the probability that (7.12) holds for every $m' \in \mathcal{M}$ is larger than $1 - \sum_{m' \in \mathcal{M}} e^{-x_{m'}}$ because $P(A \cap B) \ge 1 - a - b$, whenever $P(A) \ge 1 - a$ and $P(B) \ge 1 - b$. Clearly,

$$\begin{aligned}\left\|\frac{s_{m'}^\perp - s_m^\perp}{T}\right\|_\mu^2 &= \iint\limits_{[0,T]\times D} \left(\frac{s_{m'}^\perp(x) - s_m^\perp(x)}{T}\right)^2 s(x)dt\eta(dx)\\ &\le \|s\|_\infty \frac{\|s_{m'}^\perp - s_m^\perp\|^2}{T}.\end{aligned}$$

Using (7.10)(i), the first term on the right hand side of (7.12) is then bounded as follows:

$$\left\|\frac{s_{m'}^\perp - s_m^\perp}{T}\right\|_\mu \sqrt{2x'_{m'}} \le x_1\|s_{m'}^\perp - s_m^\perp\|^2 + \frac{\|s\|_\infty x'_{m'}}{2Tx_1}, \tag{7.13}$$

for any $x_1 > 0$. Using (4.4) and (7.10-i),

$$\begin{aligned}\|s_{m'}^\perp - s_m^\perp\|_\infty x'_{m'} &\le \left(\|s_{m'}^\perp\|_\infty + \|s_m^\perp\|_\infty\right) x'_{m'}\\ &\le \left(\sqrt{D_{m'}}\|s_{m'}^\perp\| + \sqrt{D_m}\|s_m^\perp\|\right) x'_{m'}\\ &\le \sqrt{D_{m'}}\|s\|x'_{m'} + \sqrt{D_m}\|s\|x'_{m'}\\ &\le 3x_2D_{m'} + 3x_3D_m + \frac{\|s\|^2 {x'}_{m'}^2}{12}\left(\frac{1}{x_2} + \frac{1}{x_3}\right),\end{aligned}$$

for all $x_2 > 0$, $x_3 > 0$. It follows that, for any $x_1 > 0$, $x_2 > 0$, and $x_3 > 0$,

$$\begin{aligned}\nu_D\left(s_{m'}^\perp - s_m^\perp\right) \le x_1\|s_{m'}^\perp - s_m^\perp\|^2 + x_2\frac{D_{m'}}{T} + x_3\frac{D_m}{T}\\ + \frac{\|s\|_\infty x'_{m'}}{2Tx_1} + \frac{\|s\|^2 {x'}_{m'}^2}{36T\bar{x}},\end{aligned}$$

where we set $\frac{1}{\bar{x}} = \frac{1}{x_2} + \frac{1}{x_3}$. Next, take

$$x'_{m'} \equiv x_4\sqrt{d_{m'}}\left(\frac{1}{\|s\|} \wedge \frac{1}{\|s\|_\infty}\right) + \xi.$$

Then, for any positive x_1, x_2, x_3, and x_4, there is a $\mathcal{K}$ and a function f such that, with probability greater than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{aligned}\nu_D\left(s_{m'}^\perp - s_m^\perp\right) \le x_1\|s_{m'}^\perp - s_m^\perp\|^2 + x_2\frac{D_{m'}}{T} + x_3\frac{D_m}{T}\\ + \left(\frac{x_4^2}{18\bar{x}} + \frac{x_4}{2x_1}\right)\frac{d_{m'}}{T} + \frac{f(\xi)}{T}, \quad \forall m' \in \mathcal{M}.\end{aligned} \tag{7.14}$$

Concretely,

$$
\begin{aligned}
f(\xi) &= \frac{\|s\|}{18\bar{x}}\xi^2 + \frac{\|s\|_\infty}{2x_1}\xi, \\
\mathcal{K} &= \Gamma \sum_{n=1}^{\infty} n^R \exp\left(-\sqrt{n}x_4\left(\frac{1}{\|s\|} \wedge \frac{1}{\|s\|_\infty}\right)\right).
\end{aligned} \tag{7.15}
$$

Here, we used the assumption of polynomial models (Definition 2) to come up with the constant $\mathcal{K}$. Plugging (7.14) in (7.1), and renaming the coefficient of $d_{m'}/T$, we can corroborate inequality 1. □

Inequality 2. *For any positive constants $y_1 > 1$, x_2, x_3, and x_4, there are positive constants $C_1 < 1$, $C_1' > 1$, and $\mathcal{K}$, and a strictly increasing quadratic polynomial f (all independent of the class of linear models and of T) such that with probability larger than $1 - \mathcal{K}e^{-\xi}$,*

$$
\begin{aligned}
C_1\|s - \tilde{s}\|^2 \le C_1'\|s - s_m^\perp\|^2 &+ y_1\chi^2_{\hat{m}} \\
&+ x_2\frac{D_{\hat{m}}}{T} + x_3\frac{D_m}{T} + x_4\frac{d_{\hat{m}}}{T} \\
&+ \operatorname{pen}(m) - \operatorname{pen}(\hat{m}) + \frac{f(\xi)}{T}.
\end{aligned} \tag{7.16}
$$

Moreover, if $1 < y_1 < 2$, then $C_1' = 3 - y_1$ and $C_1 = y_1 - 1$. If $y_1 \ge 2$, then $C_1' = 1 + 4x_1$ and $C_1 = 1 - 4x_1$, where x_1 is any positive constant related to f via to the equation (7.15).

Proof. Let us combine the term on the left hand side of (7.11) with the first three terms on the right hand side. Using the triangle inequality followed by (7.10-ii),

$$
\|s^\perp_{\hat{m}} - s^\perp_m\|^2 \le 2\|s - s^\perp_m\|^2 + 2\|s^\perp_{\hat{m}} - s\|^2.
$$

Then, since $\chi^2_{\hat{m}} = \|s^\perp_{\hat{m}} - \hat{s}_{\hat{m}}\|^2$, and $\|s^\perp_{\hat{m}} - s\|^2 = \|s - \hat{s}_{\hat{m}}\|^2 - \|s^\perp_{\hat{m}} - \hat{s}_{\hat{m}}\|^2$, it follows that

$$
\begin{aligned}
&\|s - s^\perp_m\|^2 + 2\chi^2_{\hat{m}} + 2x_1\|s^\perp_{\hat{m}} - s^\perp_m\|^2 - \|s - \tilde{s}\|^2 \\
&\quad \le (1 + 4x_1)\|s - s^\perp_m\|^2 + (2 - 4x_1)\|s^\perp_{\hat{m}} - \hat{s}_{\hat{m}}\|^2 \\
&\qquad + (4x_1 - 1)\|s - \tilde{s}\|^2,
\end{aligned}
$$

for every $x_1 > 0$. Then, for any $y_1 > 1$, there are positive constants $C > 0$, $C_1' > 1$, and $C_1 < 1$ such that

$$
\begin{aligned}
&\|s - s^\perp_m\|^2 + 2\chi^2_{\hat{m}} + 2C\|s^\perp_{\hat{m}} - s^\perp_m\|^2 - \|s - \tilde{s}\|^2 \\
&\quad \le C_1'\|s - s^\perp_m\|^2 + y_1\chi^2_{\hat{m}} - C_1\|s - \tilde{s}\|^2.
\end{aligned} \tag{7.17}
$$

Combining (7.11) and (7.17), we obtain (7.16). □

Inequality 3. *For any $y_2 > 1$ and positive constants x_i, $i = 2, 3, 4$, there exist positive reals $C_1 < 1$, $C_1' > 1$, an increasing quadratic polynomial of the form $f_2(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K}_2 > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}_2e^{-\xi}$,*

$$
\begin{aligned}
C_1\|s - \tilde{s}\|^2 \le C_1'\|s - s^\perp_m\|^2 & \\
&+ y_2\frac{V_{\hat{m}}}{T} + x_2\frac{D_{\hat{m}}}{T} + x_3\frac{d_{\hat{m}}}{T} - \operatorname{pen}(\hat{m}) \\
&+ x_4\frac{D_m}{T} + \operatorname{pen}(m) + \frac{f(\xi)}{T}.
\end{aligned} \tag{7.18}
$$

Proof. We bound $\chi^2_{m'}$ using Lemma 7.3 with $\mathrm{V} = \mathbb{R}_+ \times D$ and $\mu(d\mathbf{x}) = s(x)dt\eta(dx)$. We regard the linear model $\mathcal{S}_m$ as a subspace of $L^2(\mathbb{R}_+ \times D, dt\eta(dx))$ with orthonormal basis $\{\frac{\varphi_{1,m}}{\sqrt{T}}, \ldots, \frac{\varphi_{d_m,m}}{\sqrt{T}}\}$. Recall that

$$\chi^2_m = \|s^\perp_m - \hat{s}_m\|^2 = \sum_{i=1}^{d} \left[\iint_{[0,T]\times D} \varphi_{i,m}(x) \frac{\mathcal{J}(dt,dx) - s(x)dt\eta(dx)}{T} \right]^2 .$$

Then, with probability larger than $1 - \sum_{m' \in \mathcal{M}} e^{-x'_{m'}}$,

$$\sqrt{T}\chi_{m'} \leq (1 + x_1)\sqrt{V_{m'}} + \sqrt{2kM_{m'}x'_{m'}} + k(x_1)B_{m'}x'_{m'}, \tag{7.19}$$

for every $m' \in \mathcal{M}$, where $B_{m'} = \sqrt{D_{m'}/T}$,

$$\begin{aligned} V_{m'} &\equiv \int_D \left(\sum_{i=1}^{d_m} \varphi^2_{i,m}(x) \right) s(x)\eta(dx), \\ M_{m'} &\equiv \sup \left\{ \int_D f^2(x)s(x)\eta(dx) : f \in \mathcal{S}_{m'}, \|f\| = 1 \right\}. \end{aligned} \tag{7.20}$$

Now, by Cauchy-Schwarz $\int_D f^2(x)s(x)\eta(dx) \leq \|f\|_\infty \|s\|$, when $\|f\| = 1$, and so the constant $M_{m'}$ above is bounded by $\|s\|\sqrt{D_{m'}}$. In that case, we can use (7.10-i) to obtain

$$\sqrt{2kM_{m'}x'_{m'}} \leq x_2\sqrt{D_{m'}} + \frac{k\|s\|}{2x_2}x'_{m'},$$

for any $x_2 > 0$. On the other hand, by hypothesis $D_{m'} \leq T$, and (7.19) implies that

$$\sqrt{T}\chi_{m'} \leq (1 + x_1)\sqrt{V_{m'}} + x_2\sqrt{D_{m'}} + \left(\frac{k\|s\|}{2x_2} + k(x_1) \right) x'_{m'}.$$

Choosing the constant $x'_{m'}$ as

$$x'_{m'} = \frac{x_3\sqrt{d_{m'}}}{\frac{k\|s\|}{2x_2} + k(x_1)} + \xi,$$

we get that for any $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, and $\xi > 0$,

$$\sqrt{T}\chi_{m'} \leq (1 + x_1)\sqrt{V_{m'}} + x_2\sqrt{D_{m'}} + x_3\sqrt{d_{m'}} + f_1(\xi), \tag{7.21}$$

with probability larger than $1 - \mathcal{K}_1 e^{-\xi}$, where

$$\begin{aligned} f_1(\xi) &= \left(\frac{k\|s\|}{2x_2} + k(x_1) \right)\xi, \\ \mathcal{K}_1 &= \Gamma \sum_{n=1}^{\infty} n^R \exp\left(-\sqrt{n}x_3 \Big/ \left(\frac{k\|s\|}{2x_2} + k(x_1) \right) \right). \end{aligned} \tag{7.22}$$

Squaring (7.21) and using (7.10-ii) repeatedly, we conclude that, for any $y > 1$, $x_2 > 0$, and $x_3 > 0$, there exist both a constant $\mathcal{K}_1 > 0$ and a quadratic function of the form $f_2(\xi) = a\xi^2$ (independent of T, m', and of the family of linear models) such that, with probability greater than $1 - \mathcal{K}_1 e^{-\xi}$,

$$\chi^2_{m'} \leq y\frac{V_{m'}}{T} + x_2\frac{D_{m'}}{T} + x_3\frac{d_{m'}}{T} + \frac{f_2(\xi)}{T}, \quad \forall m' \in \mathcal{M}. \tag{7.23}$$

Then, (7.18) immediately follows from (7.23) and (7.16). $\square$

Proof of (4.5) for the case (c). By the inequality (7.4), we can upper bound $V_{m'}$ by $\hat{V}_{m'}$ on an event of large probability. Namely, for every $x'_{m'} > 0$ and $x > 0$, with probability greater than $1 - \sum_{m' \in \mathcal{M}} e^{-x'_{m'}}$,

$$(1+x)\left(\hat{V}_{m'} + \left(\frac{1}{2x} + \frac{5}{6}\right)\frac{D_{m'}}{T}x'_{m'}\right) \geq V_{m'}, \quad \forall m' \in \mathcal{M}, \tag{7.24}$$

(recall that $D_m = \| \sum_{i=1}^{d_m} \varphi_{i,m}^2 \|_\infty$). Since by hypothesis $D_{m'} < T$, and choosing

$$x'_{m'} = x' d_{m'} + \xi, \quad (x' > 0),$$

it is seen that for any $x > 0$ and $x_4 > 0$, there exist a positive constant $\mathcal{K}_2$ and a function $f(\xi) = b\xi$ (independent of T and of the linear models) such that with probability greater than $1 - \mathcal{K}_2 e^{-\xi}$

$$(1+x)\hat{V}_{m'} + x_4 d_{m'} + f(\xi) \geq V_{m'}, \quad \forall m' \in \mathcal{M}. \tag{7.25}$$

Here, we get $\mathcal{K}_2$ from the polynomial assumption on the class of models. Combining (7.25) and (7.18), it is clear that for any $y_2 > 1$, and positive x_i, $i = 1, 2, 3$, we can choose a pair of positive constants $C_1 < 1$, $C'_1 > 1$, an increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}e^{-\xi}$

$$\begin{aligned} C_1 \|s - \tilde{s}\|^2 \leq C'_1 \|s - s_m^\perp\|^2 & \\ + y_2 \frac{\hat{V}_{\hat{m}}}{T} + x_1 \frac{D_{\hat{m}}}{T} &+ x_2 \frac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ + x_3 \frac{D_m}{T} &+ \mathrm{pen}(m) + \frac{f(\xi)}{T}. \end{aligned} \tag{7.26}$$

Next, we take $y_2 = c$, $x_1 = c'$, and $x_2 = c''$ to cancel $-pen(\hat{m})$ in (7.26). By Lemma 7.4, it follows that

$$C_1 \mathbb{E}\left[\|s - \tilde{s}\|^2\right] \leq C'_1 \|s - s_m^\perp\|^2 + \left(1 + \frac{x_3}{c'}\right)\mathbb{E}\left[\mathrm{pen}(m)\right] + \frac{C''_1}{T}. \tag{7.27}$$

Since m is arbitrary, we obtain the case (c) of (4.5). □

Proof of (4.5) for the case (a). One can bound $V_{m'}$, as given in (7.20), by $D_{m'}\rho$ (assuming that $\rho < \infty$). On the other hand, (7.4) implies that

$$(1+x_1)\left(\frac{\mathcal{N}}{T} + \left(\frac{1}{2x_1} + \frac{5}{6}\right)\frac{\xi}{T}\right) \geq \rho, \tag{7.28}$$

with probability greater than $1 - e^{-\xi}$. Using these bounds for $V_{m'}$ and the assumption that $D_{m'} \leq T$, (7.18) reduces to

$$\begin{aligned} C_1 \|s - \tilde{s}\|^2 \leq C'_1 \|s - s_m^\perp\|^2 & \\ + y \frac{D_{\hat{m}}\mathcal{N}}{T^2} &+ x_1 \frac{d_{\hat{m}}}{T} - \mathrm{pen}(\hat{m}) \\ + x_2 \frac{D_m \mathcal{N}}{T^2} &+ \mathrm{pen}(m) + \frac{f(\xi)}{T}, \end{aligned} \tag{7.29}$$

which is valid with probability $1 - \mathcal{K}e^{-\xi}$. In (7.29), $y > 1$, $x_1 > 0$ and $x_2 > 0$ are arbitrary, while C_1, C'_1, the increasing quadratic polynomial of the form $f(\xi) =$

$a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ are determined by y, x_1, and x_2 independently of the family of linear models and of T. We point out that we divided and multiplied by ρ the terms $D_{\hat{m}}/T$ and D_m/T in (7.18), and then applied (7.28) to get (7.29). It is now clear that $y = c$, and $x_1 = c'$ will produce the desired cancelation. □

Proof of (4.5) for the case (b). We first upper bound $D_{\hat{m}}$ by $\beta^{-1}V_{\hat{m}}$ and $d_{\hat{m}}$ by $(\beta\phi)^{-1}V_{\hat{m}}$ in the inequality (7.18):

$$\begin{aligned} C_1\|s-\tilde{s}\|^2 \le C_1'\|s-s_m^{\perp}\|^2 + \left(y + x_1\beta^{-1} + x_2(\beta\phi)^{-1}\right)\frac{V_{\hat{m}}}{T} \\ - \operatorname{pen}(\hat{m}) + x_3\beta^{-1}\frac{V_m}{T} + \operatorname{pen}(m) + \frac{f(\xi)}{T}. \end{aligned} \tag{7.30}$$

Then, using $d_{m'} \le (\beta\phi)^{-1}V_{m'}$ in (7.25) and letting $x_4(\beta\phi)^{-1}$ vary between 0 and 1, we verify that for any $x' > 0$, a positive constant $\mathcal{K}_4$ and a polynomial f can be found so that with probability greater than $1 - \mathcal{K}_4 e^{-\xi}$,

$$(1+x')\hat{V}_{m'} + f(\xi) \ge V_{m'}, \quad \forall m' \in \mathcal{M}. \tag{7.31}$$

Putting together (7.31) and (7.30), it is clear that for any $y > 1$ and $x_1 > 0$, we can find a pair of positive constants $C_1 < 1$, $C_1' > 1$, an increasing quadratic polynomial of the form $f(\xi) = a\xi^2 + b\xi$, and a constant $\mathcal{K} > 0$ (all independent of the family of linear models and of T) so that, with probability greater than $1 - \mathcal{K}e^{-\xi}$,

$$\begin{aligned} C_1\|s-\tilde{s}\|^2 \le C_1'\|s-s_m^{\perp}\|^2 + y\tfrac{\hat{V}_{\hat{m}}}{T} - \operatorname{pen}(\hat{m}) \\ + x_1\tfrac{V_m}{T} + \operatorname{pen}(m) + \tfrac{f(\xi)}{T}. \end{aligned} \tag{7.32}$$

In particular, by taking $y = c$, the term $-pen(\hat{m})$ cancels out. Lemma 7.4 implies that

$$C_1\mathbb{E}\left[\|s-\tilde{s}\|^2\right] \le C_1'\|s-s_m^{\perp}\|^2 + (1+x_1)\,\mathbb{E}\left[\operatorname{pen}(m)\right] + \frac{C_1''}{T}. \tag{7.33}$$

Finally, (4.5) (b) follows since m is arbitrary. □

Remark 7.5. Let us analyze more carefully the values that the constants C and C' can take in the inequality (4.5). For instance, consider the penalty function of part (c). As we saw in (7.27), the constants C and C' are determined by C_1, C_1', C_1'', and x_3. The constant C_1 was proved to be $y_1 - 1$ if $1 < y_1 < 2$, while it can be made arbitrarily close to one otherwise (see the comment immediately after (7.16)). On the other hand, y_1 itself can be made arbitrarily close to the penalization parameter c since $c = y_2 = y_1(1+x)y$, where x is as in (7.24) and y is in (7.23). Then, when $c \ge 2$, C_1 can be made arbitrarily close to one at the cost of increasing C_1'' in (7.27). Similarly, paying a similar cost, we are able to select C_1' as close to one as we wish and x_3 arbitrarily small. Therefore, it is possible to find for any $\varepsilon > 0$, a constant $C'(\varepsilon)$ (increasing in ε) so that

$$\mathbb{E}\|s-\tilde{s}\|^2 \le (1+\varepsilon)\inf_{m\in\mathcal{M}}\left\{\|s-s_m^{\perp}\|^2 + \mathbb{E}\left[\operatorname{pen}(m)\right]\right\} + \frac{C'(\varepsilon)}{T}. \tag{7.34}$$

A more thorough inspection shows that

$$\lim_{\varepsilon\to 0} C'(\varepsilon)\varepsilon = K,$$

where K depends only c, c', c'', Γ, R, $\|s\|$, and $\|s\|_\infty$. The same reasoning applies to the other two types of penalty functions when $c \ge 2$. In particular, we point out that $\widetilde{C}$ can be made arbitrarily close to 2 in the oracle inequality (4.8) at the price of having a large constant $\widetilde{C}'$.

7.2. Some additional proofs

Proof of Corollary 5.1. The idea is to estimate the bias and the penalized term in (4.5). Clearly, the dimension d_m of $\mathcal{S}_m^k$ is $m(k+1)$. Also, D_m is bounded by $(k+1)^2 m/(b-a)$ (see (7) in [8]), and

$$\mathbb{E}\left[\hat{V}_m\right] = \int_a^b \left(\sum_i \varphi_{i,m}^2(x)\right) s(x)dx \le (k+1)m\|s\|_\infty,$$

since the functions $\varphi_{i,m}$ are orthonormal. On the other hand, by (10.1) in Chapter 2 of [15], if $s \in \mathcal{B}_\infty^\alpha\left(L^p([a,b])\right)$, there is a polynomial $q \in \mathcal{S}_m^k$ such that

$$\|s-q\|_{L^p} \le c_{[\alpha]} |s|_{\mathcal{B}_\infty^\alpha(L^p)} (b-a)^\alpha m^{-\alpha}.$$

Thus,

$$\|s - s_m^\perp\| \le c_{[\alpha]} (b-a)^{\frac{1}{2}-\frac{1}{p}+\alpha} |s|_{\mathcal{B}_\infty^\alpha(L^p)} m^{-\alpha}.$$

By (4.5)), there is a constant M (depending on C, c, c', c'', α, k, $b-a$, p, $|s|_{\mathcal{B}_\infty^\alpha(L^p)}$, and $\|s\|_\infty$), for which

$$\mathbb{E}\left[\|s - \tilde{s}_T\|^2\right] \le M \inf_{m\in\mathcal{M}_T} \left\{m^{-2\alpha} + \frac{m}{T}\right\} + \frac{C'}{T}.$$

It is not hard to see that, for large enough T, the infimum on the above right hand side is $O_\alpha(T^{-2\alpha/(2\alpha+1)})$ (where O_α means that the ratio of the terms is bounded by a constant depending only on α). Since M is monotone in $|s|_{\mathcal{B}_\infty^\alpha(L^p)}$ and $\|s\|_\infty$, (5.2) is verified. $\square$

Proof of Lemma 7.1. Let

$$\gamma_D(f) \equiv -\frac{2}{T} \iint\limits_{[0,T]\times D} f(x)\, \mathcal{J}(dt,dx) + \int_D f^2(x)\, \eta(dx), \tag{7.35}$$

which is well defined for any function $f \in L^2((D,\eta))$, where $D \in \mathcal{B}(\mathbb{R}_0)$ and η is as in (2.6)-(2.8). The projection estimator is the unique minimizer of the contrast function γ_D over S. Indeed, plugging $f = \sum_{i=1}^d \beta_i \varphi_i$ in (7.35) gives $\gamma_D(f) = \sum_{i=1}^d (-2\beta_i\hat{\beta}_i + \beta_i^2)$, and thus, $\gamma_D(f) \ge -\sum_{i=1}^d \hat{\beta}_i^2$, for all $f \in S$. Clearly,

$$\gamma_D(f) = \|f\|^2 - 2\langle f, s\rangle - 2\nu_D(f) = \|f-s\|^2 - \|s\|^2 - 2\nu_D(f).$$

By the very definition of $\tilde{s}$, as the penalized projection estimator,

$$\gamma_D(\tilde{s}) + \mathrm{pen}(\hat{m}) \le \gamma_D(\hat{s}_m) + \mathrm{pen}(m) \le \gamma_D(s_m^\perp) + \mathrm{pen}(m),$$

for any $m \in \mathcal{M}$. Using the above results,

$$\begin{aligned}
\|\tilde{s} - s\|^2 &= \gamma_D(\tilde{s}) + \|s\|^2 + 2\nu_D(\tilde{s}) \\
&\le \gamma(s_m^\perp) + \|s\|^2 + 2\nu_D(\tilde{s}) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}) \\
&= \|s_m^\perp - s\|^2 + 2\nu_D(\tilde{s} - s_m^\perp) + \mathrm{pen}(m) - \mathrm{pen}(\hat{m}).
\end{aligned}$$

Finally, notice that $\nu_D(\tilde{s}-s_m^\perp) = \nu_D(\tilde{s}-s_{\hat{m}}^\perp) + \nu_D(s_{\hat{m}}^\perp - s_m^\perp)$ and that $\nu_D(\hat{s}_m - s_m^\perp) = \chi_m^2$. $\square$

Proof of inequality (7.4). Just note that for any $a, b, \varepsilon > 0$:

$$a - \sqrt{2ab} - \frac{1}{3}b \geq \frac{a}{1+\varepsilon} - \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right) b. \tag{7.36}$$

Evaluating the integral in (7.3) for $-f$, we can write

$$\mathbb{P}\left[\int_{\mathrm{X}} f(x)N(dx) \geq \int_{\mathrm{X}} f(x)\mu(dx) - \|f\|_\mu \sqrt{2u} - \frac{1}{3}\|f\|_\infty u\right] \geq 1 - e^{-u}.$$

Using $\|f\|_\mu^2 \leq \|f\|_\infty \int_{\mathrm{X}} |f(x)|\mu(dx)$ and (7.36), lead to

$$\mathbb{P}\left[\int_{\mathrm{X}} f(x)N(dx) \geq \frac{1}{1+\varepsilon}\int_{\mathrm{X}} f(x)\mu(dx) - \left(\frac{1}{2\varepsilon} + \frac{5}{6}\right)\|f\|_\infty u\right] \geq 1 - e^{-u},$$

which is precisely the inequality (7.4). □

Proof of Lemma 7.4. Let Z^+ be the positive part of Z. First,

$$\mathbb{E}[Z] \leq \mathbb{E}\left[Z^+\right] = \int_0^\infty \mathbb{P}[Z > x]dx.$$

Since h is continuous and strictly increasing, $\mathbb{P}[Z > x] \leq K \exp(-h^{-1}(x))$, where h^{-1} is the inverse of h. Then, changing variables to $u = h^{-1}(x)$,

$$\int_0^\infty \mathbb{P}[Z > x]dx \leq K \int_0^\infty e^{-h^{-1}(x)}dx = K \int_0^\infty e^{-u}h'(u)du.$$

Finally, an integration by parts yields $\int_0^\infty e^{-u}h'(u)du = \int_0^\infty h(u)e^{-u}du$. □

Acknowledgments. The authors are grateful to an anonymous referee for helpful comments and suggestions. It is also a pleasure to thank P. Reynaud-Bouret for helpful discussions.

References

[1] Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. In *Proceeding 2nd International Symposium on Information Theory* (P.N. Petrov & F. Csaki, eds), pp. 267–281.

[2] Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics* **2** 41–68.

[3] Barndorff-Nielsen, O. E. and Shephard, N. (2001). Modelling by Lévy processess for financial economics. In *Lévy Processes. Theory and Applications* (O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick), pp. 283–318.

[4] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related fields* **113** 301–413.

[5] Bertoin, J. (1996). *Lévy Processes.* Cambridge University Press.

[6] Bibby, B. M. and Sorensen, M. (2003). Hyperbolic processes in finance. In *Handbook of Heavy Tailed Distributions in Finance* (S. Rachev ed.), pp. 211–248.

[7] Birgé, L. and Massart, P. (1994). Minimum contrast estimation on Sieves. Technical Report 34, Université Paris-Sud.

[8] Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pp. 55–87.

[9] Breton, J. C., Houdré, C. and Privault, N. (2004). Dimension free and infinite variance tail estimates on Poisson space. Available at ArXiv.

[10] Carr, P., Geman, H., Madan, D. and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business*, April, 305–332.

[11] Carr, P., Geman, H., Madan, D. and Yor, M. (2003). Stochastic volatility for Lévy processes. *Mathematical Finance* **13** 345–382.

[12] Carr, P., Madan, D. and Chang, E. (1998). The variance Gamma process and option pricing. *European Finance Review* **2** 79–105.

[13] Carr, P. and Wu, L. (2004). Time-changed Levy processes and option pricing. *Journal of Financial Economics* **71** (1) 113–141.

[14] Cont, R. and Tankov, P. (2003). *Financial Modelling with Jump Processes.* Chapman & Hall.

[15] DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation.* Springer-Verlag.

[16] Eberlein, E. and Keller, U. (1995). Hyperbolic distribution in finance. *Bernoulli* **1** 281–299.

[17] Grenander, U. (1981). *Abstract Inference.* John Wiley & Sons.

[18] Houdré, C. and Privault, N. (2002). Concentration and deviation inequalities in infinite dimensions via covariance representations. *Bernoulli* **8** (6) 697–720.

[19] Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer-Verlag, Berlin, New York, Heidelberg.

[20] Karr, A. F. (1991). *Point Processes and Their Statistical Inference.* Marcel Dekker Inc.

[21] Kutoyants, Y. A. (1998). *Statistical Inference for Spatial Poisson Processes.* Springer.

[22] Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15** 661–675.

[23] Mancini, C. Estimating the integrated volatility in stochastic volatility models with Lévy type jumps. Technical report, Dipartimento di Matematica per le Decisioni, Universita di Firenze, July 2004. Presented at the Bachelier Finance Society, 3rd World Congress.

[24] Prause, K. (1999). The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. PhD thesis, University of Freiburg, October.

[25] Reynaud-Bouret, P. (2003). Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Related Fields* **126** (1) 103–153.

[26] Rüschendorf, L. and Woerner, J. (2002). Expansion of transition distributions of Lévy processes in small time. *Bernoulli* **8** 81–96.

[27] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press.

[28] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6** 461–464.

[29] Todorov, V. (2005). Econometric analysis of jump-driven stochastic volatility models. Working Paper. Duke University.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 117–127

DOI: 10.1214/074921706000000798

Random walk models associated with distributed fractional order differential equations

Sabir Umarov[1,*,†] **and Stanly Steinberg**[1]

University of New Mexico

Abstract: In this paper the multi-dimensional random walk models governed by distributed fractional order differential equations and multi-term fractional order differential equations are constructed. The scaling limits of these random walks to a diffusion process in the sense of distributions is proved.

1. Introduction

In this paper we construct new random walks connected with fractional order differential equations. Namely, the governing equations corresponding to the constructed random walks are multi-term or distributed fractional order differential equations. Nowadays the connection between random walk and fractional order dynamics is well known, see, for instance [1, 17, 26, 33, 38]. A number of constructive random walk models governed by fractional differential equations in the one-dimensional case were studied by Gillis, et al. [12], Chechkin, et al. [7], Gorenflo, et al. [15, 16], and in the n-dimensional case by Umarov [35], Umarov, et al. [36], Andries, et al. [3]. The governing equation in these studies depends on parameters $\beta \in (0, 1]$ and $\alpha \in (0, 2]$, and is given by the fractional order differential equation

$$\mathcal{D}^{\beta} u(t, x) = D_0^{\alpha} u(t, x),\ t > 0,\ x \in \mathbb{R}^N, \tag{1}$$

where $\mathcal{D}^{\beta}$ is the time-fractional derivative in some sense, and $D_0^{\alpha}, 0 < \alpha < 2$, is the pseudo-differential operator with the symbol $-|\xi|^{\alpha}$. The precise definitions will be given below.

In the present paper we construct the random walks the governing equation of which is a distributed space fractional order differential equation

$$\frac{\partial}{\partial t} u(t, x) = \int_0^2 a(\alpha) D_0^{\alpha} u(t, x) d\alpha,\ t > 0,\ x \in \mathbb{R}^N, \tag{2}$$

where $a(\alpha)$ is a positive integrable function (positively defined distribution).

The study of properties of distributed order differential operators and their applications to diffusion processes has been developed extensively in recent years, although such operators were first mentioned by Caputo [5, 6] in 1960th. The distributed order differential equations have been used by Bagley, et al. [4] to model

[1]Mathematics and Statistics Department, University of New Mexico, Albuquerque, NM, e-mail: sabir@math.unm.edu; stanly@math.unm.edu
*Corresponding author.
†Supported by Fulbright Program.
AMS 2000 subject classifications: primary 60G50; secondary 26A33, 35S05.
Keywords and phrases: random walk, distributed order differential equation, fractional order operator, pseudo-differential operator.

the input-output relationship of linear time-variant system, by Lorenzo, et al. [22] to study of rheological properties of composite materials, by Chechkin, et al. [8] to model some ultraslow and lateral diffusion processes. Diethelm, et al. [9] studied the numerical aspects of such equations. Umarov, et al. [37] studied general properties of distributed order differential equations and solvability problems of the Cauchy and multipoint value problems.

The method used in this paper for construction of multi-dimensional random walks are essentially based on the symbolic calculus of pseudo-differential operators and on the convergence properties of some simple cubature formulas. This method is new even in the one-dimensional case and was suggested recently in [35, 36]. We note that the scaling limits are obtained in terms of characteristic functions of transition probabilities. The equivalence of corresponding convergence notions is well-known (see, [11, 13]). See also the recent book by M. Meerschaert and Scheffler where the multi-dimensional operator stable probability distributions are studied and analogs of different type limits considered. Multi-dimensional random walks are frequently used in modeling various processes in different areas [1, 2, 24–26, 31].

The present report is organized as follows. In Section 2 we give preliminaries simultaneously introducing the terminology that will be used in the paper. We also recall some properties of pseudo-differential operators with constant symbols and lay out some elementary properties of symbols. These properties play an essential role later in the study of the diffusion limits of random walks. In Section 3 we formulate our random walk problem in terms of sequences of i.i.d. (independent identically distributed) random vectors. In this Section we also formulate the main results.

2. Preliminaries

We use the following notation. $\mathbb{R}^N$ is the N-dimensional Euclidean space with coordinates $x = (x_1, \ldots, x_N)$; $\mathbb{Z}^N$ is the N-dimensional integer-valued lattice with nodes $j = (j_1, \ldots, j_N)$. We denote by $x_j = (hj_1, \ldots, hj_N), j \in \mathbb{Z}^N$, the nodes of the uniform lattice $\mathbb{Z}_h^N$ defined as $(h\mathbb{Z})^N$ with a positive number h, the mesh width.

We assume that a walker is located at the origin $x_0 = (0, \ldots, 0)$ at the initial time $t = 0$. In our random walk, at every time instants $t_1 = \tau, t_2 = 2\tau, \ldots, t_n = n\tau, \ldots$ the walker jumps through the nodes of the lattice $\mathbb{Z}_h^N$. By $p_j, j \in \mathbb{Z}^N$, we denote transition probabilities. Namely, p_j means a probability of jumping of the walker from a point $x_k \in \mathbb{Z}_h^N$ to a point $x_{j+k} \in \mathbb{Z}_h^N$, where j and k are in $\mathbb{Z}^N$. Transition probabilities satisfy the non-negativity and normalization conditions:

1. $p_j \geq 0, j \in \mathbb{Z}^N$;
2. $\sum_{j\in\mathbb{Z}^N} p_j = 1$.

Transition probabilities $\{p_j, j \in \mathbb{Z}^N\}$ are associated with a discrete function $p : \mathbb{Z}^N \to [0, 1]$. For given two transition probabilities, p and q we define the convolution operation $p * q$ by

$$(p * q)_j = \sum_{k\in\mathbb{Z}^N} p_k q_{j-k}, j \in \mathbb{Z}^N.$$

Let f be a continuous function integrable over $\mathbb{R}^N$. Then, as is known [32], the rectangular cubature formula

$$\int_{\mathbb{R}^N} f(x)dx = h^N \sum_{j\in\mathbb{Z}^N} f(x_j) + o(1) \tag{3}$$

is valid.

The operators in our random walk models have a close relationship to pseudo-differential operators with the symbols depending only on the dual variables. Symbols are allowed to have singularities. For general orientation to the theory of such operators we refer to [10, 14, 18–21, 34].

Let $A(D), D = (D_1, \ldots, D_N), D_j = \partial/i\partial x_j, j = 1, \ldots, N$, be a pseudo-differential operator with a symbol $A(\xi)$ not depending on x, and defined in $\mathbb{R}^N$. We refer to the variable ξ as a dual variable. Both type of variables, x and ξ belong to $\mathbb{R}^N$ (more precisely, ξ belongs to the conjugate $(\mathbb{R}^N)^* = \mathbb{R}^N$). To avoid confusion sometimes we write $\mathbb{R}^N_x$ and $\mathbb{R}^N_\xi$, indicating their relevance to the variables x and ξ respectively. Further, for a test function $\varphi(x)$ taken from the classical space $S(\mathbb{R}^N_x)$ of rapidly decreasing functions, the Fourier transform

$$\hat{\varphi}(\xi) = F[\varphi](\xi) = \int_{\mathbf{R}^N} \varphi(x) e^{-ix\xi} dx$$

is well defined and belongs again to $S(\mathbb{R}^N_\xi)$. Let $S'(\mathbb{R}^N)$ be the space of tempered distributions, i.e. the dual space to $S(\mathbb{R}^N)$. The Fourier transform for distributions $f \in S'(\mathbb{R}^N_x)$ is usually defined by the extension formula $(\hat{f}(\xi), \varphi(\xi)) = (f(x), \hat{\varphi}(x))$, with the duality pairing $(.,.)$ of $S'(\mathbb{R}^N_\xi)$ and $S(\mathbb{R}^N_\xi)$.

Definition 2.1. Assume G to be an open domain in $\mathbb{R}^N_\xi$. Let a function f be continuous and bounded on $\mathbb{R}^N_x$ and have a Fourier transform (taken in the sense of distributions) $\hat{f}(\xi)$ with compact support in G. We denote by $\Psi_G(\mathbb{R}^N_x)$ the set of all such functions endowed with the following convergence. A sequence of functions $f_m \in \Psi_G(\mathbb{R}^N_x)$ is said to converge to an element $f_0 \in \Psi_G({\mathbb{R}_x}^N)$ iff:

1. there exists a compact set $K \subset G$ such that supp $\hat{f}_m \subset K$ for all $m = 1, 2, \ldots$;
2. $\|f_m - f_0\| = \sup |f_m - f_0| \to 0$ for $m \to \infty$.

In the case $G = \mathbb{R}^N_\xi$ we write simply $\Psi({\mathbb{R}_x}^N)$ omitting $\mathbb{R}^N_\xi$ in the index of $\Psi_G(\mathbb{R}^N_x)$.

Note that according to the Paley-Wiener theorem functions in $\Psi_G(\mathbb{R}^N_x)$ are entire functions of finite exponential type (see [27], [10]).

Denote by $H^s(\mathbb{R}^N_x), s \in (-\infty, +\infty)$ the Sobolev space of elements $f \in S'(\mathbb{R}^N_x)$ for which $(1 + |\xi|^2)^{s/2} |\hat{f}(\xi)| \in L_2(\mathbb{R}^N_\xi)$. It is known [18] that if $f \in L_p({\mathbb{R}_x}^N)$ with $p > 2$, then its Fourier transform $\hat{f}$ belongs to $H^{-s}(\mathbb{R}^N_\xi)$, $s > N(\frac{1}{2} - \frac{1}{p})$. Letting $p \to \infty$ we get $\hat{f} \in H^{-s}(\mathbb{R}^N_\xi), s > \frac{N}{2}$ for $f \in L_\infty({\mathbb{R}_x}^N)$. It follows from this fact and the Paley-Wiener theorem that the Fourier transform of $f \in \Psi_G(\mathbb{R}^N)$ belongs to the space

$$\bigcap_{s > \frac{N}{2}} H_c^{-s}(G),$$

where $H_c^{-s}(G)$ is a negative order Sobolev space of functionals with compact support on G. Hence $\hat{f}$ is a distribution, which is well defined on continuous functions.

Let $\Psi'_{-G}(\mathbb{R}^N)$ be the space of all linear bounded functionals defined on the space $\Psi_G(\mathbb{R}^N)$ endowed with the weak (dual with respect to $\Psi_G(\mathbb{R}^N)$) topology. By the weak topology we mean that a sequence of functionals $g_m \in \Psi'_{-G}(\mathbb{R}^N)$ converges to an element $g_0 \in \Psi'_{-G}(\mathbb{R}^N)$ in the weak sense if for all $f \in \Psi_G(\mathbb{R}^N)$ the sequence of numbers $\langle g_m, f \rangle$ converges to $\langle g_0, f \rangle$ as $m \to \infty$. By $\langle g, f \rangle$ we mean the value of $g \in \Psi'_{-G}(\mathbb{R}^N)$ on an element $f \in \Psi_G(\mathbb{R}^N)$.

Definition 2.2. Let $A(\xi)$ be a continuous function defined in $G \subset \mathbb{R}^N_\xi$. A pseudo-differential operator $A(D)$ with the symbol $A(\xi)$ is defined by the formula

$$A(D)\varphi(x) = \frac{1}{(2\pi)^N}(\hat{\varphi}, A(\xi)e^{-ix\xi}),\ \varphi \in \Psi_G(\mathbb{R}^N). \tag{4}$$

Obviously, the function $A(\xi)e^{-ix\xi}$ is continuous in G. Thus, $A(D)$ in Eq. (4) is well defined on $\Psi_G(\mathbb{R}^N)$. If $\hat{\varphi}$ is an integrable function with $\operatorname{supp}\hat{\varphi} \subset G$, then (4) takes the usual form of pseudo-differential operator

$$A(D)\varphi(x) = \frac{1}{(2\pi)^N}\int A(\xi)\hat{\varphi}(\xi)e^{-ix\xi}d\xi,$$

with the integral taken over G. Note that in general this integral may not make sense even for infinitely differentiable functions with finite support (see [14]).

Now we define the operator $A(-D)$ acting in the space $\Psi'_{-G}(\mathbb{R}^N)$ by the extension formula

$$< A(-D)f, \varphi > = < f, A(D)\varphi >,\ f \in \Psi'_{-G}(\mathbb{R}^N),\ \varphi \in \Psi_G(\mathbb{R}^N). \tag{5}$$

We recall (see [14]) some basic properties of pseudo-differential operators introduced above.

Lemma 2.3. *The pseudo-differential operators $A(D)$ and $A(-D)$ with a continuous symbol $A(\xi)$ act as*

1. $A(D) : \Psi_G(\mathbb{R}^N) \to \Psi_G(\mathbb{R}^N)$,
2. $A(-D) : \Psi'_{-G}(\mathbb{R}^N) \to \Psi'_{-G}(\mathbb{R}^N)$

respectively, and are continuous.

Lemma 2.4. *Let $A(\xi)$ be a function continuous on $\mathbb{R}^N$. Then for $\xi \in \mathbb{R}^N$*

$$A(D)\{e^{-ix\xi}\} = A(\xi)e^{-ix\xi}.$$

Proof. For any fixed $\xi \in \mathbb{R}^N$ the function $e^{-ix\xi}$ is in $\Psi(\mathbb{R}^N)$. We have

$$A(D)\{e^{-ix\xi}\} = \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} A(\eta)e^{-ix\eta}d\mu_\xi(\eta),$$

where $d\mu_\xi(\eta) = F_\eta[e^{-ix\xi}]d\eta = (2\pi)^N\delta(\eta - \xi)d\eta$. Hence $A(D)\{e^{-ix\xi}\} = A(\xi)e^{-ix\xi}$. $\square$

Corollary 2.5. 1. $A(\xi) = (A(D)e^{-ix\xi})e^{ix\xi}$;
2. $A(\xi) = (A(D)e^{-ix\xi})_{|x=0}$;
3. $A(\xi) =< A(-D)\delta(x), e^{-ix\xi} >$, *where δ is the Dirac distribution.*

Remark 2.6. Since the function $e^{-ix\xi}$ does not belong to $S(\mathbb{R}^N)$ and $D(\mathbb{R}^N)$, the representations for the symbol obtained in Lemma 2.4 and Corollary 2.5 are not directly applicable in these spaces.

Denote by D_0^α, $0 < \alpha \leq 2$, the pseudo-differential operator with the symbol $-|\xi|^\alpha$. It is evident that D_0^α coincides with the Laplace operator Δ for $\alpha = 2$. For $\alpha < 2$ it can be considered as a fractional power of the Laplace operator, namely $D_0^\alpha = -(-\Delta)^{\alpha/2}$. D_0^α can also be represented as a hypersingular integral (see, e.g. [29])

$$D_0^\alpha f(x) = b(\alpha)\int_{\mathbb{R}^N_y} \frac{\Delta_y^2 f(x)}{|y|^{N+\alpha}}dy, \tag{6}$$

where Δ_y^2 is the second order centered finite difference in the y direction, and $b(\alpha)$ is norming constant defined as

$$b(\alpha) = \frac{\alpha\Gamma(\frac{\alpha}{2})\Gamma(\frac{N+\alpha}{2})\sin\frac{\alpha\pi}{2}}{2^{2-\alpha}\pi^{1+N/2}}. \tag{7}$$

It is seen from (7) that in the representation (6) of D_0^α the value $\alpha = 2$ is singular.

Lemma 2.7. *For the symbol of D_0^α the following equalities hold true:*

$$\left(D_0^\alpha e^{ix\xi}\right)|_{x=0} = b(\alpha)\int_{\mathbb{R}_y^N} \frac{\Delta_y^2 e^{ix\xi}}{|y|^{N+\alpha}}dy|_{x=0} = -|\xi|^\alpha, 0 < \alpha < 2. \tag{8}$$

Proof. This statement is a direct implication of Corollary 2.5 applied to the operator D_0^α. ☐

The cubature formula (3) yields for the integral in the right hand side of (6)

$$\int_{\mathbb{R}^N} \frac{\Delta_y^2 f(x_j)}{|y|^{N+\alpha}}dy = h^\alpha \sum_{k\in\mathbb{Z}^N} \frac{\Delta_k^2 f_j}{|k|^{N+\alpha}} + o(1),\ j\in\mathbb{Z}^N, \tag{9}$$

where $f_j = f(x_j)$ and $|k|$ is Euclidean norm of $k = (k_1, \dots, k_N) \in \mathbb{Z}^N$.

Consider the distributed space fractional order differential equation

$$\frac{\partial}{\partial t}u(t,x) = \int_0^2 a(\alpha)D_0^\alpha u(t,x)d\alpha,\ t>0,\ x\in\mathbb{R}^N, \tag{10}$$

where $a(\alpha)$ is a positive (in general, generalized) function defined in $(0,2]$. A distribution $G(t,x)$, which satisfies the equation (10) in the weak sense and the condition

$$G(0,x) = \delta(x), x\in\mathbb{R}^N, \tag{11}$$

where $\delta(x)$ is the Dirac's distribution, is called a fundamental solution of the Cauchy problem (10), (11). In the particular case of

$$a(\alpha) = \sum_{m=1}^M a_m\delta(\alpha-\alpha_m),\ 0<\alpha_1<\dots<\alpha_M\le 2, \tag{12}$$

with positive constants a_m we get a multiterm space fractional differential equation

$$\frac{\partial}{\partial t}u(t,x) = \sum_{m=1}^M a_m D_0^{\alpha_m}u(t,x)\, t>0,\ x\in\mathbb{R}^N. \tag{13}$$

Denote the operator on the right hand side of the equation (10) by $\mathcal{B}(D)$. It can be represented as a pseudo-differential operator with the symbol

$$\mathcal{B}(\xi) = -\int_0^2 a(\alpha)|\xi|^\alpha d\alpha. \tag{14}$$

It is not hard to verify that the fundamental solution of equation (10) is

$$G(t,x) = F^{-1}\left(e^{t\mathcal{B}(\xi)}\right), \tag{15}$$

where F^{-1} stands for the inverse Fourier transform. In the particular case of $a(\alpha) = \delta(\alpha - 2)$ we have the classical heat conduction equation

$$\frac{\partial}{\partial t}u(t,x) = \Delta u(t,x),\ t > 0,\ x \in \mathbf{R}^N,$$

whose fundamental solution is the Gaussian probability density function evolving in time

$$G_2(t,x) = \frac{1}{(4\pi t)^{n/2}} e^{\frac{-|x|^2}{4t}}.$$

For $a(\alpha) = \delta(\alpha - \alpha_0),\ 0 < \alpha_0 < 2$, the corresponding fundamental solution is the Levy α_0-stable probability density function [30]

(16) $$G_{\alpha_0}(t,x) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-t|\xi|^{\alpha_0}} e^{ix\xi} d\xi.$$

The power series representation of the stable Levy probability density functions is studied in [3, 23, 33]. Recall also that $\alpha_0 = 1$ corresponds to the Cauchy-Poisson probability density (see [28])

$$G_1(t,x) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{1}{(|x|^2 + t^2)^{(n+1)/2}}.$$

3. Main results: construction of random walks

In this Section we construct random walks associated with distributed space fractional order differential equations (10). More precisely, we show that the special scaling limit of the constructed random walk is a diffusion process whose probability density function is the fundamental solution of (10).

Let $\mathbf{X}$ be an N-dimensional random vector [25] which takes values in $\mathbb{Z}^N$. Let the random vectors $\mathbf{X}_1, \mathbf{X}_2, \dots$ also be N-dimensional independent identically distributed random vectors, all having the same probability distribution, common with $\mathbf{X}$. We introduce a spatial grid $\{x_j = jh, j \in \mathbb{Z}^N\}$, with $h > 0$ and temporal grid $\{t_n = n\tau, n = 0, 1, 2, \dots\}$ with a step $\tau > 0$. Consider the sequence of random vectors

$$\mathbf{S}_n = h\mathbf{X}_1 + h\mathbf{X}_2 + \cdots + h\mathbf{X}_n, n = 1, 2, \dots$$

taking $\mathbf{S}_0 = \mathbf{0}$ for convenience. We interpret $\mathbf{X}_1, \mathbf{X}_2, \dots,$ as the jumps of a particle sitting in $x = x_0 = \mathbf{0}$ at the starting time $t = t_0 = 0$ and making a jump $\mathbf{X}_n$ from $\mathbf{S}_{n-1}$ to $\mathbf{S}_n$ at the time instant $t = t_n$. Then the position $\mathbf{S}(t)$ of the particle at time t is

$$\sum_{1 \le k \le t/\tau} \mathbf{X}_k.$$

Denote by $y_j(t_n)$ the probability of sojourn of the walker at x_j at the time t_n. Taking into account the recursion $\mathbf{S}_{n+1} = \mathbf{S}_n + h\mathbf{X}_n$ we have

(17) $$y_j(t_{n+1}) = \sum_{k \in \mathbb{Z}^N} p_k y_{j-k}(t_n), j \in \mathbb{Z}^N, n = 0, 1, \dots$$

The convergence of the sequence $\mathbf{S}_n$ when $n \to \infty$ means convergence of the discrete probability law $(y_j(t_n))_{j \in \mathbb{Z}^N}$, properly rescaled as explained below, to the probability law with a density $u(t,x)$ in the sense of distributions (in law). This is equivalent to the locally uniform convergence of the corresponding characteristic functions (see for details [25]). We use this idea to prove the convergence of the sequence of characteristic functions of the constructed random walks to the fundamental solution of distributed order diffusion equations.

In order to construct a random walk relevant to (10) we use the approximation (3) for the integral on the right hand side of (6), namely

$$D_0^\alpha u(t,x_j) \approx b(\alpha) \sum_{k\in\mathbb{Z}^N} \frac{u_{j+k}(t) - 2u_j(t) + u_{j-k}(t)}{|k|^{N+\alpha} h^\alpha},$$

and the first order difference ratio

$$\frac{\partial u}{\partial t} \approx \frac{u_j(t_{n+1}) - u_j(t_n)}{\tau}$$

for $\frac{\partial u}{\partial t}$ with the time step $\tau = t/n$. Then from (10) we derive the relation (17) with the transition probabilities

$$p_k = \begin{cases} 1 - 2\tau \sum_{m\neq 0} \frac{Q_m(h)}{|m|^N}, & \text{if } k = 0; \\ 2\tau \frac{Q_m(h)}{|k|^N}, & \text{if } k \neq 0, \end{cases} \tag{18}$$

where

$$Q_m(h) = \int_0^2 |m|^{-\alpha} \rho(\alpha, h) d\alpha, \;\; \rho(\alpha,h) = \frac{a(\alpha)b(\alpha)}{h^\alpha}. \tag{19}$$

Assume that the condition

$$\sigma(\tau, h) := 2\tau \sum_{m\neq 0} \frac{Q_m(h)}{|m|^N} \leq 1. \tag{20}$$

is fulfilled. Then, obviously, the transition probabilities satisfy the properties:

1. $\sum_{k\in\mathbb{Z}^N} p_k = 1$;
2. $p_k \geq 0, k \in \mathbb{Z}^N$.

Introduce the function

$$\mathcal{R}(\alpha) = \sum_{k\neq 0} \frac{1}{|k|^{N+\alpha}} = \sum_{m=1}^{\infty} \frac{M_m}{m^{N+\alpha}}, \;\; 0 < \alpha \leq 2,$$

where $M_m = \sum_{|k|=m} 1$. (In the one-dimensional case $\mathcal{R}(\alpha)$ coincides with the Riemann's zeta-function, $\mathcal{R}(\alpha) = 2\zeta(1+\alpha)$.) The Eq. (20) can be rewritten as

$$\sigma(\tau, h) = 2\tau \int_0^2 \frac{a(\alpha)b(\alpha)\mathcal{R}(\alpha)}{h^\alpha} d\alpha \leq 1. \tag{21}$$

It follows from the latter inequality that $h \to 0$ yelds $\tau \to 0$. This, in turn, yields $n = t/\tau \to \infty$ for any finite t.

Now we assume that the singular support of a does not contain 2, i.e., $\{2\} \notin$ *singsupp* a [1].

[1]This condition relates only to distributions.

Theorem 3.1. *Let* $\mathbf{X}$ *be a random vector with the transition probabilities* $p_k = P(\mathbf{X} = x_k), k \in \mathbb{Z}^N$, *defined in Eqs:* (18), (19) *and, which satisfy the condition* (20) *(or, the same,* (21)*). Then the sequence of random vectors* $\mathbf{S}_n = h\mathbf{X}_1 + \cdots + h\mathbf{X}_n$, *converges as* $n \to \infty$ *in law to the random vector whose probability density function is the fundamental solution of the distributed space fractional order differential equation* (10)*.*

Proof. We have to show that the sequence of random vectors $\mathbf{S}_n$ tends to the random vector with pdf $G(t,x)$ in Eq. (15), or the same, the discrete function $y_j(t_n)$ tends to $G(t,x)$ as $n \to \infty$. It is obvious that the Fourier transform of $G(t,x)$ with respect to the variable x is the function $\hat{G}(t,\xi) = e^{t\mathcal{B}(\xi)}$, where $\mathcal{B}(\xi)$ is defined in Eq. (14). Let $\hat{p}(-\xi)$ be the characteristic function corresponding to the discrete function p_k, $k \in \mathbb{Z}^N$, that is

$$\hat{p}(-\xi) = \sum_{k \in \mathbb{Z}^N} p_k e^{ik\xi}.$$

It follows from the recursion formula (17) (which exhibits the convolution) and the well known fact that convolution goes over in multiplication by the Fourier transform, the characteristic function of $y_j(t_n)$ can be represented in the form

$$\hat{y}_j(t_n, -\xi) = \hat{p}^n(-\xi).$$

Taking this into account it suffices to show that

$$\hat{p}^n(-h\xi) \to e^{t\mathcal{B}(\xi)}, n \to \infty. \tag{22}$$

The next step of the proof is based on the following simple fact: if a sequence s_n converges to s for $n \to \infty$, then

$$\lim (1 + \frac{s_n}{n})^n = e^s. \tag{23}$$

We have

$$\begin{aligned}
\hat{p}^n(-h\xi) &= (1 - \tau \sum_{k \neq 0} \frac{Q_k}{|k|^N}(1 - e^{ik\xi h}))^n \\
&= (1 - \tau \sum_{k \neq 0} \frac{1}{|k|^N} \int_0^2 \frac{a(\alpha)b(\alpha)d\alpha}{|k|^\alpha h^\alpha}(1 - e^{ik\xi h}))^n \\
&= (1 + \frac{t\int_0^2 a(\alpha)\{b(\alpha)\sum \frac{\Delta^2 e^{ik\xi h}}{|kh|^{N+\alpha}} h^N\}d\alpha}{n})^n
\end{aligned} \tag{24}$$

It follows from (3) and Corollary 2.5 that

$$b(\alpha) \sum_{k \in Z^N} \frac{\Delta^2 e^{ik\xi h}}{|kh|^{N+\alpha}} h^N$$

tends to $(D_0^\alpha e^{ix\xi})_{|x=0} = -|\xi|^\alpha$ as $h \to 0$ (or, the same, $n \to \infty$) for all $\alpha \in (0,2]$. Hence

$$s_n = \int_0^2 a(\alpha)\{b(\alpha)\sum \frac{\Delta^2 e^{ik\xi h}}{|kh|^{N+\alpha}} h^N\}d\alpha \to \mathcal{B}(\xi),\ n \to \infty\ (h \to 0).$$

Thus, in accordance with (23) we have

$$\hat{p}^n(-h\xi) \to e^{t\mathcal{B}(\xi)}, n \to \infty.$$

□

The random walk related to the multiterm fractional diffusion equation can be derived from Theorem 3.1. Assume that $a(\alpha)$ has the form (12) with $0 < \alpha_1 < \cdots < \alpha_M < 2$. So, we again exclude the case $\{2\} \in singsupp\, a$.

Theorem 3.2. *Let the transition probabilities* $p_k = P(\mathbf{X} = x_k), k \in \mathbb{Z}^N$, *of the random vector* $\mathbf{X}$ *be given as follows:*

$$p_k = \begin{cases} 1 - \sum_{j \neq 0} \frac{1}{|j|^N} \sum_{m=1}^{M} \frac{\mu_m a_m b(\alpha_m)}{|j|^{\alpha_m}}, & \text{if } k = 0; \\ \frac{1}{|k|^N} \sum_{m=1}^{M} \frac{\mu_m a_m b(\alpha_m)}{|j|^{\alpha_m}}, & \text{if } k \neq 0, \end{cases}$$

where $\mu_m = \frac{2\tau}{h^{\alpha_m}}, m = 1, \ldots, M$. *Assume,*

$$\sum_{m=1}^{M} a_m b(\alpha_m) \mathcal{R}(\alpha_m) \mu_m \leq 1.$$

Then the sequence of random vectors $\mathbf{S}_n = h\mathbf{X}_1 + \cdots + h\mathbf{X}_n$, *converges as* $n \to \infty$ *in law to the random vector whose probability density function is the fundamental solution of the multiterm fractional order differential equation (13).*

Remark 3.3. The condition $\{2\} \notin singsupp\, a$ is required due to singularity of the value $\alpha = 2$ in the definition of D_0^{α} (see (7)). The particular case $a(\alpha) = \delta(\alpha - 2)$ reduces to the classic heat conduction equation and corresponding random walk is the classic Brownian motion. In more general case of $a(\alpha) = \sum_{l=0}^{m} c_l \delta^{(l)}(\alpha - 2)$ this condition leads to the scaling limit with $\sigma(\tau, h) = h^2 ln \frac{1}{h}$ (see, also [16]).

This work was supported in part by NIH grant P20 GMO67594.

References

[1] Adler, R., Feldman, R. and Taqqu, M. (1998). *A Practical Guide to Heavy Tails.* Birkhäuser, Boston. MR1652283

[2] Anderson, P. and Meerschaert, M. M. (1998). Modeling river flows with heavy tails. *Water Resour. Res.* **34** 2271–2280.

[3] Andries, E., Steinberg, S. and Umarov, S. Fractional space-time differential equations: theoretical and numerical aspects, in preparation.

[4] Bagley, R. L. and Torvic, P. J. (2000). On the existance of the order domain and the solution of distributed order equations I, II. *Int. J. Appl. Math* **2** 865–882, 965–987. MR1760547

[5] Caputo, M. (1967). Linear models of dissipation whose Q is almost frequency independent. II. *Geophys. J. R. Astr. Soc.* **13** 529–539.

[6] Caputo, M. (2001). Distributed order differential equations modeling dielectric induction and diffusion. *Fract. Calc. Appl. Anal.* **4** 421–442. MR1874477

[7] Chechkin, A. V. and Gonchar, V. Yu. (1999). A model for ordinary Levy motion. *ArXiv:cond-mat/9901064* **v1**, 14 pp.

[8] Chechkin, A. V., Gorenflo, R., Sokolov, I. M. and Gonchar, V. Yu. (2003). Distributed order time fractional diffusion equation. *FCAA* **6** 259–279. MR2035651

[9] Diethelm, K. and Ford, N. J. (2001). Numerical solution methods for distributed order differential equations. *FCAA* **4** 531–542. MR1874482

[10] Dubinskii, Yu. A. (1991). *Analytic Pseudo-differential Operators and Their Applications.* Kluwer Academic Publishers, Dordrecht. MR1175753

[11] Feller, W. (1974). *An Introduction to Probability Theory and Its Applications.* John Wiley and Sons, New York–London–Sydney. MR

[12] Gillis, G. E. and Weiss, G. H. (1970). Expected number of distinct sites visited by a random walk with an infinite variance. *J. Math. Phys.* **11** 1307–1312. MR0260036

[13] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading.

[14] Gorenflo, R., Luchko, Yu. and Umarov, S. (2000). On the Cauchy and multi-point value problems for partial pseudo-differential equations of fractional order. *FCAA* **3** 250–275. MR1788164

[15] Gorenflo, R. and Mainardi, F. (1999). Approximation of Lévy-Feller diffusion by random walk. *ZAA* **18** (2) 231–246. MR1701351

[16] Gorenflo, R. and Mainardi, M. (2001). Random walk models approximating symmetric space-fractional diffusion processes. In: Elschner, Gohberg and Silbermann (Eds), *Problems in Mathematical Physics* (Siegfried Prössdorf Memorial Volume). Birkhäuser Verlag, Boston-Basel-Berlin, pp. 120–145.

[17] Gorenflo, R., Mainardi, F., Moretti, D., Pagnini, G. and Paradisi, P. (2002). Discrete random walk models for space-time fractional diffusion. *Chemical Physics.* **284** 521–541.

[18] Hörmander, L. (1983). *The Analysis of Linear Partial Differential Operators: I. Distribution Theory and Fourier Analysis.* Springer-Verlag, Berlin–Heidelberg–New York–Tokyo. MR0717035

[19] Jacob, N. (2001). *Pseudo-differential Operators and Markov Processes.* Vol. **I**: *Fourier Analysis and Semigroups*. Imperial College Press, London. MR1873235

[20] Jacob, N. (2002). *Pseudo-differential Operators and Markov Processes.* Vol. **II**: *Generators and Their Potential Theory.* Imperial College Press, London. MR1917230

[21] Jacob, N. (2005). *Pseudo-differential Operators and Markov Processes.* Vol. **III**: *Markov Processes and Applications.* Imperial College Press, London. MR2158336

[22] Lorenzo, C. F. and Hartley, T. T. (2002). Variable order and distributed order fractional operators. *Nonlinear Dynamics* **29** 57–98. MR1926468

[23] Mainardi, F., Luchko, Yu. and Pagnini, G. (2001). The fundamental solution of the space-time fractional diffusion equation. *FCAA* **4** 153–192. MR

[24] McCulloch, J. (1996). Financial applications of stable distributions. In: *Statistical Methods in Finance: Handbook of Statistics* **14** (G. Madfala and C.R. Rao, Eds). Elsevier, Amsterdam, pp. 393–425. MR1602156

[25] Meerschaert, M. M. and Scheffler, P.-H. (2001). *Limit Distributions for Sums of Independent Random Vectors. Heavy Tails in Theory and Practice.* John Wiley and Sons, Inc. MR1840531

[26] Metzler, R. and Klafter, J. (2000). The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports* **339** 1–77. MR1809268

[27] Nikol'skii, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems.* Springer-Verlag. MR0374877

[28] Rubin, B. (1996). *Fractional Integrals and Potentials.* Addison Wesley, Longman Ltd. MR1428214

[29] Samko, S. G., Kilbas, A. A. and Marichev, O. I. (1993). *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science

Publishers, New York and London (Originally published in 1987 in Russian). MR1347689

[30] Samorodnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York–London.

[31] Schmitt, F. G. and Seuront, L. (2001). Multifractal random walk in copepod behavior. *Physica* A **301** 375–396.

[32] Sobolev, S. L. (1992). *Introduction to the Theory of Cubature Formulas.* Nauka, Moscow (1974) (in Russian). Translated into English: S. L. Sobolev, *Cubature Formulas and Modern Analysis: An Introduction.* Gordon and Breach Science Publishers. MR1248825

[33] Uchaikin, V. V. and Zolotarev, V. M. (1999). *Chance and Stability. Stable Distributions and Their Applications*. VSP, Utrecht. MR1745764

[34] Umarov, S. (1997, 1998). Non-local boundary value problems for pseudo-differential and differential-operator equations I, II. *Differential Equations* **33, 34** 831–840, 374–381. MR1615099, MR1668214

[35] Umarov, S. (2003). Multidimensional random walk model approximating fractional diffusion processes. *Docl. Ac. Sci. of Uzbekistan.*

[36] Umarov, S. and Gorenflo, R. (2005). On multi-dimensional symmetric random walk models approximating fractional diffusion processes. *FCAA* **8** 73–88.

[37] Umarov, S. and Gorenflo, R. (2005). The Cauchy and multipoint problem for distributed order fractional differential equations. *ZAA* **24** 449–466.

[38] Zaslavsky, G. (2002). Chaos, fractional kinetics, and anomalous transport. *Physics Reports* **371** 461–580. MR1937584

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 128–147

DOI: 10.1214/074921706000000806

Fractal properties of the random string processes

Dongsheng Wu[1,*] **and Yimin Xiao**[1,†]

Department of Statistics and Probability, Michigan State University

Abstract: Let $\{u_t(x), t \geq 0, x \in \mathbb{R}\}$ be a random string taking values in $\mathbb{R}^d$, specified by the following stochastic partial differential equation [Funaki (1983)]:

$$\frac{\partial u_t(x)}{\partial t} = \frac{\partial^2 u_t(x)}{\partial x^2} + \dot{W},$$

where $\dot{W}(x,t)$ is an $\mathbb{R}^d$-valued space-time white noise.

Mueller and Tribe (2002) have proved necessary and sufficient conditions for the $\mathbb{R}^d$-valued process $\{u_t(x) : t \geq 0, x \in \mathbb{R}\}$ to hit points and to have double points. In this paper, we continue their research by determining the Hausdorff and packing dimensions of the level sets and the sets of double times of the random string process $\{u_t(x) : t \geq 0, x \in \mathbb{R}\}$. We also consider the Hausdorff and packing dimensions of the range and graph of the string.

1. Introduction and preliminaries

Consider the following model of a random string introduced by Funaki [5]:

$$\frac{\partial u_t(x)}{\partial t} = \frac{\partial^2 u_t(x)}{\partial x^2} + \dot{W}, \tag{1}$$

where $\dot{W}(x,t)$ is a space-time white noise in $\mathbb{R}^d$, which is assumed to be adapted with respect to a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, where $\mathcal{F}$ is complete and the filtration $\{\mathcal{F}_t, t \geq 0\}$ is right continuous. The components $\dot{W}_1(x,t), \ldots, \dot{W}_d(x,t)$ of $\dot{W}(x,t)$ are independent space-time white noises, which are generalized Gaussian processes with covariance given by

$$\mathbb{E}\big[\dot{W}_j(x,t)\dot{W}_j(y,s)\big] = \delta(x-y)\delta(t-s), \quad (j = 1, \ldots, d).$$

That is, for every $1 \leq j \leq d$, $W_j(f)$ is a random field indexed by functions $f \in L^2([0,\infty) \times \mathbb{R})$ and, for all $f, g \in L^2([0,\infty) \times \mathbb{R})$, we have

$$\mathbb{E}\big[W_j(f)W_j(g)\big] = \int_0^\infty \int_{\mathbb{R}} f(t,x)g(t,x)\, dx dt.$$

Hence $W_j(f)$ can be represented as

$$W_j(f) = \int_0^\infty \int_{\mathbb{R}} f(t,x)\, W_j(dx\, dt).$$

[1]Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA, e-mail: wudongsh@msu.edu; xiao@sst.msu.edu, url: www.msu.edu/~wudongsh; www.stt.msu.edu/~xiaoyimi

[*]Research partially supported by NSF grant DMS-0417869.

[†]Research partially supported by NSF grant DMS-0404729.

AMS 2000 subject classifications: primary 60H15, 60G15, 60G17; secondary 28A80.

Keywords and phrases: random string process, stationary pinned string, Hausdorff dimension, packing dimension, range, graph, level set, double times.

Note that $W(f)$ is $\mathcal{F}_t$-measurable whenever f is supported on $[0,t] \times \mathbb{R}$.

Recall from Mueller and Tribe [9] that a solution of (1) is defined as an $\mathcal{F}_t$-adapted, continuous random field $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ with values in $\mathbb{R}^d$ satisfying the following properties:

(i) $u_0(\cdot) \in \mathcal{E}_{\exp}$ almost surely and is adapted to $\mathcal{F}_0$, where $\mathcal{E}_{\exp} = \cup_{\lambda>0}\mathcal{E}_\lambda$ and

$$\mathcal{E}_\lambda = \Big\{ f \in C(\mathbb{R}, \mathbb{R}^d) : \ |f(x)|\, e^{-\lambda|x|} \to 0 \ \text{ as } |x| \to \infty \Big\};$$

(ii) For every $t > 0$, there exists $\lambda > 0$ such that $u_s(\cdot) \in \mathcal{E}_\lambda$ for all $s \leq t$, almost surely;

(iii) For every $t > 0$ and $x \in \mathbb{R}$, the following Green's function representation holds

$$u_t(x) = \int_{\mathbb{R}} G_t(x-y)u_0(y)dy + \int_0^t G_{t-r}(x-y)\, W(dy\, dr), \tag{2}$$

where $G_t(x) = \frac{1}{\sqrt{4\pi t}}\, e^{-\frac{x^2}{4t}}$ is the fundamental solution of the heat equation.

We call each solution $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ of (1) a random string process with values in $\mathbb{R}^d$, or simply a random string as in [9]. Note that, whenever the initial conditions u_0 are deterministic, or are Gaussian fields independent of $\mathcal{F}_0$, the random string processes are Gaussian. We recall briefly some basic properties about the solutions of (1), and refer to Mueller and Tribe [9] and Funaki [5] for further information on stochastic partial differential equations (SPDEs) related to random motion of strings.

Funaki [5] investigated various properties of the solutions of semi-linear type SPDEs which are more general than (1). In particular, his results (cf. Lemma 3.3 in [5]) imply that every solution $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ of (1) is Hölder continuous of any order less than $\frac{1}{2}$ in space and $\frac{1}{4}$ in time. This anisotropic property of the process $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ makes it a very interesting object to study. Recently Mueller and Tribe [9] have found necessary and sufficient conditions [in terms of the dimension d] for a random string in $\mathbb{R}^d$ to hit points or to have double points of various types. They have also studied the question of recurrence and transience for $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$. Note that, in general, a random string may not be Gaussian, a powerful step in the proofs of Mueller and Tribe [9] is to reduce the problems about a general random string process to those of the stationary pinned string $U = \{U_t(x), t \geq 0,\, x \in \mathbb{R}\}$, obtained by taking the initial functions $U_0(\cdot)$ in (2) to be defined by

$$U_0(x) = \int_0^\infty \int (G_r(x-z) - G_r(z))\, \widetilde{W}(dzdr), \tag{3}$$

where $\widetilde{W}$ is a space-time white noise independent of the white noise $\dot{W}$. One can verify that $U_0 = \{U_0(x) : x \in \mathbb{R}\}$ is a two-sided $\mathbb{R}^d$ valued Brownian motion satisfying $U_0(0) = 0$ and $\mathbb{E}[(U_0(x) - U_0(y))^2] = |x-y|$. We assume, by extending the probability space if needed, that U_0 is $\mathcal{F}_0$-measurable. As pointed out by Mueller and Tribe [9], the solution to (1) driven by the noise $W(x,s)$ is then given by

$$\begin{aligned} U_t(x) &= \int G_t(x-z)U_0(z)dz + \int_0^t \int G_r(x-z)W(dzdr) \\ &= \int_0^\infty (G_{t+r}(x-z) - G_r(z))\, \widetilde{W}(dzdr) + \int_0^t \int G_r(x-z)W(dzdr). \end{aligned} \tag{4}$$

A continuous version of the above solution is called *a stationary pinned string.*

The components $\{U_t^j(x) : t \geq 0,\, x \in \mathbb{R}\}$ for $j = 1, \ldots, d$ are independent and identically distributed Gaussian processes. In the following we list some basic properties of the processes $\{U_t^j(x) : t \geq 0,\, x \in \mathbb{R}\}$, which will be needed for proving the results in this paper. Lemma 1.1 below is Proposition 1 of Mueller and Tribe [9].

Lemma 1.1. *The components $\{U_t^j(x) : t \geq 0,\, x \in \mathbb{R}\}$ $(j = 1, \ldots, d)$ of the stationary pinned string are mean-zero Gaussian random fields with stationary increments. They have the following covariance structure: for $x, y \in \mathbb{R}$, $t \geq 0$,*

$$\mathbb{E}\Big[\big(U_t^j(x) - U_t^j(y)\big)^2\Big] = |x - y|, \tag{5}$$

and for all $x, y \in \mathbb{R}$ and $0 \leq s < t$,

$$\mathbb{E}\Big[\big(U_t^j(x) - U_s^j(y)\big)^2\Big] = (t - s)^{1/2} F\big(|x - y|(t - s)^{-1/2}\big), \tag{6}$$

where

$$F(a) = (2\pi)^{-1/2} + \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} G_1(a - z) G_1(a - z') \big(|z| + |z'| - |z - z'|\big)\, dz dz'.$$

$F(x)$ is a smooth function, bounded below by $(2\pi)^{-1/2}$, and $F(x)/|x| \to 1$ as $|x| \to \infty$. Furthermore there exists a positive constant $c_{1,1}$ such that for all $s, t \in [0, \infty)$ and all $x, y \in \mathbb{R}$,

$$c_{1,1} \big(|x - y| + |t - s|^{1/2}\big) \leq \mathbb{E}\Big[\big(U_t^j(x) - U_s^j(y)\big)^2\Big] \leq 2\big(|x - y| + |t - s|^{1/2}\big). \tag{7}$$

It follows from (6) that the stationary pinned string has the following scaling property [or operator-self-similarity]: For any constant $c > 0$,

$$\{c^{-1} U_{c^4 t}(c^2 x) : t \geq 0, x \in \mathbb{R}\} \overset{d}{=} \{U_t(x) : t \geq 0,\, x \in \mathbb{R}\}, \tag{8}$$

where $\overset{d}{=}$ means equality in finite dimensional distributions; see Corollary 1 in [9].

We will also need more precise information about the asymptotic property of the function $F(x)$. By a change of variables we can write it as

$$F(x) = -(2\pi)^{-1/2} + \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} G_1(z) G_1(z') \big(|z - x| + |z' - x|\big)\, dz dz'. \tag{9}$$

Denote the above double integral by $H(x)$. Then it can be written as

$$H(x) = \int_{\mathbb{R}} G_1(z)\, |z - x|\, dz. \tag{10}$$

The following lemma shows that the behavior of $H(x)$ is similar to that of $F(x)$, and the second part describes how fast $H(x)/|x| \to 1$ as $x \to \infty$.

Lemma 1.2. *There exist positive constants $c_{1,2}$ and $c_{1,3}$ such that*

$$c_{1,2} \big(|x - y| + |t - s|^{1/2}\big) \leq |t - s|^{1/2} H\big(|x - y||t - s|^{-1/2}\big) \leq c_{1,3} \big(|x - y| + |t - s|^{1/2}\big). \tag{11}$$

Moreover, we have the limit:

$$\lim_{x \to \infty} |H(x) - x| = 0. \tag{12}$$

Proof. The inequality (11) follows from the proof of (7) in [9], p. 9. Hence we only need to prove (12).

By (10), we see that for $x > 0$,

$$
\begin{aligned}
H(x) - x &= \int_{\mathbb{R}} G_1(z)\big(|z-x| - x\big)dz \\
&= \int_x^{\infty} (z-2x)\, G_1(z)\, dz - \int_{-\infty}^{x} z\, G_1(z)\, dz \\
&= 2\int_x^{\infty} (z-x)\, G_1(z)\, dz.
\end{aligned} \tag{13}
$$

Since the last integral tends to 0 as $x \to \infty$, (12) follows. $\square$

The following lemmas indicate that, for every $j \in \{1, 2, \ldots, d\}$, the Gaussian process $\{U_t^j(x), t \geq 0, x \in \mathbb{R}\}$ satisfies some preliminary forms of sectorial local nondeterminism; see [13] for more information on the latter. Lemma 1.3 is implied by the proof of Lemma 3 in [9], p. 15, and Lemma 1.4 follows from the proof of Lemma 4 in [9], p. 21.

Lemma 1.3. *For any given $\varepsilon \in (0,1)$, there exists a positive constant $c_{1,4}$, which depends on ε only, such that*

$$
\mathrm{Var}\Big(U_t^j(x)\Big|U_s^j(y)\Big) \geq c_{1,4}\Big(|x-y| + |t-s|^{1/2}\Big) \tag{14}
$$

for all (t,x), $(s,y) \in [\varepsilon, \varepsilon^{-1}] \times [-\varepsilon^{-1}, \varepsilon^{-1}]$.

Lemma 1.4. *For any given constants $\varepsilon \in (0,1)$ and $L > 0$, there exists a constant $c_{1,5} > 0$ such that*

$$
\begin{aligned}
&\mathrm{Var}\Big(U_{t_2}^j(x_2) - U_{t_1}^j(x_1)\Big|U_{s_2}^j(y_2) - U_{s_1}^j(y_1)\Big) \\
&\qquad \geq c_{1,5}\Big(|x_1-y_1| + |x_2-y_2| + |t_1-s_1|^{1/2} + |t_2-s_2|^{1/2}\Big)
\end{aligned} \tag{15}
$$

for all (t_k, x_k), $(s_k, y_k) \in [\varepsilon, \varepsilon^{-1}] \times [-\varepsilon^{-1}, \varepsilon^{-1}]$, where $k \in \{1,2\}$, such that $|t_2 - t_1| \geq L$ and $|s_2 - s_1| \geq L$.

Note that in Lemma 1.4, the pairs t_1 and t_2, s_1 and s_2, are well separated. The following lemma is concerned with the case when $t_1 = t_2$ and $s_1 = s_2$.

Lemma 1.5. *Let $\varepsilon \in (0,1)$ and $L > 0$ be given constants. Then there exist positive constants $h_0 \in (0, \frac{L}{2})$ and $c_{1,6}$ such that*

$$
\mathrm{Var}\Big(U_t^j(x_2) - U_t^j(x_1)\Big|U_s^j(y_2) - U_s^j(y_1)\Big) \geq c_{1,6}\big(|s-t|^{1/2} + |x_1-y_1| + |x_2-y_2|\big) \tag{16}
$$

for all $s, t \in [\varepsilon, \varepsilon^{-1}]$ with $|s-t| \leq h_0$ and all x_k, $y_k \in [-\varepsilon^{-1}, \varepsilon^{-1}]$, where $k \in \{1,2\}$, such that $|x_2 - x_1| \geq L$, $|y_2 - y_1| \geq L$ and $|x_k - y_k| \leq \frac{L}{2}$ for $k = 1, 2$.

Remark 1.6. Note that, in the above, it is essential to only consider those $s, t \in [\varepsilon, \varepsilon^{-1}]$ such that $|s-t|$ is small. Otherwise (16) does not hold as indicated by (5). In this sense, Lemma 1.5 is more restrictive than Lemma 1.4. But it is sufficient for the proof of Theorem 4.2.

Proof. Using the notation similar to that in [9], we let $(X, Y) = \big(U_t^j(x_2) - U_t^j(x_1), U_s^j(y_2) - U_s^j(y_1)\big)$ and write $\sigma_X^2 = \mathbb{E}(X^2)$, $\sigma_Y^2 = \mathbb{E}(Y^2)$ and $\rho_{X,Y}^2 = \mathbb{E}\big[(X-Y)^2\big]$. Recall that, for the Gaussian vector (X, Y), we have

$$\operatorname{Var}(X|Y) = \frac{\big(\rho_{X,Y}^2 - (\sigma_X - \sigma_Y)^2\big)\big((\sigma_X + \sigma_Y)^2 - \rho_{X,Y}^2\big)}{4\sigma_Y^2}. \tag{17}$$

Lemma 1.1 and the separation condition on x_k and y_k imply that both σ_X^2 and σ_Y^2 are bounded from above and below by positive constants. Similar to the proofs of Lemmas 3 and 4 in [9], we only need to derive a suitable lower bound for $\rho_{X,Y}^2$. By using the identity

$$(a-b+c-d)^2 = (a-b)^2 + (c-d)^2 + (a-d)^2 + (b-c)^2 - (a-c)^2 - (b-d)^2$$

and (5) we have

$$\begin{aligned} \rho_{X,Y}^2 &= |t-s|^{1/2} F\big(|x_2 - y_2||t-s|^{-1/2}\big) + |t-s|^{1/2} F\big(|y_1 - x_1||t-s|^{-1/2}\big) \\ &\quad + |x_2 - x_1| - |t-s|^{1/2} F\big(|x_2 - y_1||t-s|^{-1/2}\big) \\ &\quad + |y_1 - y_2| - |t-s|^{1/2} F\big(|x_1 - y_2||t-s|^{-1/2}\big). \end{aligned} \tag{18}$$

By (9), we can rewrite the above equation as

$$\begin{aligned} \rho_{X,Y}^2 &= |t-s|^{1/2} H\big(|x_2 - y_2||t-s|^{-1/2}\big) \\ &\quad + |t-s|^{1/2} H\big(|y_1 - x_1||t-s|^{-1/2}\big) \\ &\quad + |x_2 - x_1| - |t-s|^{1/2} H\big(|x_2 - y_1||t-s|^{-1/2}\big) \\ &\quad + |y_1 - y_2| - |t-s|^{1/2} H\big(|x_1 - y_2||t-s|^{-1/2}\big). \end{aligned} \tag{19}$$

Denote the algebraic sum of the last four terms in (19) by S and we need to derive a lower bound for it. Note that, under the conditions of our lemma, $|x_2 - y_1| \ge \frac{L}{2}$ and $|x_1 - y_2| \ge \frac{L}{2}$. Hence Lemma 1.2 implies that, for any $0 < \delta < \frac{c_{1,2}}{2}$, there exists a constant $h_0 \in (0, \frac{L}{2})$ such that

$$|t-s|^{1/2} H\big(|x_2 - y_1||t-s|^{-1/2}\big) \le |x_2 - y_1| + \frac{\delta}{2}\,|t-s|^{1/2} \tag{20}$$

whenever $|t-s| \le h_0$; and the same inequality holds when $|x_2 - y_1|$ is replaced by $|x_1 - y_2|$. It follows that

$$\begin{aligned} S &\ge \big(|x_2 - x_1| - |x_2 - y_1| + |y_1 - y_2| - |x_1 - y_2|\big) - \delta\,|t-s|^{1/2} \\ &= -\delta\,|t-s|^{1/2}, \end{aligned} \tag{21}$$

because the sum of the four terms in the parentheses equals 0 under the separation condition. Combining (19) and (11) yields

$$\rho_{X,Y}^2 \ge \frac{c_{1,2}}{2}\,\big(|t-s|^{1/2} + |x_1 - y_1| + |x_2 - y_2|\big) \tag{22}$$

whenever x_k, y_k $(k = 1, 2)$ satisfy the above conditions.

By (5), we have $(\sigma_X - \sigma_Y)^2 \le c\,\big(|y_1 - x_1| + |x_2 - y_2|\big)^2$. It follows from (17) and (22) that (16) holds whenever $|y_1 - x_1| + |x_2 - y_2|$ is sufficiently small. Finally, a continuity argument as in [9], p. 15 removes this last restriction. This finishes the proof of Lemma 1.5. □

The present paper is a continuation of the paper of Mueller and Tribe [9]. Our objective is to study the fractal properties of various random sets generated by the random string processes. In Section 2, we determine the Hausdorff and packing dimensions of the range $u\big([0,1]^2\big)$ and the graph $\mathrm{Gr}u\big([0,1]^2\big)$. We also consider the Hausdorff dimension of the range $u(E)$, where $E \subseteq [0,\infty)\times\mathbb{R}$ is an arbitrary Borel set. In Section 3, we consider the existence of the local times of the random string process and determine the Hausdorff and packing dimensions of the level set $L_{\mathbf{u}} = \{(t,x) \in (0,\infty)\times\mathbb{R} : u_t(x) = \mathbf{u}\}$, where $\mathbf{u} \in \mathbb{R}^d$. Finally, we conclude our paper by determining the Hausdorff and packing dimensions of the sets of two kinds of double times of the random string in Section 4.

2. Dimension results of the range and graph

In this section, we study the Hausdorff and packing dimensions of the range $u\big([0,1]^2\big) = \big\{u_t(x) : (t,x) \in [0,1]^2\big\} \subset \mathbb{R}^d$ and the graph $\mathrm{Gr}u\big([0,1]^2\big) = \big\{ ((t,x),u_t(x)) : (t,x) \in [0,1]^2\big\} \subset \mathbb{R}^{2+d}$. We refer to Falconer [4] for the definitions and properties of Hausdorff dimension $\dim_{\mathrm{H}}(\cdot)$ and packing dimension $\dim_{\mathrm{P}}(\cdot)$.

Theorem 2.1. *Let $\{u_t(x) : t \geq 0,\ x \in \mathbb{R}\}$ be a random string process taking values in $\mathbb{R}^d$. Then with probability* 1*,*

$$\dim_{\mathrm{H}} u\big([0,1]^2\big) = \min\big\{d;\, 6\big\} \tag{23}$$

and

$$\dim_{\mathrm{H}} \mathrm{Gr}u\big([0,1]^2\big) = \begin{cases} 2+\frac{3}{4}d & \text{if } 1 \leq d < 4,\\ 3+\frac{1}{2}d & \text{if } 4 \leq d < 6,\\ 6 & \text{if } 6 \leq d. \end{cases} \tag{24}$$

Proof. Corollary 2 of Mueller and Tribe [9] states that the distributions of $\{u_t(x) : t \geq 0,\ x \in \mathbb{R}\}$ and the stationary pinned string $U = \{U_t(x) : t \geq 0,\ x \in \mathbb{R}\}$ are mutually absolutely continuous. Hence it is enough for us to prove (23) and (24) for the stationary pinned string $U = \{U_t(x) : t \geq 0,\ x \in \mathbb{R}\}$. This is similar to the proof of Theorem 4 of Ayache and Xiao [2]. We include a self-contained proof for reader's convenience.

As usual, the proof is divided into proving the upper and lower bounds separately. For the upper bound in (23), we note that clearly $\dim_{\mathrm{H}} U\big([0,1]^2\big) \leq d$ a.s., so we only need to prove the following inequality:

$$\dim_{\mathrm{H}} U\big([0,1]^2\big) \leq 6 \qquad \text{a.s.} \tag{25}$$

Because of Lemma 1.1, one can use the standard entropy method for estimating the tail probabilities of the supremum of a Gaussian process to establish the modulus of continuity of $U = \{U_t(x) : t \geq 0,\ x \in \mathbb{R}\}$. See, for example, Kôno [8]. It follows that, for any constants $0 < \gamma_1 < \gamma_1' < 1/4$ and $0 < \gamma_2 < \gamma_2' < 1/2$, there exist a random variable $A > 0$ of finite moments of all orders and an event Ω_1 of probability 1 such that for all $\omega \in \Omega_1$,

$$\sup_{(s,y),(t,x)\in[0,1]^2} \frac{|U_s(y,\omega) - U_t(x,\omega)|}{|s-t|^{\gamma_1'} + |x-y|^{\gamma_2'}} \leq A(\omega). \tag{26}$$

Let $\omega \in \Omega_1$ be fixed and then suppressed. For any integer $n \geq 2$, we divide $[0,1]^2$ into n^6 sub-rectangles $\{R_{n,i}\}$ with sides parallel to the axes and side-lengths

n^{-4} and n^{-2}, respectively. Then $U\big([0,1]^2\big)$ can be covered by the sets $U(R_{n,i})$ $(1 \le i \le n^6)$. By (26), we see that the diameter of the image $U(R_{n,i})$ satisfies

$$\operatorname{diam} U(R_{n,i}) \le c_{2,1}\, n^{-1+\delta}, \tag{27}$$

where $\delta = \max\{1-4\gamma_1', 1-2\gamma_2'\}$. We choose $\gamma_1' \in (\gamma_1, 1/4)$ and $\gamma_2' \in (\gamma_2, 1/2)$ such that

$$(1-\delta)\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}\right) > 6.$$

Hence, for $\gamma = \frac{1}{\gamma_1}+\frac{1}{\gamma_2}$, it follows from (27) that

$$\sum_{i=1}^{n^6}\Big[\operatorname{diam} U(R_{n,i})\Big]^\gamma \le c_{2,2}\, n^6\, n^{-(1-\delta)\gamma} \to 0 \tag{28}$$

as $n \to \infty$. This implies that $\dim_{\rm H} U\big([0,1]^2\big) \le \gamma$ a.s. By letting $\gamma_1 \uparrow 1/4$ and $\gamma_2 \uparrow 1/2$ along rational numbers, respectively, we derive (25).

Now we turn to the proof of the upper bound in (24) for the stationary pinned string U. We will show that there are three different ways to cover $\mathrm{Gr} U\big([0,1]^2\big)$, each of which leads to an upper bound for $\dim_{\rm H} \mathrm{Gr} U\big([0,1]^2\big)$.

- For each fixed integer $n \ge 2$, we have

$$\mathrm{Gr} U\big([0,1]^2\big) \subseteq \bigcup_{i=1}^{n^6} R_{n,i} \times U(R_{n,i}). \tag{29}$$

  It follows from (27) and (29) that $\mathrm{Gr} U\big([0,1]^2\big)$ can be covered by n^6 cubes in $\mathbb{R}^{2+d}$ with side-lengths $c_{2,3}\, n^{-1+\delta}$ and the same argument as the above yields

$$\dim_{\rm H} \mathrm{Gr} U\big([0,1]^2\big) \le 6 \quad \text{a.s.} \tag{30}$$

- Observe that each $R_{n,i} \times U(R_{n,i})$ can be covered by $\ell_{n,1}$ cubes in $\mathbb{R}^{2+d}$ of sides n^{-4}, where by (26)

$$\ell_{n,1} \le c_{2,4}\, n^2 \times \left(\frac{n^{-1+\delta}}{n^{-4}}\right)^d.$$

  Hence $\mathrm{Gr} U\big([0,1]^2\big)$ can be covered by $n^6 \times \ell_{n,1}$ cubes in $\mathbb{R}^{2+d}$ with sides n^{-4}. Denote

$$\eta_1 = 2 + (1-\gamma_1)d.$$

  Recall from the above that we can choose the constants γ_1, γ'_1 and γ'_2 such that $1-\delta > 4\gamma_1$. Therefore

$$n^6 \times \ell_{n,1} \times \big(n^{-4}\big)^{\eta_1} \le c_{2,5}\, n^{-(1-\delta-4\gamma_1)d} \to 0$$

  as $n \to \infty$. This implies that $\dim_{\rm H} \mathrm{Gr} U\big([0,1]^2\big) \le \eta_1$ almost surely. Hence,

$$\dim_{\rm H} \mathrm{Gr} U\big([0,1]^2\big) \le 2 + \frac{3}{4}d, \qquad \text{a.s.} \tag{31}$$

- We can also cover each $R_{n,i} \times U(R_{n,i})$ by $\ell_{n,2}$ cubes in $\mathbb{R}^{2+d}$ of sides n^{-2}, where by (26)
$$\ell_{n,2} \le c_{2,6} \left(\frac{n^{-1+\delta}}{n^{-2}}\right)^d.$$
Hence $\mathrm{Gr}U\big([0,1]^2\big)$ can be covered by $n^6 \times \ell_{n,2}$ cubes in $\mathbb{R}^{2+d}$ with sides n^{-2}. Denote $\eta_2 = 3 + (1-\gamma_2)d$. Recall from the above that we can choose the constants γ_2, γ'_1 and γ'_2 such that $1-\delta > 2\gamma_2$. Therefore
$$n^6 \times \ell_{n,2} \times \big(n^{-2}\big)^{\eta_2} \le c_{2,7}\, n^{-(1-\delta-2\gamma_2)d} \to 0$$
as $n \to \infty$. This implies that $\dim_{\mathrm{H}} \mathrm{Gr}U\big([0,1]^2\big) \le \eta_2$ almost surely. Hence,
$$\dim_{\mathrm{H}} \mathrm{Gr}U\big([0,1]^2\big) \le 3 + \frac{1}{2}d, \qquad \text{a.s.} \tag{32}$$

Combining (30), (31) and (32) yields
$$\dim_{\mathrm{H}} \mathrm{Gr}U\big([0,1]^2\big) \le \min\left\{6,\, 2+\frac{3}{4}d,\, 3+\frac{1}{2}d\right\}, \quad \text{a.s.} \tag{33}$$
and the upper bounds in (24) follow from (33).

To prove the lower bound in (23), by Frostman's theorem it is sufficient to show that for any $0 < \gamma < \min\{d, 6\}$,
$$\mathcal{E}_\gamma = \int_{[0,1]^2}\int_{[0,1]^2} \mathbb{E}\left(\frac{1}{|U_s(y) - U_t(x)|^\gamma}\right) dsdydtdx < \infty. \tag{34}$$
See, e.g., [7], Chapter 10. Since $0 < \gamma < d$, we have $0 < \mathbb{E}(|\Xi|^{-\gamma}) < \infty$, where Ξ is a standard d-dimensional normal vector. Combining this fact with Lemma 1.1, we have
$$\mathcal{E}_\gamma \le c_{2,8} \int_0^1 ds \int_0^1 dt \int_0^1 dy \int_0^1 \frac{1}{\big(|s-t|^{1/2} + |x-y|\big)^{\gamma/2}}\, dx. \tag{35}$$
Recall the weighted arithmetic-mean and geometric-mean inequality: for all integer $n \ge 2$ and $x_i \ge 0$, $\beta_i > 0$ $(i = 1, \ldots, n)$ such that $\sum_{i=1}^n \beta_i = 1$, we have
$$\prod_{i=1}^n x_i^{\beta_i} \le \sum_{i=1}^n \beta_i x_i. \tag{36}$$
Applying (36) with $n = 2$, $\beta_1 = 2/3$ and $\beta_2 = 1/3$, we obtain
$$|s-t|^{1/2} + |x-y| \ge \frac{2}{3}|s-t|^{1/2} + \frac{1}{3}|x-y| \ge |s-t|^{1/3}|x-y|^{1/3}. \tag{37}$$
Therefore, the denominator in (35) can be bounded from below by $|s-t|^{\gamma/6}|x-y|^{\gamma/6}$. Since $\gamma < 6$, by (35), we have $\mathcal{E}_\gamma < \infty$, which proves (34).

For proving the lower bound in (24), we need the following lemma from Ayache and Xiao [2].

Lemma 2.2. *Let α, β and η be positive constants. For $a > 0$ and $b > 0$, let*
$$J := J(a,b) = \int_0^1 \frac{dt}{(a+t^\alpha)^\beta (b+t)^\eta}. \tag{38}$$
Then there exist finite constants $c_{2,9}$ and $c_{2,10}$, depending on α, β, η only, such that the following hold for all reals $a, b > 0$ satisfying $a^{1/\alpha} \le c_{2,9}\, b$:

(i) *if* $\alpha\beta > 1$, *then*

$$J \le c_{2,10}\, \frac{1}{a^{\beta-\alpha^{-1}} b^{\eta}}; \tag{39}$$

(ii) *if* $\alpha\beta = 1$, *then*

$$J \le c_{2,10}\, \frac{1}{b^{\eta}} \log\big(1 + b a^{-1/\alpha}\big); \tag{40}$$

(iii) *if* $0 < \alpha\beta < 1$ *and* $\alpha\beta + \eta \neq 1$, *then*

$$J \le c_{2,10}\, \Big(\frac{1}{b^{\alpha\beta+\eta-1}} + 1\Big). \tag{41}$$

Now we prove the lower bound in (24). Since $\dim_{\mathrm{H}} \mathrm{Gr}U\big([0,1]^2\big) \ge \dim_{\mathrm{H}} U\big([0,1]^2\big)$ always holds, we only need to consider the cases $1 \le d < 4$ and $4 \le d < 6$, respectively.

Since the proof of the two cases are almost identical, we only prove the case when $1 \le d < 4$ here. Let $0 < \gamma < 2 + \frac{3}{4}d$ be a fixed, but arbitrary, constant. Since $1 \le d < 4$, we may and will assume $\gamma > 1+d$. In order to prove $\dim_{\mathrm{H}} \mathrm{Gr}U([0,1]^2) \ge \gamma$ a.s., again by Frostman's theorem, it is sufficient to show

$$\begin{aligned} \mathcal{G}_\gamma = \int_{[0,1]^2}\int_{[0,1]^2} \mathbb{E}\Bigg[\frac{1}{\big(|s-t|^2 + |x-y|^2 + |U_s(y) - U_t(x)|^2\big)^{\gamma/2}}\Bigg]\, dsdydtdx \\ < \infty. \end{aligned} \tag{42}$$

Since $\gamma > d$, we note that for a standard normal vector Ξ in $\mathbb{R}^d$ and any number $a \in \mathbb{R}$,

$$\mathbb{E}\Bigg[\frac{1}{\big(a^2 + |\Xi|^2\big)^{\gamma/2}}\Bigg] \le c_{2,11}\, a^{-(\gamma-d)},$$

see e.g. [7], p. 279. Consequently, by Lemma 1.1, we derive that

$$\mathcal{G}_\gamma \le c_{2,12} \int_{[0,1]^2}\int_{[0,1]^2} \frac{1}{\big(|s-t|^{1/2} + |x-y|\big)^{d/2}\, \big(|s-t| + |x-y|\big)^{\gamma-d}}\, dsdydtdx. \tag{43}$$

By Lemma 2.2 and a change of variable and noting that $d < 4$, we can apply (41) to derive

$$\begin{aligned} \mathcal{G}_\gamma &\le c_{2,13} \int_0^1 dx \int_0^1 \frac{1}{(t^{1/2} + x)^{d/2} (t+x)^{\gamma-d}} dt \\ &\le c_{2,14} \int_0^1 \Big(\frac{1}{x^{d/4+\gamma-d-1}} + 1\Big)\, dx < \infty, \end{aligned} \tag{44}$$

where the last inequality follows from $\gamma - \frac{3}{4}d - 1 < 1$. This completes the proof of Theorem 2.1. □

By using the relationships among the Hausdorff dimension, packing dimension and the box dimension (see Falconer [4]), Theorem 2.1 and the proof of the upper bounds, we derive the following analogous result on the packing dimensions of $u\big([0,1]^2\big)$ and $\mathrm{Gr}u\big([0,1]^2\big)$.

Theorem 2.3. *Let* $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ *be a random string process taking values in* $\mathbb{R}^d$*. Then with probability* 1,

$$\dim_{\mathrm{P}} u\big([0,1]^2\big) = \min\big\{d;\, 6\big\} \tag{45}$$

and

$$\dim_{\mathrm{P}} \mathrm{Gr}u\big([0,1]^2\big) = \begin{cases} 2 + \frac{3}{4}d & \text{if } 1 \leq d < 4, \\ 3 + \frac{1}{2}d & \text{if } 4 \leq d < 6, \\ 6 & \text{if } 6 \leq d. \end{cases} \tag{46}$$

Theorems 2.1 and 2.3 show that the random fractals $u\big([0,1]^2\big)$ and $\mathrm{Gr}u\big([0,1]^2\big)$ are rather regular because they have the same Hausdorff and packing dimensions.

Now we will turn our attention to find the Hausdorff dimension of the range $u(E)$ for an arbitrary Borel set $E \subseteq [0,\infty) \times \mathbb{R}$.

For this purpose, we mention the related results of Wu and Xiao [13] for an (N,d)-fractional Brownian sheet $B^H = \{B^H(\mathbf{t}) : \mathbf{t} \in \mathbb{R}^N_+\}$ with Hurst index $H = (H_1, \ldots, H_N) \in (0,1)^N$. What the random string process $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ and a $(2,d)$-fractional Broanian sheet B^H with $H = (\frac{1}{4}, \frac{1}{2})$ have in common is that they are both anisotropic.

As Wu and Xiao [13] pointed out, the Hausdorff dimension of the image $B^H(F)$ cannot be determined by $\dim_{\mathrm{H}} F$ and H alone for an arbitrary fractal set F, and more information about the geometry of F is needed. To capture the anisotropic nature of B^H, they have introduced a new notion of dimension, namely, the *Hausdorff dimension contour*, for finite Borel measures and Borel sets and showed that $\dim_{\mathrm{H}} B^H(F)$ is determined by the Hausdorff dimension contour of F. It turns out that we can use the same technique to study the images of the random string.

We start with the following Proposition 2.4 which determines $\dim_{\mathrm{H}} u(E)$ when E belongs to a special class of Borel sets in $[0,\infty) \times \mathbb{R}$. Its proof is the same as that of Proposition 3.1 in [13].

Proposition 2.4. *Let* $\{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ *be a random string in* $\mathbb{R}^d$*. Assume that* E_1 *and* E_2 *are Borel sets in* $[0,\infty)$ *and* $\mathbb{R}$*, respectively, which satisfy* $\dim_{\mathrm{H}} E_1 = \dim_{\mathrm{P}} E_1$ *or* $\dim_{\mathrm{H}} E_2 = \dim_{\mathrm{P}} E_2$*. Let* $E = E_1 \times E_2 \subset [0,\infty) \times \mathbb{R}$*, then we have*

$$\dim_{\mathrm{H}} u(E) = \min\{d;\, 4\dim_{\mathrm{H}} E_1 + 2\dim_{\mathrm{H}} E_2\}, \quad a.s. \tag{47}$$

In order to determine $\dim_{\mathrm{H}} u(E)$ for an arbitrary Borel set $E \subset [0,\infty) \times \mathbb{R}$, we recall from [13] the following definition. Denote by $\mathcal{M}^+_c(E)$ the family of finite Borel measures with compact support in E.

Definition 2.5. Given $\mu \in \mathcal{M}^+_c(E)$, we define the set $\Lambda_\mu \subseteq \mathbb{R}^2_+$ by

$$\begin{aligned} \Lambda_\mu = \Big\{\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2_+ : \limsup_{r \to 0^+} \frac{\mu\left(R((t,x),r)\right)}{r^{4\lambda_1 + 2\lambda_2}} = 0, \\ \text{for } \mu\text{-a.e. } (t,x) \in [0,\infty) \times \mathbb{R}\Big\}, \end{aligned} \tag{48}$$

where $R((t,x),r) = [t - r^4, t + r^4] \times [x - r^2, x + r^2]$.

The properties of set Λ_μ can be found in Lemma 3.6 of Wu and Xiao [13]. The boundary of Λ_μ, denoted by $\partial\Lambda_\mu$, is called the Hausdorff dimension contour of μ.

Define

$$\Lambda(E) = \bigcup_{\mu \in \mathcal{M}_c^+(E)} \Lambda_\mu.$$

and define the Hausdorff dimension contour of E by $\bigcup_{\mu \in \mathcal{M}_c^+(E)} \partial\Lambda_\mu$. It can be verified that, for every $\mathbf{b} \in (0,\infty)^2$, the supremum $\sup_{\lambda \in \Lambda(E)} \langle \lambda, \mathbf{b} \rangle$ is achieved on the Hausdorff dimension contour of E (Lemma 3.6, [13]).

Theorem 2.6. *Let $u = \{u_t(x) : t \geq 0,\, x \in \mathbb{R}\}$ be a random string process with values in $\mathbb{R}^d$. Then, for any Borel set $E \subset [0,\infty) \times \mathbb{R}$,*

$$\dim_{\mathrm{H}} u(E) = \min\big\{d;\, s(E)\big\}, \qquad a.s. \tag{49}$$

where $s(E) = \sup_{\lambda \in \Lambda(E)} (4\lambda_1 + 2\lambda_2)$.

Proof. By Corollary 2 of Mueller and Tribe [9], one only needs to prove (49) for the stationary pinned string $U = \{U_t(x) : t \geq 0,\, x \in \mathbb{R}\}$. The latter follows from the proof of Theorem 3.10 in [13]. □

3. Existence of the local times and dimension results for level sets

In this section, we will first give a sufficient condition for the existence of the local times of a random string process on any rectangle $I \in \mathcal{A}$, where $\mathcal{A}$ is the collection of all the rectangles in $[0,\infty) \times \mathbb{R}$ with sides parallel to the axes. Then, we will determine the Hausdorff and packing dimensions for the level set $L_{\mathbf{u}} = \{(t,x) \in [0,\infty) \times \mathbb{R} : u_t(x) = \mathbf{u}\}$, where $\mathbf{u} \in \mathbb{R}^d$ is fixed.

We start by briefly recalling some aspects of the theory of local times. For an excellent survey on local times of random and deterministic vector fields, we refer to Geman and Horowitz [6].

Let $X(\mathbf{t})$ be a Borel vector field on $\mathbb{R}^N$ with values in $\mathbb{R}^d$. For any Borel set $T \subseteq \mathbb{R}^N$, the occupation measure of X on T is defined as the following measure on $\mathbb{R}^d$:

$$\mu_T(\bullet) = \lambda_N\big\{\mathbf{t} \in T : X(\mathbf{t}) \in \bullet\big\}.$$

If μ_T is absolutely continuous with respect to λ_d, the Lebesgue measure on $\mathbb{R}^d$, we say that $X(\mathbf{t})$ has *local times* on T, and define its local time $l(\bullet, T)$ as the Radon–Nikodým derivative of μ_T with respect to λ_d, i.e.,

$$l(\mathbf{u}, T) = \frac{d\mu_T}{d\lambda_d}(\mathbf{u}), \qquad \forall \mathbf{u} \in \mathbb{R}^d.$$

In the above, $\mathbf{u}$ is the so-called *space variable*, and T is the *time* variable. Sometimes, we write $l(\mathbf{u}, \mathbf{t})$ in place of $l(\mathbf{u}, [0, \mathbf{t}])$. It is clear that if X has local times on T, then for every Borel set $S \subseteq T$, $l(\mathbf{u}, S)$ also exists.

By standard martingale and monotone class arguments, one can deduce that the local times have a measurable modification that satisfies the following *occupation density formula*: for every Borel set $T \subseteq \mathbb{R}^N$, and for every measurable function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\int_T f(X(\mathbf{t}))\, d\mathbf{t} = \int_{\mathbb{R}^d} f(\mathbf{u}) l(\mathbf{u}, T)\, d\mathbf{u}. \tag{50}$$

The following theorem is concerned with the existence of local times of the random string.

Theorem 3.1. *Let $\{u_t(x):\ t\geq 0,\ x\in\mathbb{R}\}$ be a random string process in $\mathbb{R}^d$. If $d<6$, then for every $I\in\mathcal{A}$, the string has local times $\{l(\mathbf{u},I),\ \mathbf{u}\in\mathbb{R}^d\}$ on I, and $l(\mathbf{u},I)$ admits the following L^2 representation:*

$$l(\mathbf{u},I)=(2\pi)^{-d}\int_{\mathbb{R}^d}e^{-i\langle \mathbf{v},\,\mathbf{u}\rangle}\int_I e^{i\langle \mathbf{v},\,u_t(x)\rangle}dtdxd\mathbf{v},\quad \forall\ \mathbf{u}\in\mathbb{R}^d. \tag{51}$$

Proof. Because of Corollary 2 of Mueller and Tribe [9], we only need to prove the existence for the stationary pinned string $U=\{U_t(x):\ t\geq 0,\ x\in\mathbb{R}\}$.

Let $I\in\mathcal{A}$ be fixed. Without loss of generality, we may assume $I=[\varepsilon,1]^2$. By (21.3) in [6] and using the characteristic functions of Gaussian random variables, it suffices to prove

$$\mathcal{J}(I):=\int_I dtdx\int_I dsdy\int_{\mathbb{R}^d}d\mathbf{u}\int_{\mathbb{R}^d}\big|\mathbb{E}\exp\left(i\langle\mathbf{u},\,U_t(x)\rangle+i\langle\mathbf{v},\,U_s(y)\rangle\right)\big|d\mathbf{v}<\infty. \tag{52}$$

Since the components of U are i.i.d., it is easy to see that

$$\mathcal{J}(I)=(2\pi)^d\int_I dtdx\int_I\left[\mathrm{detCov}\left(U_t^1(x),\,U_s^1(y)\right)\right]^{-d/2}dsdy. \tag{53}$$

By Lemma 1.3 and noting that $I=[\varepsilon,1]^2$, we can see that

$$\begin{aligned}\mathrm{detCov}\big(U_t^j(x),\,U_s^j(y)\big)&=\mathrm{Var}\big(U_s^j(y)\big)\mathrm{Var}\big(U_t^j(x)\big|U_s^j(y)\big)\\&\geq c_{3,1}\left(|x-y|+|t-s|^{1/2}\right).\end{aligned} \tag{54}$$

The above inequality, (37) and the fact that $d<6$ lead to

$$\mathcal{J}(I)\leq c_{3,2}\int_\varepsilon^1\int_\varepsilon^1|s-t|^{-d/6}dtds\int_\varepsilon^1\int_\varepsilon^1|x-y|^{-d/6}dxdy<\infty, \tag{55}$$

which proves (52), and therefore Theorem 3.1. □

Remark 3.2. It would be interesting to study the regularity properties of the local times $l(\mathbf{u},\mathbf{t})$, ($\mathbf{u}\in\mathbb{R}^d,\mathbf{t}\in[0,\infty)\times\mathbb{R}$) such as joint continuity and moduli of continuity. One way to tackle these problems is to establish sectorial local nondeterminism (see [13]) for the stationary pinned string $U=\{U_t(x):\ t\geq 0,\ x\in\mathbb{R}\}$. This will have to be pursued elsewhere. Some results of this nature for certain isotropic Gaussian random fields can be found in [15].

Mueller and Tribe [9] proved that for every $\mathbf{u}\in\mathbb{R}^d$,

$$\mathbb{P}\big\{u_t(x)=\mathbf{u}\ \text{ for some }\ (t,x)\in[0,\infty)\times\mathbb{R}\big\}>0$$

if and only if $d<6$. Now we study the Hausdorff and packing dimensions of the level set $L_{\mathbf{u}}=\{(t,x)\in[0,\infty)\times\mathbb{R}:\ u_t(x)=\mathbf{u}\}$.

Theorem 3.3. *Let $\{u_t(x):\ t\geq 0,\ x\in\mathbb{R}\}$ be a random string process in $\mathbb{R}^d$ with $d<6$. Then for every $\mathbf{u}\in\mathbb{R}^d$, with positive probability,*

$$\dim_{\mathrm{H}}\big(L_{\mathbf{u}}\cap[0,1]^2\big)=\dim_{\mathrm{P}}\big(L_{\mathbf{u}}\cap[0,1]^2\big)=\begin{cases}2-\frac14 d & \text{if }\ 1\leq d<4,\\ 3-\frac12 d & \text{if }\ 4\leq d<6.\end{cases} \tag{56}$$

Proof. As usual, it is sufficient to prove (56) for the stationary pinned string $U = \{U_t(x) : t \geq 0, x \in \mathbb{R}\}$. We first prove the almost sure upper bound

$$\dim_{\mathrm{P}}\big(L_{\mathbf{u}} \cap [0,1]^2\big) \leq \begin{cases} 2 - \frac{1}{4}d & \text{if } 1 \leq d < 4, \\ 3 - \frac{1}{2}d & \text{if } 4 \leq d < 6. \end{cases} \tag{57}$$

By the σ-stability of $\dim_{\mathrm{P}}$, it is sufficient to show (57) holds for $L_{\mathbf{u}} \cap [\varepsilon, 1]^2$ for every $\varepsilon \in (0,1)$. For this purpose, we construct coverings of $L_{\mathbf{u}} \cap [0,1]^2$ by cubes of the same side length.

For any integer $n \geq 2$, we divide the square $[\varepsilon, 1]^2$ into n^6 sub-rectangles $R_{n,\ell}$ of side lengths n^{-4} and n^{-2}, respectively. Let $0 < \delta < 1$ be fixed and let $\tau_{n,\ell}$ be the lower-left vertex of $R_{n,\ell}$. Then

$$\begin{aligned} \mathbb{P}\big\{\mathbf{u} \in U(R_{n,\ell})\big\} &\leq \mathbb{P}\bigg\{\max_{(s,y),(t,x) \in R_{n,\ell}} |U_s(y) - U_t(x)| \leq n^{-(1-\delta)};\ \mathbf{u} \in U(R_{n,\ell})\bigg\} \\ &\quad + \mathbb{P}\bigg\{\max_{(s,y),(t,x) \in R_{n,\ell}} |U_s(y) - U_t(x)| > n^{-(1-\delta)}\bigg\} \\ &\leq \mathbb{P}\Big\{|U(\tau_{n,\ell}) - \mathbf{u}| \leq n^{-(1-\delta)}\Big\} + e^{-c\, n^{2\delta}} \\ &\leq c_{3,3}\, n^{-(1-\delta)d}. \end{aligned} \tag{58}$$

In the above we have applied Lemma 1.1 and the Gaussian isoperimetric inequality (cf. Lemma 2.1 in [11]) to derive the second inequality.

Since we can deal with the cases $1 \leq d < 4$ and $4 \leq d < 6$ almost identically, we will only consider the case $1 \leq d < 4$ here and leave the case $4 \leq d < 6$ to the interested readers.

Define a covering $\{R'_{n,\ell}\}$ of $L_{\mathbf{u}} \cap [\varepsilon, 1]^2$ by $R'_{n,\ell} = R_{n,\ell}$ if $\mathbf{u} \in U(R_{n,\ell})$ and $R'_{n,\ell} = \emptyset$ otherwise. Note that each $R'_{n,\ell}$ can be covered by n^2 squares of side length n^{-4}. Thus, for every $n \geq 2$, we have obtained a covering of the level set $L_{\mathbf{u}} \cap [\varepsilon, 1]^2$ by squares of side length n^{-4}. Consider the sequence of integers $n = 2^k$ $(k \geq 1)$, and let N_k denote the minimum number of squares of side-length 2^{-4k} that are needed to cover $L_{\mathbf{u}} \cap [\varepsilon, 1]^2$. It follows from (58) that

$$\mathbb{E}(N_k) \leq c_{3,3}\, 2^{6k} \cdot 2^{2k} \cdot 2^{-k(1-\delta)d} = c_{3,3}\, 2^{k(8-(1-\delta)d)}. \tag{59}$$

By (59), Markov's inequality and the Bore-Cantelli lemma we derive that for any $\delta' \in (0, \delta)$, almost surely for all k large enough,

$$N_k \leq c_{3,3}\ 2^{k(8-(1-\delta')d)}. \tag{60}$$

By the definition of box dimension and its relation to $\dim_{\mathrm{P}}$ (cf. [4]), (60) implies that $\dim_{\mathrm{P}}\big(L_{\mathbf{u}} \cap [\varepsilon, 1]^2\big) \leq 2 - (1-\delta')d/4$ a.s. Since $\varepsilon > 0$ is arbitrary, we obtain the desired upper bound for $\dim_{\mathrm{P}}\big(L_{\mathbf{u}} \cap [\varepsilon, 1]^2\big)$ in the case $1 \leq d < 4$.

Since $\dim_{\mathrm{H}} E \leq \dim_{\mathrm{P}} E$ for all Borel sets $E \subset \mathbb{R}^2$, it remains to prove the following lower bound: for any $\varepsilon \in (0,1)$, with positive probability

$$\dim_{\mathrm{P}}\big(L_{\mathbf{u}} \cap [\varepsilon, 1]^2\big) \geq \begin{cases} 2 - \frac{1}{4}d & \text{if } 1 \leq d < 4, \\ 3 - \frac{1}{2}d & \text{if } 4 \leq d < 6. \end{cases} \tag{61}$$

We only prove (61) for $1 \leq d < 4$. The other case is similar and is omitted. Let $\delta > 0$ such that

$$\gamma := 2 - \frac{1}{4}(1+\delta)d > 1. \tag{62}$$

Note that if we can prove that there is a constant $c_{3,4} > 0$ such that

$$\mathbb{P}\big\{\dim_{\rm H}\big(L_{\mathbf{u}} \cap [\varepsilon, 1]^2\big) \geq \gamma\big\} \geq c_{3,4}, \tag{63}$$

then the lower bound in (61) will follow by letting $\delta \downarrow 0$.

Our proof of (63) is based on the capacity argument due to Kahane (see, e.g., [7]). Similar methods have been used by Adler [1], Testard [12], Xiao [14], Ayache and Xiao [2] to various types of stochastic processes.

Let $\mathcal{M}_\gamma^+$ be the space of all non-negative measures on $[0, 1]^2$ with finite γ-energy. It is known (cf. [1]) that $\mathcal{M}_\gamma^+$ is a complete metric space under the metric

$$\|\mu\|_\gamma = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \frac{\mu(dt, dx)\mu(ds, dy)}{\left(|t-s|^2 + |x-y|^2\right)^{\gamma/2}}. \tag{64}$$

We define a sequence of random positive measures μ_n on the Borel sets of $[\varepsilon, 1]^2$ by

$$\begin{aligned}\mu_n(C) &= \int_C (2\pi n)^{d/2} \exp\Big(-\frac{n\,|U_t(x) - \mathbf{u}|^2}{2}\Big) dtdx \\ &= \int_C\int_{\mathbb{R}^d} \exp\Big(-\frac{|\xi|^2}{2n} + i\langle \xi, U_t(x) - \mathbf{u}\rangle\Big) d\xi\, dtdx,\ \forall\ C \in \mathcal{B}([\varepsilon, 1]^2).\end{aligned} \tag{65}$$

It follows from Kahane [7] or Testard [12] that if there are positive constants $c_{3,5}$ and $c_{3,6}$, which may depend on $\mathbf{u}$, such that

$$\mathbb{E}\big(\|\mu_n\|\big) \geq c_{3,5}, \qquad \mathbb{E}\big(\|\mu_n\|^2\big) \leq c_{3,6}, \tag{66}$$

$$\mathbb{E}\big(\|\mu_n\|_\gamma\big) < +\infty, \tag{67}$$

where $\|\mu_n\| = \mu_n([\varepsilon, 1]^2)$, then there is a subsequence of $\{\mu_n\}$, say $\{\mu_{n_k}\}$, such that $\mu_{n_k} \to \mu$ in $\mathcal{M}_\gamma^+$ and μ is strictly positive with probability $\geq c_{3,5}^2/(2c_{3,6})$. It follows from (65) and the continuity of U that μ has its support in $L_{\mathbf{u}} \cap [\varepsilon, 1]^2$ almost surely. Hence Frostman's theorem yields (63).

It remains to verify (66) and (67). By Fubini's theorem we have

$$\begin{aligned}\mathbb{E}\big(\|\mu_n\|\big) &= \int_{[\varepsilon,1]^2}\int_{\mathbb{R}^d} e^{-i\langle \xi, \mathbf{u}\rangle} \exp\Big(-\frac{|\xi|^2}{2n}\Big)\,\mathbb{E}\exp\Big(i\langle \xi, U_t(x)\rangle\Big) d\xi\, dtdx \\ &= \int_{[\varepsilon,1]^2}\int_{\mathbb{R}^d} e^{-i\langle \xi, \mathbf{u}\rangle} \exp\Big(-\frac{1}{2}(n^{-1} + \sigma^2(t,x))|\xi|^2\Big)\, d\xi\, dtdx \\ &= \int_{[\varepsilon,1]^2} \Big(\frac{2\pi}{n^{-1} + \sigma^2(t,x)}\Big)^{d/2} \exp\Big(-\frac{|\mathbf{u}|^2}{2(n^{-1} + \sigma^2(t,x))}\Big)\, dtdx \\ &\geq \int_{[\varepsilon,1]^2} \Big(\frac{2\pi}{1 + \sigma^2(t,x)}\Big)^{d/2} \exp\Big(-\frac{|\mathbf{u}|^2}{2\sigma^2(t,x)}\Big)\, dt := c_{3,5},\end{aligned} \tag{68}$$

where $\sigma^2(t,x) = \mathbb{E}\Big[\big(U_t^1(x)\big)^2\Big]$.

Denote by I_{2d} the identity matrix of order $2d$ and by $\mathrm{Cov}(U_s(y), U_t(x))$ the covariance matrix of the Gaussian vector $(U_s(y), U_t(x))$. Let $\Gamma = n^{-1}I_{2d} + \mathrm{Cov}(U_s(y), U_t(x))$ and let $(\xi, \eta)'$ be the transpose of the row vector (ξ, η). As in the proof of

(52), we apply (14) in Lemma 1.3 and the inequality (36) to derive

$$
\begin{aligned}
&\mathbb{E}(\|\mu_n\|^2) \\
&= \int_{[\varepsilon,1]^2}\int_{[\varepsilon,1]^2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-i\langle \xi+\eta,\mathbf{u}\rangle}\,\exp\Big(-\frac{1}{2}(\xi,\eta)\,\Gamma\,(\xi,\eta)'\Big)\,d\xi d\eta\, dsdydtdx \\
&= \int_{[\varepsilon,1]^2}\int_{[\varepsilon,1]^2} \frac{(2\pi)^d}{\sqrt{\det\Gamma}}\,\exp\Big(-\frac{1}{2}(\mathbf{u},\mathbf{u})\,\Gamma^{-1}\,(\mathbf{u},\mathbf{u})'\Big)\,dsdy\,dtdx \\
&\le \int_{[\varepsilon,1]^N}\int_{[\varepsilon,1]^N} \frac{(2\pi)^d}{\big[\det\mathrm{Cov}(U_s^1(y),U_t^1(x))\big]^{d/2}}\,dsdy\,dtdx \\
&\le c_{3,7}\int_{\varepsilon}^1\int_{\varepsilon}^1 |s-t|^{-d/6}dtds\int_{\varepsilon}^1\int_{\varepsilon}^1 |x-y|^{-d/6}dxdy := c_{3,6} < \infty.
\end{aligned} \tag{69}
$$

Similar to (69) and by the same method as in proving (43), we have

$$
\begin{aligned}
\mathbb{E}(\|\mu_n\|_\gamma) &= \int_{[\varepsilon,1]^2}\int_{[\varepsilon,1]^2} \frac{dsdy\,dtdx}{(|s-t|^2+|x-y|^2)^{\gamma/2}} \\
&\qquad\times\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-i\langle \xi+\eta,\mathbf{u}\rangle}\,\exp\Big(-\frac{1}{2}(\xi,\eta)\,\Gamma\,(\xi,\eta)'\Big)\,d\xi d\eta \\
&\le c_{3,8}\int_{[\varepsilon,1]^2}\int_{[\varepsilon,1]^2} \frac{dsdy\,dtdx}{\big(|s-t|^{1/2}+|x-y|\big)^{d/2}\big(|s-t|+|x-y|\big)^{\gamma}} \\
&< \infty,
\end{aligned} \tag{70}
$$

where the last inequality follows from Lemma 2.2 and the facts that $d<4$ and $d/4+\gamma-1<1$. This proves (67) and thus the proof of Theorem 3.3 is finished. □

4. Hausdorff and packing dimensions of the sets of double times

Mueller and Tribe [9] found necessary and sufficient conditions for an $\mathbb{R}^d$-valued string process to have double points. In this section, we determine the Hausdorff and packing dimensions of the sets of double times of the random string.

As in [9], we consider the following two kinds of double times for the string process $\{u_t(x) : t\ge 0,\, x\in\mathbb{R}\}$.

- Type I double times:

$$
L_{\mathrm{I},2} = \Big\{\big((t_1,x_1),(t_2,x_2)\big)\in\big((0,\infty)\times\mathbb{R}\big)^2_{\neq} :\ u_{t_1}(x_1)=u_{t_2}(x_2)\Big\}, \tag{71}
$$

where

$$
\big((0,\infty)\times\mathbb{R}\big)^2_{\neq} = \big\{\big((t_1,x_1),(t_2,x_2)\big)\in\big((0,\infty)\times\mathbb{R}\big)^2 :\ (t_1,x_1)\neq(t_2,x_2)\big\}.
$$

In order to determine the Hausdorff and packing dimensions of $L_{\mathrm{I},2}$, we introduce a $(4,d)$-random field $\Delta u = \{\Delta u(t_1,x_1;t_2,x_2)\}$ defined by

$$
\Delta u(t_1,x_1;t_2,x_2) = u_{t_2}(x_2)-u_{t_1}(x_1),\ \forall (t_1,x_1,t_2,x_2)\in\big((0,\infty)\times\mathbb{R}\big)^2. \tag{72}
$$

Then $L_{\mathrm{I},2}$ can be viewed as the zero set of $\Delta u(t_1,x_1;t_2,x_2)$, denoted by $(\Delta u)^{-1}(0)$; and its Hausdorff and packing dimensions can be studied by using the method in Section 3.

- Type II double times:

$$L_{\mathrm{II},2} = \Big\{(t, x_1, x_2) \in (0, \infty) \times \mathbb{R}^2_{\neq} :\ u_t(x_1) = u_t(x_2)\Big\},$$

where $\mathbb{R}^2_{\neq} = \{(x_1, x_2) \in \mathbb{R}^2 :\ x_1 \neq x_2\}$.
In order to determine the Hausdorff and packing dimensions of $L_{\mathrm{II},2}$, we will consider the $(3, d)$-random field $\widetilde{\Delta} u = \{\widetilde{\Delta} u(t; x_1, x_2)\}$ defined by

$$\widetilde{\Delta} u(t; x_1, x_2) = u_t(x_2) - u_t(x_1), \qquad \forall\ (t, x_1, x_2) \in (0, \infty) \times \mathbb{R}^2. \tag{73}$$

Then we can see that $L_{\mathrm{II},2}$ is nothing but the zero set of $\widetilde{\Delta} u$:

$$L_{\mathrm{II},2} = \Big\{(t, x_1, x_2) \in (0, \infty) \times \mathbb{R}^2_{\neq} :\ \widetilde{\Delta} u(t; x_1, x_2) = 0\Big\}.$$

For any constants $0 < a_1 < a_2$ and $b_1 < b_2$, consider the squares $J_\ell = [a_\ell, a_\ell + h] \times [b_\ell, b_\ell + h]$ $(\ell = 1, 2)$. Let $J = \prod_{\ell=1}^2 J_\ell \subset ((0, \infty) \times \mathbb{R})^2$ denote the corresponding hypercube. We choose $h > 0$ small enough, say,

$$h < \min\left\{\frac{a_2 - a_1}{3},\ \frac{b_2 - b_1}{3}\right\} \equiv L.$$

Thus $|t_2 - t_1| > L$ for all $t_2 \in [a_2,\ a_2 + h]$ and $t_1 \in [a_1,\ a_1 + h]$. We will use this assumption together with Lemma 1.4 to prove Theorem 4.1 below. We denote the collection of the hypercubes having the above properties by $\mathcal{J}$.

The following theorem gives the Hausdorff and packing dimensions of the Type I double times of a random string.

Theorem 4.1. *Let $u = \{u_t(x) :\ t \geq 0,\ x \in \mathbb{R}\}$ be a random string process in $\mathbb{R}^d$. If $d \geq 12$, then $L_{\mathrm{I},2} = \emptyset$ a.s. If $d < 12$, then, for every $J \in \mathcal{J}$, with positive probability,*

$$\dim_{\mathrm{H}}\big(L_{\mathrm{I},2} \cap J\big) = \dim_{\mathrm{P}}\big(L_{\mathrm{I},2} \cap J\big) = \begin{cases} 4 - \frac{1}{4}d & \text{if } 1 \leq d < 8, \\ 6 - \frac{1}{2}d & \text{if } 8 \leq d < 12. \end{cases} \tag{74}$$

Proof. The first statement is due to Mueller and Tribe [9]. Hence, we only need to prove the dimension result (74).

Thanks to Corollary 2 of Mueller and Tribe [9], it is sufficient to prove (74) for the stationary pinned string U. This will be done by working with the zero set of the $(4, d)$-Gaussian field $\Delta U = \{\Delta U(t_1, x_1; t_2, x_2)\}$ define by (72). That is, we will prove (74) with $L_{\mathrm{I},2}$ replaced by the zero set $(\Delta U)^{-1}(0)$. The proof is a modification of that of Theorem 3.3. Hence, we only give a sketch of it.

For an integer $n \geq 2$, we divide the hypercube J into n^{12} sub-domains $T_{n,p} = R^1_{n,p} \times R^2_{n,p}$, where $R^1_{n,p}$, $R^2_{n,p} \subset (0, \infty) \times \mathbb{R}$ are rectangles of side lengths $n^{-4}h$ and $n^{-2}h$, respectively. Let $0 < \delta < 1$ be fixed and let $\tau^k_{n,p}$ be the lower-left vertex of $R^k_{n,p}$ $(k = 1, 2)$. Denote

$$\Delta V^{t_1, x_1; t_2, x_2}_{s_1, y_1; s_2, y_2} = \Delta U(t_1, x_1; t_2, x_2) - \Delta U(s_1, y_1; s_2, y_2),$$

then the probability $\mathbb{P}\big\{0 \in \Delta U(T_{n,p})\big\}$ is at most

$$\begin{aligned}
&\mathbb{P}\bigg\{\max_{(t_1,x_1;t_2,x_2),(s_1,y_1;s_2,y_2)\in T_{n,p}} \big|\Delta V^{t_1,x_1;t_2,x_2}_{s_1,y_1;s_2,y_2}\big| \le n^{-(1-\delta)};\ 0 \in \Delta U(T_{n,p})\bigg\} \\
&\quad + \mathbb{P}\bigg\{\max_{(t_1,x_1;t_2,x_2),(s_1,y_1;s_2,y_2)\in T_{n,p}} \big|\Delta V^{t_1,x_1;t_2,x_2}_{s_1,y_1;s_2,y_2}\big| > n^{-(1-\delta)}\bigg\} \\
&\le \mathbb{P}\Big\{\big|\Delta U(\tau^1_{n,p};\tau^2_{n,p})\big| \le n^{-(1-\delta)}\Big\} \\
&\quad + \mathbb{P}\bigg\{\max_{(t_1,x_1;t_2,x_2),(s_1,y_1;s_2,y_2)\in T_{n,p}} \big|\Delta V^{t_1,x_1;t_2,x_2}_{s_1,y_1;s_2,y_2}\big| > n^{-(1-\delta)}\bigg\}.
\end{aligned} \tag{75}$$

By the definition of J, we see that $\Delta U(\tau^1_{n,p}, \tau^2_{n,p})$ is a Gaussian random variable with mean 0 and variance at least $c\,L^{1/2}$. Hence the first term in (75) is at most $c_{4,1}\, n^{-(1-\delta)d}$.

On the other hand, since

$$\big|\Delta V^{t_1,x_1;t_2,x_2}_{s_1,y_1;s_2,y_2}\big| \le c \sum_{k=1}^{2} \big|U_{s_k}(y_k) - U_{t_k}(x_k)\big|,$$

we have

$$\begin{aligned}
&\mathbb{P}\bigg\{\max_{(t_1,x_1;t_2,x_2),(s_1,y_1;s_2,y_2)\in T_{n,p}} \big|\Delta V^{t_1,x_1;t_2,x_2}_{s_1,y_1;s_2,y_2}\big| > n^{-(1-\delta)}\bigg\} \\
&\quad \le \sum_{k=1}^{2} \mathbb{P}\bigg\{\max_{(s_k,y_k),(t_k,x_k)\in R^k_{n,p}} \big|U_{s_k}(y_k) - U_{t^k}(x_k)\big| > \frac{n^{-(1-\delta)}}{2c}\bigg\} \\
&\quad \le e^{-c_{4,2}\, n^{2\delta}},
\end{aligned} \tag{76}$$

where the last inequality follows from Lemma 1.1 and the Gaussian isoperimetric inequality (cf. Lemma 2.1 in [11]).

Combine (75) and (76), we have

$$\mathbb{P}\Big\{0 \in \Delta U(T_{n,p})\Big\} \le c_{4,1}\, n^{-(1-\delta)d} + e^{-c_{4,2}\, n^{2\delta}}. \tag{77}$$

Hence the same covering argument as in the proof of Theorem 3.3 yields the desired upper bound for $\dim_{\mathrm{P}}\big((\Delta U)^{-1}(0) \cap J\big)$. This proves the upper bounds in (74).

Now we prove the lower bound for the Hausdorff dimension of $(\Delta U)^{-1}(0) \cap J$. We will only consider the case $1 \le d < 8$ here and leave the case $8 \le d < 12$ to the interested readers.

Let $\delta > 0$ such that

$$\gamma := 4 - \frac{1}{4}(1+\delta)d > 2. \tag{78}$$

As in the proof of Theorem 3.3, it is sufficient to prove that there is a constant $c_{4,3} > 0$ such that

$$\mathbb{P}\big\{\dim_{\mathrm{H}}\big(L_{\mathrm{I},2} \cap J\big) \ge \gamma\big\} \ge c_{4,3}. \tag{79}$$

Let $\mathcal{N}^+_\gamma$ be the space of all non-negative measures on $[0,1]^4$ with finite γ-energy. Then $\mathcal{N}^+_\gamma$ is a complete metric space under the metric

$$\|\nu\|_\gamma = \int_{\mathbb{R}^4}\int_{\mathbb{R}^4} \frac{\nu(dt_1dx_1dt_2dx_2)\nu(ds_1dy_1ds_2dy_2)}{\big(|t_1-s_1|^2+|x_1-y_1|^2+|t_2-s_2|^2+|x_2-y_2|^2\big)^{\gamma/2}}; \tag{80}$$

see [1]. We define a sequence of random positive measures ν_n on the Borel set J by

$$\begin{aligned}\nu_n(C) &= \int_C (2\pi n)^{\frac{d}{2}} \exp\Big(- \frac{n\,|\Delta U(t_1,x_1;t_2,x_2)|^2}{2}\Big)\, dt_1dx_1dt_2dx_2 \\ &= \int_C \int_{\mathbb{R}^d} \exp\Big(- \frac{|\xi|^2}{2n} + i\langle \xi, \Delta U(t_1,x_1;t_2,x_2)\rangle\Big)\, d\xi\, dt_1dx_1dt_2dx_2.\end{aligned} \tag{81}$$

It follows from Kahane [7] or Testard [12] that (79) will follow if there are positive constants $c_{4,4}$ and $c_{4,5} > 0$ such that

$$\mathbb{E}\big(\|\nu_n\|\big) \ge c_{4,4}, \qquad \mathbb{E}\big(\|\nu_n\|^2\big) \le c_{4,5}, \tag{82}$$

$$\mathbb{E}\big(\|\nu_n\|_\gamma\big) < +\infty, \tag{83}$$

where $\|\nu_n\| = \nu_n(J)$.

The verifications of (82) and (83) are similar to those in the proof of Theorem 3.3. By Fubini's theorem we have

$$\begin{aligned}&\mathbb{E}(\|\nu_n\|)\\ &= \int_J \int_{\mathbb{R}^d} \exp\Big(- \frac{|\xi|^2}{2n}\Big)\, \mathbb{E}\exp\Big(i\langle \xi, \Delta U(t_1,x_1;t_2,x_2)\rangle\Big)\, d\xi\, dt_1dx_1dt_2dx_2 \\ &= \int_J \int_{\mathbb{R}^d} \exp\Big(- \frac{1}{2}\xi\big(n^{-1}\mathrm{I}_d + \mathrm{Cov}(\Delta U(t_1,x_1;t_2,x_2))\big)\xi'\Big)\, d\xi\, dt_1dx_1dt_2dx_2 \\ &= \int_J \frac{(2\pi)^{\frac{d}{2}}}{\sqrt{\det\big(n^{-1}\mathrm{I}_d + \mathrm{Cov}(\Delta U(t_1,x_1;t_2,x_2))\big)}}\, dt_1dx_1dt_2dx_2 \\ &\ge \int_J \frac{(2\pi)^{\frac{d}{2}}}{\sqrt{\det\big(\mathrm{I}_d + \mathrm{Cov}(\Delta U(t_1,x_1;t_2,x_2))\big)}}\, dt_1dx_1dt_2dx_2 := c_{4,4}.\end{aligned} \tag{84}$$

Denote by $\mathrm{Cov}\big(\Delta U(s_1,y_1;s_2,y_2), \Delta U(t_1,x_1;t_2,x_2)\big)$ the covariance matrix of the Gaussian vector $\big(\Delta U(s_1,y_1;s_2,y_2), \Delta U(t_1,x_1;t_2,x_2)\big)$ and let

$$\Gamma = n^{-1}\mathrm{I}_{2d} + \mathrm{Cov}\big(\Delta U(s_1,y_1;s_2,y_2), \Delta U(t_1,x_1;t_2,x_2)\big).$$

Then by the definition of J and (15) in Lemma 1.4, we have

$$\begin{aligned}&\mathbb{E}\big(\|\nu_n\|^2\big)\\ &= \int_J\int_J\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp\Big(- \frac{1}{2}(\xi,\eta)\,\Gamma\,(\xi,\eta)'\Big)\, d\xi d\eta\, ds_1dy_1ds_2dy_2dt_1dx_1dt_2dx_2 \\ &= \int_J\int_J \frac{(2\pi)^d}{\sqrt{\det\Gamma}}\, ds_1dy_1ds_2dy_2dt_1dx_1dt_2dx_2 \\ &\le \int_J\int_J \frac{(2\pi)^d ds_1dy_1ds_2dy_2dt_1dx_1dt_2dx_2}{\big[\det\mathrm{Cov}(\Delta U^1(s_1,y_1;s_2,y_2), \Delta U^1(t_1,x_1;t_2,x_2))\big]^{d/2}} \\ &\le c_{4,6} \int_J\int_J \frac{ds_1dy_1ds_2dy_2dt_1dx_1dt_2dx_2}{\big[|x_1-y_1| + |x_2-y_2| + |t_1-s_1|^{1/2} + |t_2-s_2|^{1/2}\big]^{d/2}} \\ &\le c_{4,7} \int_J\int_J \frac{dx_1dy_1dx_2dy_2dt_1ds_1dt_2ds_2}{\big[|x_1-y_1||x_2-y_2||t_1-s_1||t_2-s_2|\big]^{d/12}} := c_{4,5} < \infty,\end{aligned} \tag{85}$$

where the last inequality follows from $d < 12$. In the above, we have also applied the inequality (36) with $\beta_1 = \beta_2 = 1/6$ and $\beta_3 = \beta_4 = 1/3$.

Similar to (85) and by the same method as in proving (43), we have that $\mathbb{E}(\|\nu_n\|_\gamma)$ is, up to a constant factor, bounded by

$$
\begin{aligned}
&\int_J\int_J \frac{ds_1dy_1ds_2dy_2dt_1dx_1dt_2dx_2}{(|x_1-y_1|+|x_2-y_2|+|t_1-s_1|+|t_2-s_2|)^\gamma}\\
&\qquad\qquad \times \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp\Big(-\frac{1}{2}(\xi,\eta)\,\Gamma\,(\xi,\eta)'\Big)\,d\xi d\eta\\
&\le c_{4,8}\int_J\int_J \frac{1}{(|x_1-y_1|+|x_2-y_2|+|t_1-s_1|+|t_2-s_2|)^\gamma}\\
&\qquad\qquad \times \frac{dx_1dy_1dx_2dy_2dt_1ds_1dt_2ds_2}{\big(|x_1-y_1|+|x_2-y_2|+|t_1-s_1|^{1/2}+|t_2-s_2|^{1/2}\big)^{d/2}}\\
&\le c_{4,9}\int_0^1 dx_2\int_0^1 dx_1\int_0^1 dt_2\int_0^1 \frac{dt_1}{(t_1^{1/2}+t_2^{1/2}+x_1+x_2)^{d/2}(t_1+t_2+x_1+x_2)^\gamma}\\
&<\infty,
\end{aligned}
\tag{86}
$$

where the last inequality follows from Lemma 2.2, $d<8$ and the definition of γ [We need to consider three cases: $d<4$, $d=4$ and $4<d<8$, respectively]. This proves (83) and hence Theorem 4.1. □

For $a>0$ and $b_1<b_2$, let $K=[a,a+h]\times[b_1,b_1+h]\times[b_2,b_2+h]\subset(0,\infty)\times\mathbb{R}^2$. We choose $h>0$ small enough, say,

$$h<\frac{b_2-b_1}{3}\equiv\kappa.$$

Then $|x_2-x_1|>\kappa$ for all $x_2\in[b_2,\,b_2+h]$ and $x_1\in[b_1,\,b_1+h]$. We denote the collection of all the cubes K having the above properties by $\mathcal{K}$.

By using Lemma 1.5 and a similar argument as in the proof of Theorem 4.1, we can prove the following dimension result on $L_{\text{II},2}$. We leave the proof to the interested readers.

Theorem 4.2. *Let $u=\{u_t(x):\ t\ge 0,\ x\in\mathbb{R}\}$ be a random string process in $\mathbb{R}^d$. If $d\ge 8$, then $L_{\text{II},2}=\emptyset$ a.s. If $d<8$, then for every $K\in\mathcal{K}$, with positive probability,*

$$\dim_{\text{H}}\big(L_{\text{II},2}\cap K\big)=\dim_{\text{P}}\big(L_{\text{II},2}\cap K\big)=\begin{cases}3-\frac{1}{4}d & \text{if } 1\le d<4,\\ 4-\frac{1}{2}d & \text{if } 4\le d<8.\end{cases}\tag{87}$$

Remark 4.3. Rosen [10] studied k-multiple points of the Brownian sheet and multiparameter fractional Brownian motion by using their self-intersection local times. It would be interesting to establish similar results for the random string processes.

References

[1] Adler, R. J. (1981). *The Geometry of Random Fields.* Wiley, New York.

[2] Ayache, A. and Xiao, Y. (2005). Asymptotic properties and Hausdorff dimensions of fractional Brownian sheets. *J. Fourier Anal. Appl.* **11** 407–439.

[3] Ehm, W. (1981). Sample function properties of multi-parameter stable processes. *Z. Wahrsch. verw Gebiete* **56** 195–228.

[4] Falconer, K. J. (1990). *Fractal Geometry.* John Wiley & Sons Ltd., Chichester.

[5] Funaki, T. (1983). Random motion of strings and related stochastic evolution equations. *Nagoya Math. J.* **89** 129–193.

[6] Geman, D. and Horowitz, J. (1980). Occupation densities. *Ann. Probab.* **8** 1–67.

[7] Kahane, J.-P. (1985). *Some Random Series of Functions*, 2nd edition. Cambridge University Press, Cambridge.

[8] Kôno, N. (1975) Asymptotoc behavior of sample functions of Gaussian random fields. *J. Math. Kyoto Univ.* **15** 671–707.

[9] Mueller, C. and Tribe, R. (2002). Hitting probabilities of a random string. *Electronic J. Probab.* **7** (10) 1–29.

[10] Rosen, J. (1984). Self-intersections of random fields. *Ann. Probab.* **12** 108–119.

[11] Talagrand, M. (1995). Hausdorff measure of trajectories of multiparameter fractional Brownian motion. *Ann. Probab.* **23** 767–775.

[12] Testard, F. (1986). Polarité, points multiples et géométrie de certain processus gaussiens. Publ. du Laboratoire de Statistique et Probabilités de l'U.P.S., Toulouse, 01–86.

[13] Wu, D. and Xiao, Y. (2006). Geometric properties of fractional Brownian sheets. Submitted.

[14] Xiao, Y. (1995). Dimension results for Gaussian vector fields and index-α stable fields. *Ann. Probab.* **23** 273–291.

[15] Xiao, Y. (1997). Hölder conditions for the local times and the Hausdorff measure of the level sets of Gaussian random fields. *Probab. Theory Relat. Fields* **109** 129–157.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 148–154

DOI: 10.1214/074921706000000815

Random sets of isomorphism of linear operators on Hilbert space

Roman Vershynin[1,*]

University of California, Davis

Abstract: This note deals with a problem of the probabilistic Ramsey theory in functional analysis. Given a linear operator T on a Hilbert space with an orthogonal basis, we define the *isomorphic structure* $\Sigma(T)$ as the family of all subsets of the basis so that T restricted to their span is a nice isomorphism. Our main result is a dimension-free optimal estimate of the size of $\Sigma(T)$. It improves and extends in several ways the principle of restricted invertibility due to Bourgain and Tzafriri. With an appropriate notion of randomness, we obtain a randomized principle of restricted invertibility.

1. Introduction

1.1. Randomized Ramsey-type problems

Finding a nice structure in a big unstructured object is a recurrent theme in mathematics. This direction of thought is often called Ramsey theory, although Ramsey theory was originally only associated with combinatorics. One celebrated example is Van der Waerden's theorem: for any partition of the integers into two sets, one of these sets contains arbitrary long arithmetic progressions.

Ramsey theory meets probability theory when one asks about the quality of *most* sub-structures of a given structure. Can one improve the quality of a structure by passing to its *random* sub-structure? (a random subgraph, for example). A remarkable example of the randomized Ramsey theory is Dvoretzky's theorem in geometric functional analysis in the form of V.Milman (see [4], 4.2). One of its corollaries states that, for any n-dimensional finite-dimensional Banach space, a random $O(\log n)$-dimensional subspace (with respect to some natural measure) is well isomorphic to a Hilbert space.

1.2. The isomorphism structure of a linear operator

In this note we are trying to find a nice structure in an arbitrary bounded linear operator on a separable Hilbert space. Let T be a bounded linear operator on a Hilbert space H with an orthonormal basis $(e_i)_{i\in\mathbb{N}}$. We naturally think of T as being nice if it is a nice isomorphism on H. However, this situation is rather rare; instead, T may be a nice isomorphism on the subspace spanned by some subsets of the basis. So, instead of being a "global" isomorphism, T may be a "local" isomorphism when restricted to certain subspaces of H. A central question is then – how many such subspaces are there? Let us call these subspaces an isomorphism structure of T:

[1]Department of Mathematics, University of California, Davis, CA 95616, USA.

*The author is an Alfred P. Sloan Research Fellow. He was also supported by NSF DMS 0401032.

AMS 2000 subject classifications: 46B09, 47D25.

Keywords and phrases: restricted invertibility, operators on Hilbert spaces.

Definition 1.1. Let T be a bounded linear operator on a Hilbert space H, and $(e_i)_{i\in\mathbb{N}}$ be an orthonormal basis of H. Let $0 < \varepsilon < 1$. A set σ of $\mathbb{N}$ is called a *set of ε-isomorphism* of T if the equivalence

$$(1-\varepsilon)\sum_{i\in\sigma}\|a_iTe_i\|^2 \le \|\sum_{i\in\sigma}a_iTe_i\|^2 \le (1+\varepsilon)\sum_{i\in\sigma}\|a_iTe_i\|^2 \tag{1}$$

holds for every choice of scalars $(a_i)_{i\in\sigma}$. The *ε-isomorphism structure* $\Sigma(T,\varepsilon)$ consists of all such sets σ.

How big is the isomorphism structure? From the probabilistic point of view, we can ask for the probability that a random subset of (a finite interval of) the basis is the set of isomorphism. Unfortunately, this probability is in general exponentially small. For example, if T acts as $Te_i = e_{\lceil (i+1)/2\rceil}$, then every set of isomorphism contains no pairs of the form $\{2i-1,\, 2i\}$. Hence a random subset of a finite interval is unlikely to be a set of isomorphism of T. However, an appropriate notion of randomness yields a clean optimal bound on the size of the isomorphic structure. This is the main result of this note, which extends in several ways the Bourgain-Tzafriri's principle of the restricted invertibility [1], as we will see shortly.

Theorem 1.2. *Let T be a norm-one linear operator on a Hilbert space H, and let $0 < \varepsilon < 1$. Then there exists a probability measure ν on the isomorphism structure $\Sigma(T,\varepsilon)$, such that*

$$\nu\{\sigma \in \Sigma(T,\varepsilon) \mid i \in \sigma\} \ge c\varepsilon^2\|Te_i\|^2 \quad \textit{for all } i. \tag{2}$$

Here and thereafter $c, C, c_1, \dots$ denote positive absolute constants.

Theorem 1.2 gives a lower bound on the average of the characteristic functions of the sets of the isomorphism. Indeed, the left hand side in (2) clearly equals $\int_{\Sigma(T,\varepsilon)} \chi_\sigma(i)\, d\nu(\sigma)$. Thus, in absence of "true" randomness in the isomorphic structure $\Sigma(T,\varepsilon)$, we can still measure the size of $\Sigma(T,\varepsilon)$ by bounding below the *average of the characteristic functions* of its sets. It might be that considering this weak type of randomness might help in other problems, in which the usual, strong randomness, fails.

1.3. Principle of restricted invertibility

One important consequence of Theorem 1.2 is that *there always exists a big set of isomorphism of T*. This extends and strengthens a well known result due to Bourgain and Tzafriri, known under the name of the principle of restricted invertibility [1]. We will show how to find a big set of isomorphism; its size can be measured with respect to an arbitrary measure μ on $\mathbb{N}$. For the rest of the paper, we denote the measure of the singletons $\mu(\{i\})$ by μ_i. Summing over i with weights μ_i in (2) and using Theorem 1.2, we obtain

$$\begin{aligned}\int_{\Sigma(T,\varepsilon)} \mu(\sigma)\, d\nu(\sigma) &= \sum_i \mu_i \int_{\Sigma(T,\varepsilon)} \chi_\sigma(i)\, d\nu(\sigma) \\ &= \sum_i \mu_i\, \nu\{\sigma \in \Sigma(T,\varepsilon) \mid i \in \sigma\} \ge c\varepsilon^2 \sum_i \mu(i)\|Te_i\|^2.\end{aligned} \tag{3}$$

Replacing the integral in the left hand side of (4) by the maximum shows that there exists a big set of isomorphism:

Corollary 1.3. *Let T be a norm-one linear operator on a Hilbert space H, and let μ be a measure on $\mathbb{N}$. Then, for every $0 < \varepsilon < 1$, there exists a set of ε-isomorphism σ of T such that*

$$\mu(\sigma) \geq c\varepsilon^2 \sum_i \mu_i \|Te_i\|^2. \tag{4}$$

Earlier, Bourgain and Tzafriri [1] proved a weaker form of Corollary 1.3 with only the *lower* bound in the definition (1) of the set of isomorphism, for a uniform measure μ on an interval, under an additional assumption on the uniform lower bound on $\|Te_i\|$, and for some fixed ε.

Theorem 1.4 (Bourgain-Tzafriri's principle of restricted invertibility). *Let T be a linear operator on an n-dimensional Hilbert space H with an orthonormal basis (e_i). Assume that $\|Te_i\| = 1$ for all i. Then there exits a subset σ of $\{1, \ldots, n\}$ such that $|\sigma| \geq cn/\|T\|^2$ and*

$$\|Tf\| \geq c\|f\|$$

for all $f \in \operatorname{span}(e_i)_{i\in\sigma}$.

This important result has found applications in Banach space theory and harmonic analysis. Corollary 1.3 immediately yields a stronger result, which is dimension-free and which yields an almost isometry:

Corollary 1.5. *Let T be a linear operator on a Hilbert space H with an orthonormal basis (e_i). Assume that $\|Te_i\| = 1$ for all i. Let μ be a measure on $\mathbb{N}$. Then, for every $0 < \varepsilon < 1$, there exits a subset σ of $\mathbb{N}$ such that $\mu(\sigma) \geq c\varepsilon^2/\|T\|^2$ and such that*

$$(1-\varepsilon)\|f\| \leq \|Tf\| \leq (1+\varepsilon)\|f\| \tag{5}$$

for all $f \in \operatorname{span}(e_i)_{i\in\sigma}$.

Szarek [5] proved a weaker form of Corollary 1.3 with only the *upper* bound in the definition (1) of the set of isomorphism, and with some fixed ε.

For the counting measure on $\mathbb{N}$, Corollary 1.3 was proved in [7]. In this case, bound (4) reads as

$$|\sigma| \geq c\varepsilon^2 \|T\|_{\mathrm{HS}}^2, \tag{6}$$

where $\|T\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm of T. (If T is not a Hilbert-Schmidt operator, then an infinite σ exists).

2. Proof of Theorem 1.2

Corollary 1.3 is a consequence of two suppression results due to Szarek [5] and Bourgain-Tzafriri [2]. We will then deduce Theorem 1.2 from Corollary 1.3 by a simple separation argument from [2].

To prove Corollary 1.3, we can assume by a straighforward approximation that our Hilbert space H is finite dimensional. We can thus identify H with the n-dimensional Euclidean space ℓ_2^n, and identify the basis $(e_i)_{i=1}^n$ of H with the canonical basis of ℓ_2^n. Given a subset σ of $\{1, \ldots, n\}$ (or of $\mathbb{N}$), by ℓ_2^σ we denote the subspace of ℓ_2^n (of ℓ_2 respectively) spanned by $(e_i)_{i\in\sigma}$. The orthogonal projection onto ℓ_2^σ is denoted by Q_σ.

With a motivaiton different from ours, Szarek proved in ([5], Lemma 4) the following suppression result for operators in ℓ_2^n.

Theorem 2.1 (Szarek). *Let T be a norm-one linear operator on ℓ_2^n. Let $\lambda_1, \dots, \lambda_n$, $\sum_{i=1}^n \lambda_i = 1$, be positive weights. Then there exists a subset σ of $\{1, \dots, n\}$ such that*

$$\sum_{i \in \sigma} \lambda_i \|Te_i\|^{-2} \geq c \tag{7}$$

and such that the inequality

$$\|\sum_{i \in \sigma} a_i T e_i\|^2 \leq C \sum_{i \in \sigma} \|a_i T e_i\|^2$$

holds for every choice of scalars $(a_i)_{i \in \sigma}$.

Remark 2.2. Inequality (7) for a probability measure λ on $\{1, \dots, n\}$ is equivalent to the inequality

$$\mu(\sigma) \geq c \sum_i \mu_i \|Te_i\|^2 \tag{8}$$

for a positive measure μ on $\{1, \dots, n\}$.

Indeed, (7) implies (8) with

$$\lambda_i = \frac{\mu_i \|Te_i\|^2}{\sum_i \mu_i \|Te_i\|^2}.$$

Conversely, (8) implies (7) with $\mu_i = \lambda_i \|Te_i\|^{-2}$.

Theorem 2.1 and Remark 2.2 yield a weaker version of Corollary 1.3 – with only the upper bound in the definition (1) of the set of isomorphism, and with some fixed ε.

To prove Corollary 1.3 in full strength, we will use the following suppression analog of Theorem 1.2 due to Bourgain and Tzafriri [2].

Theorem 2.3 (Bourgain-Tzafriri). *Let S be a linear operator on ℓ_2 whose matrix relative to the unit vector basis has zero diagonal. For a $\delta > 0$, denote by $\Sigma'(S, \delta)$ the family of all subsets σ of $\mathbb{N}$ such that $\|Q_\sigma S Q_\sigma\| \leq \delta \|S\|$. Then there exists a probability measure ν' on $\Sigma'(S, \delta)$ such that*

$$\nu'\{\sigma \in \Sigma'(S, \delta) \mid i \in \sigma\} \geq c\delta^2 \quad \textit{for all } i. \tag{9}$$

Proof of Corollary 1.3. We define a linear operator T_1 on $H = \ell_2^n$ as

$$T_1 e_i = Te_i / \|Te_i\|, \quad i = 1, \dots, n.$$

Theorem 2.1 and the remark below it yield the existence of a subset σ of $\{1, \dots, n\}$ whose measure satisfies (8) and such that the inequality

$$\|T_1 f\| \leq C \|f\|$$

holds for all $f \in \operatorname{span}(e_i)_{i \in \sigma}$. In other words, the operator

$$T_2 = T_1 Q_\sigma$$

satisfies

$$\|T_2\| \le C. \tag{10}$$

We will apply Theorem 2.3 for the operator S on ℓ_2^σ defined as

$$S = T_2^* T_2 - I \quad \text{and with} \quad \delta = \varepsilon / \|S\|. \tag{11}$$

Indeed, S has zero diagonal:

$$\langle S e_i, e_i \rangle = \|T_2 e_i\|^2 - 1 = \|T_1 e_i\|^2 - 1 = 0 \quad \text{for all } i \in \sigma.$$

Also, S has nicely bounded norm by (10):

$$\|S\| \le \|T_2\|^2 + 1 \le C^2 + 1,$$

which yields a lower bound on δ:

$$\delta \ge \varepsilon / (C^2 + 1). \tag{12}$$

So, Theorem 2.3 yields a family $\Sigma'(S, \delta)$ of subsets of σ and a measure ν' on this family. It follows as before that $\Sigma'(S, \delta)$ must contain a big set, because

$$\begin{aligned} \int_{\Sigma'(S,\delta)} \mu(\sigma') \, d\nu'(\sigma') &= \sum_{i \in \sigma} \mu_i \int_{\Sigma'(S,\delta)} \chi_{\sigma'}(i) \, d\nu'(\sigma') \\ &= \sum_{i \in \sigma} \mu_i \; \nu' \{ \sigma' \in \Sigma'(S, \delta) \mid i \in \sigma' \} \\ &\ge \sum_{i \in \sigma} \mu_i \cdot c\delta^2 \ge c' \varepsilon^2 \mu(\sigma) \end{aligned}$$

where the last inequality follows from (12) with $c' = c(C^2 + 1)^{-2}$. Thus there exists a set $\sigma' \in \Sigma'(S, \delta)$ such that by (8) we have

$$\mu(\sigma') \ge c' \varepsilon^2 \mu(\sigma) \ge c'' \varepsilon^2 \sum_{i=1}^n \mu_i \|T e_i\|^2,$$

so with the measure as required in (4).

It remains to check that σ' is a set of ε-isomorphism of T. Consider an $f \in \operatorname{span}(e_i)_{i \in \sigma'}$, $\|f\| = 1$. By the suppression estimate in Theorem 2.3 and by our choice of S and δ made in (11), we have

$$\begin{aligned} \varepsilon = \delta \|S\| &\ge |\langle Q_{\sigma'} S Q_{\sigma'} f, f \rangle| \\ &= |\langle S f, f \rangle| \quad \text{because } Q_{\sigma'} f = f \\ &= |\|T_2 f\|^2 - \|f\|^2| \quad \text{by the definition of } S \\ &= |\|T_1 f\|^2 - 1| \quad \text{because } Q_{\sigma'} f = Q_\sigma f = f \text{ as } \sigma' \subset \sigma. \end{aligned}$$

It follows by homogeneity that

$$(1 - \varepsilon) \|f\|^2 \le \|T_1 f\|^2 \le (1 + \varepsilon) \|f\|^2 \quad \text{for all } f \in \operatorname{span}(e_i)_{i \in \sigma'}.$$

By the definition of T_1, this means that σ' is a set of ε-isomorphism of T. This completes the proof. $\square$

Proof of Theorem 1.2. We deduce Theorem 1.2 from Corollary 1.3 by a separation argument, which is a minor adaptation of the proof of Corollary 1.4 in [2].

We first note that, by Remark 2.2, an equivalent form of the consequence of Corollary 1.3 is the following. For every probability measure λ on $\mathbb{N}$, there exists a set $\sigma \in \Sigma(T,\varepsilon)$ such that

$$\sum_{i\in\sigma} \lambda_i \|Te_i\|^{-2} \geq c\varepsilon^2. \tag{13}$$

We consider the space of continuous functions $C(\Sigma(T,\varepsilon))$ on the isomorphism structure $\Sigma(T,\varepsilon)$, which is compact in its natural topology (of pointwise convergence of the indicators of the sets $\sigma \in \Sigma(T,\varepsilon)$). For each $i \in \mathbb{N}$, define a function $\pi_i \in C(\Sigma(T,\varepsilon))$ by setting

$$\pi_i(\sigma) = \chi_\sigma(i)\ \|Te_i\|^{-2}, \quad \sigma \in \Sigma(T,\varepsilon).$$

Let $\mathcal{C}$ be the convex hull of the set of functions $\{\pi_i, i \in \mathbb{N}\}$. Every $\pi \in \mathcal{C}$ can be expressed a convex combination $\pi = \sum_i \lambda_i \pi_i$. By Corollary 1.3 in the form (13), there exists a set $\sigma \in \Sigma(T,\varepsilon)$ such that $\pi(\sigma) \geq c\varepsilon^2$. Thus $\|\pi\|_{C(\Sigma(T,\varepsilon))} \geq c\varepsilon^2$. We conclude by the Hahn-Banach theorem that there exists a probability measure $\nu \in C(\Sigma(T,\varepsilon))^*$ such that

$$\nu(\pi) = \int_{\Sigma(T,\varepsilon)} \pi(\sigma)\ d\nu(\sigma) \geq c\varepsilon^2 \quad \text{for all } \pi \in \mathcal{C}.$$

Applying this estimate for $\pi = \pi_i$, we obtain

$$\int_{\Sigma(T,\varepsilon)} \chi_\sigma(i)\ d\nu(\sigma) \geq c\varepsilon^2 \|Te_i\|^2,$$

which is exactly the conclusion of the theorem. □

Remark 2.4. The proof of Theorem 1.2 given above is a combination of previously known tools – two suppression results due to [5] and [2] and a separation argument from [2]. The new point was to realize that the suppression result of Szarek [5], developed with a different purpose in mind, gives a sharp estimate when combined with the results of [2]. To find a set of the isomorphism as in (1), one needs to reduce the norm of the operator with [5] *before* applying restricted invertibility principles from [2].

Acknowledgement. I am grateful to the referee for numerous comments and suggestions.

References

[1] Bourgain, J. and Tzafriri, L. (1987). Invertibility of “large” submatrices and applications to the geometry of Banach spaces and Harmonic Analysis. *Israel J. Math.* **57** 137–224.

[2] Bourgain, J. and Tzafriri, L. (1991). On a problem of Kadison and Singer, *J. Reine Angew. Math.* **420** 1–43.

[3] Kashin, B. and Tzafriri, L. Some remarks on the restrictions of operators to coordinate subspaces, unpublished.

[4] Milman, V. and Schechtman, G. (1986). Asymptotic theory of finite dimensional normed spaces. *Lecture Notes in Math.* **1200**. Springer.

[5] Szarek, S. (1991). Computing summing norms and type constants on few vectors. *Studia Mathematica* **98** 147–156.

[6] Tomczak-Jaegermann, N. (1989). *Banach–Mazur Distances and Finite Dimensional Operator Ideals*. Pitman.

[7] Vershynin, R. (2001). John's decompositions: selecting a large part. *Israel Journal of Mathematics* **122** 253–277.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 155–172

DOI: 10.1214/074921706000000824

Revisiting two strong approximation results of Dudley and Philipp

This paper is dedicated to the memory of Walter Philipp.

Philippe Berthet[1] **and David M. Mason**[2,*]

Université Rennes 1 and University of Delaware

Abstract: We demonstrate the strength of a coupling derived from a Gaussian approximation of Zaitsev (1987a) by revisiting two strong approximation results for the empirical process of Dudley and Philipp (1983), and using the coupling to derive extended and refined versions of them.

1. Introduction

Einmahl and Mason [17] pointed out in their Fact 2.2 that the Strassen–Dudley theorem (see Theorem 11.6.2 in [11]) in combination with a special case of Theorem 1.1 and Example 1.2 of Zaitsev [42] yields the following coupling. Here $|\cdot|_N$, $N \geq 1$, denotes the usual Euclidean norm on $\mathbb{R}^N$.

Coupling inequality. *Let $Y_1, \dots, Y_n$ be independent mean zero random vectors in $\mathbb{R}^N$, $N \geq 1$, such that for some $B > 0$,*

$$|Y_i|_N \leq B, \ i = 1, \dots, n.$$

If $(\Omega, \mathcal{T}, \mathbb{P})$ is rich enough then for each $\delta > 0$, one can define independent normally distributed mean zero random vectors $Z_1, \dots, Z_n$ with Z_i and Y_i having the same variance/covariance matrix for $i = 1, \dots, n$, such that for universal constants $C_1 > 0$ and $C_2 > 0$,

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n}(Y_i - Z_i)\right|_N > \delta\right\} \leq C_1 N^2 \exp\left(-\frac{C_2\delta}{N^2 B}\right). \tag{1.1}$$

(Actually Einmahl and Mason did not specify the N^2 in (1.1) and they applied a less precise result in [43], however their argument is equally valid when based upon [42].) Often in applications, N is allowed to increase with n. This result and its variations, when combined with inequalities from empirical and Gaussian processes and from probability on Banach spaces, has recently been shown to be an extremely powerful tool to establish a Gaussian approximation to the uniform empirical process on the $d-$dimensional cube (Rio [34]), strong approximations for the local empirical process (Einmahl and Mason [17]), extreme value results for the

[1]IRMAR, Université Rennes 1, Campus de Beaulieu, 35042 Rennes, France, e-mail: philippe.berthet@univ-rennes1.fr

[2]Statistics Program, University of Delaware, 206 Townsend Hall, Newark, DE 19716, USA, e-mail: davidm@udel.edu

*Research partially supported by an NSF Grant.

AMS 2000 subject classifications: primary 62E17, 62E20; secondary 60F15.

Keywords and phrases: Gaussian approximation, coupling, strong approximation.

Hopfield model (Bovier and Mason [3] and Gentz and Löwe [19]), laws of the iterated logarithm in Banach spaces (Einmahl and Kuelbs [15]), moderate deviations for Banach space valued sums (Einmahl and Kuelbs [16]), and a functional large deviation result for the local empirical process (Mason [26]). In this paper we shall further demonstrate the strength of (1.1) by revisiting two strong approximation results for the empirical process of Dudley and Philipp [14], and use (1.1) to derive extended and refined versions of them.

Dudley and Philipp [14] was a path breaking paper, which introduced a very effective technique for obtaining Gaussian approximations to sums of i.i.d. Banach space valued random variables. The strong approximation results of theirs, which we shall revisit, were derived from a much more general result in their paper. Key to this result was their Lemma 2.12, which is a special case of an extension by Dehling [8] of a Gaussian approximation in the Prokhorov distance to sums of i.i.d. multivariate random vectors due to Yurinskii [41]. In essence, we shall be substituting the application of their Lemma 2.12 by the above coupling inequality (1.1) based upon Zaitsev [42]. We shall also update and streamline the methodology by employing inequalities that were not available to Dudley and Philipp, when they wrote their paper.

1.1. The Gaussian approximation and strong approximation problems

Let us begin by describing the Gaussian approximation problem for the empirical process. For a fixed integer $n \geq 1$ let $X, X_1, \dots, X_n$ be independent and identically distributed random variables defined on the same probability space $(\Omega, \mathcal{T}, \mathbb{P})$ and taking values in a measurable space $(\mathcal{X}, \mathcal{A})$. Denote by $\mathbb{E}$ the expectation with respect to $\mathbb{P}$ of real valued random variables defined on $(\Omega, \mathcal{T})$ and write $P = \mathbb{P}^X$. Let $\mathcal{M}$ be the set of all measurable real valued functions on $(\mathcal{X}, \mathcal{A})$. In this paper we consider the following two processes indexed by a sufficiently small class $\mathcal{F} \subset \mathcal{M}$. First, define the P-empirical process indexed by $\mathcal{F}$ to be

$$\alpha_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i) - \mathbb{E}f(X)\}, \ f \in \mathcal{F}. \tag{1.2}$$

Second, define the P-Brownian bridge $\mathbb{G}$ indexed by $\mathcal{F}$ to be the mean zero Gaussian process with the same covariance function as α_n,

$$\langle f, h \rangle = cov(\mathbb{G}(f), \mathbb{G}(h)) = \mathbb{E}\left(f(X)h(X)\right) - \mathbb{E}\left(f(X)\right)\mathbb{E}(h(X)), \ f, g \in \mathcal{F}. \tag{1.3}$$

Under entropy conditions on $\mathcal{F}$, the Gaussian process $\mathbb{G}$ has a version which is almost surely continuous with respect to the intrinsic semi-metric

$$d_P(f, h) = \sqrt{\mathbb{E}\left(f(X) - h(X)\right)^2}, \ f, g \in \mathcal{F}, \tag{1.4}$$

that is, we include d_P-continuity in the definition of $\mathbb{G}$.

Our goal is to show that a version of $X_1, \dots, X_n$ and $\mathbb{G}$ can be constructed on the same underlying probability space $(\Omega, \mathcal{T}, \mathbb{P})$ in such a way that

$$\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\alpha_n(f) - \mathbb{G}(f)| \tag{1.5}$$

is very small with high probability, under useful assumptions on $\mathcal{F}$ and P. This is what we call the *Gaussian approximation problem.* We shall use our Gaussian approximation results to define on the same probability $(\Omega, \mathcal{T}, \mathbb{P})$ a sequence $X_1, X_2, \ldots$, i.i.d. X and a sequence $\mathbb{G}_1, \mathbb{G}_2, \ldots$, i.i.d. $\mathbb{G}$ so that with high probability,

$$n^{-1/2} \max_{1 \le m \le n} \left\| \sqrt{m}\alpha_m - \sum_{i=1}^{m} \mathbb{G}_i \right\|_{\mathcal{F}} \tag{1.6}$$

is small. This is what we call the *strong approximation problem.*

1.2. Basic assumptions

We shall assume that $\mathcal{F}$ satisfies the following boundedness condition (F.i) and measurability condition (F.ii).

(F.i) *For some $M > 0$, for all $f \in \mathcal{F}$, $\|f\|_{\mathcal{X}} = \sup_{x \in \mathcal{X}} |f(x)| \le M/2$.*

(F.ii) *The class $\mathcal{F}$ is point-wise measurable, i.e. there exists a countable subclass $\mathcal{F}_\infty$ of $\mathcal{F}$ such that we can find for any function $f \in \mathcal{F}$ a sequence of functions $\{f_m\}$ in $\mathcal{F}_\infty$ for which $\lim_{m \to \infty} f_m(x) = f(x)$ for all $x \in \mathcal{X}$.*

Assumption (F.i) justifies the finiteness of all the integrals that follow as well as the application of the key inequalities. The requirement (F.ii) is imposed to avoid using outer probability measures in our statements – see Example 2.3.4 in [38].

We intend to compute probability bounds for (1.5) holding for any n and some fixed M in (F.i) with ensuing constants independent of n.

2. Entropy approach based on Zaitsev [42]

We shall require that one of the following two L_2-metric entropy conditions (VC) and (BR) holds on the class $\mathcal{F}$. These conditions are commonly used in the context of weak invariance principles and many examples are available – see e.g. van der Vaart and Wellner [38] and Dudley [12]. In this section we shall state our main results. We shall prove them in Section 5.

2.1. L_2-covering numbers

First we consider polynomially scattered classes $\mathcal{F}$. Let F be an envelope function for the class $\mathcal{F}$, that is, F a measurable function such that $|f(x)| \le F(x)$ for all $x \in \mathcal{X}$ and $f \in \mathcal{F}$. Given a probability measure Q on $(\mathcal{X}, \mathcal{A})$ endow $\mathcal{M}$ with the semi-metric d_Q, where $d_Q^2(f,h) = \int_{\mathcal{X}} (f-h)^2 dQ$. Further, for any $f \in \mathcal{M}$ set $Q(f^2) = d_Q^2(f,0) = \int_{\mathcal{X}} f^2 dQ$. For any $\varepsilon > 0$ and probability measure Q denote by $N(\varepsilon, \mathcal{F}, d_Q)$ the minimal number of balls $\{f \in \mathcal{M} : d_Q(f,h) < \varepsilon\}$ of d_Q-radius ε and center $h \in \mathcal{M}$ needed to cover $\mathcal{F}$. The uniform L_2-covering number is defined to be

$$N_F(\varepsilon, \mathcal{F}) = \sup_Q N\left(\varepsilon\sqrt{Q(F^2)}, \mathcal{F}, d_Q\right), \tag{2.1}$$

where the supremum is taken over all probability measures Q on $(\mathcal{X}, \mathcal{A})$ for which $0 < Q(F^2) < \infty$. A class of functions $\mathcal{F}$ satisfying the following uniform entropy condition will be called a VC class.

(VC) *Assume that for some* $c_0 > 0$, $\nu_0 > 0$, *and envelope function* F,

$$N_F(\varepsilon, \mathcal{F}) \le c_0 \varepsilon^{-\nu_0}, \ 0 < \varepsilon < 1. \tag{2.2}$$

The name "VC class" is given to this condition in recognition to Vapnik and Červonenkis [39] who introduced a condition on classes of sets, which implies (VC). In the sequel we shall assume that $F := M/2$ as in (F.i).

Proposition 1. *Under* (F.i), (F.ii) *and* (VC) *with* $F := M/2$ *for each* $\lambda > 1$ *there exists a* $\rho(\lambda) > 0$ *such that for each integer* $n \ge 1$ *one can construct on the same probability space random vectors* $X_1, \ldots, X_n$ *i.i.d.* X *and a version of* $\mathbb{G}$ *such that*

$$\mathbb{P}\left\{\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} > \rho(\lambda)\, n^{-\tau_1} (\log n)^{\tau_2}\right\} \le n^{-\lambda}, \tag{2.3}$$

where $\tau_1 = 1/(2 + 5\nu_0)$ *and* $\tau_2 = (4 + 5\nu_0)/(4 + 10\nu_0)$.

Proposition 1 leads to the following strong approximation result. It is an indexed by functions generalization of an indexed by sets result given in Theorem 7.4 of Dudley and Philipp [14].

Theorem 1. *Under the assumptions and notation of Proposition 1 for all* $1/(2\tau_1) < \alpha < 1/\tau_1$ *and* $\gamma > 0$ *there exist a* $\rho(\alpha, \gamma) > 0$, *a sequence of i.i.d.* $X_1, X_2, \ldots,$ *and a sequence of independent copies* $\mathbb{G}_1, \mathbb{G}_2, \ldots,$ *of* $\mathbb{G}$ *sitting on the same probability space such that*

$$\mathbb{P}\left\{\max_{1 \le m \le n} \left\| \sqrt{m}\alpha_m - \sum_{i=1}^{m} \mathbb{G}_i \right\|_{\mathcal{F}} > C\rho(\alpha, \gamma)\, n^{1/2 - \tau(\alpha)} (\log n)^{\tau_2} \right\} \le n^{-\gamma} \tag{2.4}$$

and

$$\max_{1 \le m \le n} \left\| \sqrt{m}\alpha_m - \sum_{i=1}^{m} \mathbb{G}_i \right\|_{\mathcal{F}} = O\left(n^{1/2 - \tau(\alpha)} (\log n)^{\tau_2}\right), \ a.s., \tag{2.5}$$

where $\tau(\alpha) = (\alpha\tau_1 - 1/2)/(1 + \alpha) > 0$.

2.2. Bracketing numbers

A second way to measure the size of the class $\mathcal{F}$ is to use $L_2(P)$-brackets instead of $L_2(Q)$-balls. Let $l \in \mathcal{M}$ and $u \in \mathcal{M}$ be such that $l \le u$ and $d_P(l, u) < \varepsilon$. The pair of functions l, u form an ε-bracket $[l, u]$ consisting of all the functions $f \in \mathcal{F}$ such that $l \le f \le u$. Let $N_{[\,]}(\varepsilon, \mathcal{F}, d_P)$ be the minimum number of ε-brackets needed to cover $\mathcal{F}$. Notice that trivially we have $N(\varepsilon, \mathcal{F}, d_P) \le N_{[\,]}(\varepsilon/2, \mathcal{F}, d_P)$.

(BR) *Assume that for some* $b_0 > 0$ *and* $0 < r_0 < 1$,

$$\log N_{[\,]}(\varepsilon, \mathcal{F}, d_P) \le b_0^2 \varepsilon^{-2r_0}, \ 0 < \varepsilon < 1. \tag{2.6}$$

We derive the following rate of Gaussian approximation assuming an exponentially scattered index class $\mathcal{F}$, meaning that (2.6) holds. Note that we get a slower rate in Proposition 2 than that given Proposition 1.

Proposition 2. *Under* (F.i), (F.ii) *and* (BR) *for each* $\lambda > 1$ *there exists a* $\rho(\lambda) > 0$ *such that for each integer* $n \geq 1$ *one can construct on the same probability space random vectors* $X_1, \ldots, X_n$ *i.i.d.* X *and a version of* $\mathbb{G}$ *such that*

$$\mathbb{P}\left\{\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} > \rho(\lambda)(\log n)^{-\kappa}\right\} \leq n^{-\lambda}, \tag{2.7}$$

where $\kappa = (1 - r_0)/2r_0$.

Proposition 2 leads to the following indexed by functions generalization of an indexed by sets result given in Theorem 7.1 of Dudley and Philipp [14].

Theorem 2. *Under the assumptions and notation of Proposition 2, with* $\kappa < 1/2$ $(1/2 < r_0 < 1)$*, for every* $H > 0$ *there exist* $\rho(\tau, H) > 0$ *and a sequence of i.i.d.* $X_1, X_2, \ldots,$ *and a sequence of independent copies* $\mathbb{G}_1, \mathbb{G}_2, \ldots,$ *of* $\mathbb{G}$ *sitting on the same probability space such that*

$$\mathbb{P}\left\{\max_{1\leq m\leq n}\left\|\sqrt{m}\alpha_m - \sum_{i=1}^{m}\mathbb{G}_i\right\|_{\mathcal{F}} > \sqrt{n}\rho(\tau, H)(\log n)^{-\tau}\right\} \leq (\log n)^{-H} \tag{2.8}$$

and

$$\max_{1\leq m\leq n}\left\|\sqrt{m}\alpha_m - \sum_{i=1}^{m}\mathbb{G}_i\right\|_{\mathcal{F}} = O\left(\sqrt{n}(\log n)^{-\tau}\right), \text{ a.s.,} \tag{2.9}$$

where $\tau = \kappa(1/2 - \kappa)/(1 - \kappa)$.

3. Comments on the approach based on KMT

Given $\mathcal{F}$, the rates obtained in Proposition 1 and Theorem 1 are universal in P. If one specializes to particular P, the rates in Propositions 1 and 2 and Theorem 1 and 2 are far from being optimal. In such situations one can get better and even unimprovable rates by replacing the use of Zaitsev [42] by the Komlós, Major and Tusnády [KMT] [22] Brownian bridge approximation to the uniform empirical process or one based on the same dyadic scheme. (More details about this approximation are provided in [4, 13, 25, 27, 28].) This is especially the case when the underlying probability measure P is smooth. To see how this works in the empirical process indexed by functions setup refer to Koltchinskii [21] and Rio [33] and in the indexed by smooth sets situation turn to Révesz [32] and Massart [29]. One can also use the KMT–type bivariate Brownian bridge approximation to the bivariate uniform empirical process as a basis for further approximation. For a brief outline of this approximation consult Tusnády [36] and for detailed presentations refer to Castelle [5] and Castelle and Laurent-Bonvalot [6].

4. Tools needed in proofs

For convenience we shall collect here the basic tools we shall need in our proofs.

4.1. Inequalities for empirical processes

On a rich enough probability space $(\Omega, \mathcal{T}, \mathbb{P})$, let $X, X_1, X_2, \ldots, X_n$ be i.i.d. random variables with law $P = \mathbb{P}^X$ and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ be i.i.d. Rademacher random variables independent of $X_1,\dots,X_n$. By a Rademacher random variable ϵ_1, we mean that $\mathbb{P}(\epsilon_1 = 1) = \mathbb{P}(\epsilon_1 = -1) = 1/2$. Consider a point-wise measurable class $\mathcal{G}$ of bounded measurable real valued functions on $(\mathcal{X},\mathcal{A})$.

The following exponential inequality is due to Talagrand [35].

Talagrand's inequality. *If $\mathcal{G}$ satisfies* (F.i) *and* (F.ii) *then for all $n \geq 1$ and $t > 0$ we have, for suitable finite constants $A > 0$ and $A_1 > 0$,*

$$\begin{aligned}\mathbb{P}\left\{ ||\alpha_n||_{\mathcal{G}} > A\left(\mathbb{E}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i g(X_i)\right\|_{\mathcal{G}}\right) + t\right)\right\} \\ \leq 2\exp\left(-\frac{A_1 t^2}{\sigma_{\mathcal{G}}^2}\right) + 2\exp\left(-\frac{A_1 t\sqrt{n}}{M}\right),\end{aligned} \tag{4.1}$$

where $\sigma_{\mathcal{G}}^2 := \sup_{g\in\mathcal{G}} Var(g(X))$.

Moreover the constants A and A_1 are independent of $\mathcal{G}$ and M. Next we state two upper bounds for the above expectation of the supremum of the symmetrized empirical process.

We shall require two moment bounds. The first is due to Einmahl and Mason [18] – for a similar bound refer to Giné and Guillou [20].

Moment inequality for (VC). *Let $\mathcal{G}$ satisfy* (F.i) *and* (F.ii) *with envelope function G and be such that for some positive constants β, v, $c > 1$ and $\sigma \leq 1/(8c)$ the following four conditions hold,*

$$\mathbb{E}(G^2(X)) \leq \beta^2;\ N_G(\varepsilon,\mathcal{G}) \leq c\varepsilon^{-v},\ 0 < \varepsilon < 1;\ \sup_{g\in\mathcal{G}} \mathbb{E}(g^2(X)) \leq \sigma^2;$$

and

$$\sup_{g\in\mathcal{G}} \|g\|_{\mathcal{X}} \leq \frac{\sqrt{n\sigma^2/\log(\beta\vee 1/\sigma)}}{2\sqrt{v+1}}.$$

Then we have for a universal constant A_2 not depending on β,

$$\mathbb{E}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i g(X_i)\right\|_{\mathcal{G}}\right) \leq A_2\sqrt{v\sigma^2\log(\beta\vee 1/\sigma)}. \tag{4.2}$$

Next we state a moment inequality under (BR). For any $0 < \sigma < 1$, set

$$J(\sigma,\mathcal{G}) = \int_{[0,\sigma]} \sqrt{\log N_{[\,]}(s,\mathcal{G},d_P)}\, ds \tag{4.3}$$

and

$$a(\sigma,\mathcal{G}) = \frac{\sigma}{\sqrt{\log N_{[\,]}(\sigma,\mathcal{G},d_P)}}. \tag{4.4}$$

The second moment bound follows from Lemma 19.34 in [37] and a standard symmetrization inequality, and is reformulated by using (4.3).

Moment inequality for (BR). *Let $\mathcal{G}$ satisfy* (F.i) *and* (F.ii) *with envelope G and be such that* $\sup_{g\in\mathcal{G}}\mathbb{E}\left(g^2(X)\right) < \sigma^2 < 1$. *We have, for a universal constant A_3,*

$$\mathbb{E}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i g(X_i)\right\|_{\mathcal{G}}\right) \leq A_3\left(J(\sigma,\mathcal{G}) + \sqrt{n}\,\mathbb{P}\left\{G(X) > \sqrt{n}\,a(\sigma,\mathcal{G})\right\}\right). \tag{4.5}$$

4.2. Inequalities for Gaussian processes

Let $\mathbb{Z}$ be a separable mean zero Gaussian process on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ indexed by a set T. Define the intrinsic semi–metric ρ on T by

$$\rho(s,t) = \sqrt{\mathbb{E}(\mathbb{Z}_t - \mathbb{Z}_s)^2}. \tag{4.6}$$

For each $\varepsilon > 0$ let $N(\varepsilon, T, \rho)$ denote the minimal number of ρ-balls of radius ε needed to cover T. Write $\|\mathbb{Z}\|_T = \sup_{t\in T}|\mathbb{Z}_t|$ and $\sigma_T^2(\mathbb{Z}) = \sup_{t\in T}\mathbb{E}\left(\mathbb{Z}_t^2\right)$. The following large deviation probability estimate for $\|\mathbb{Z}\|_T$ is due to Borell [2]. (Also see Proposition A.2.1 in [38].)

Borell's inequality. *For all $t > 0$,*

$$\mathbb{P}\left\{\left|\|\mathbb{Z}\|_T - \mathbb{E}\left(\|\mathbb{Z}\|_T\right)\right| > t\right\} \le 2\exp\left(-\frac{t^2}{2\sigma_T^2(\mathbb{Z})}\right). \tag{4.7}$$

According to Dudley [9], the entropy condition

$$\int_{[0,1]} \sqrt{\log N(\varepsilon, T, \rho)}\, d\varepsilon < \infty \tag{4.8}$$

ensures the existence of a separable, bounded, d_P-uniformly continuous modification of $\mathbb{Z}$. Moreover the above Dudley integral (4.8) controls the modulus of continuity of $\mathbb{Z}$ (see Dudley [10]) as well as its expectation (see Marcus and Pisier [24], p. 25, Ledoux and Talagrand [23], p. 300, de la Peña and Giné [7], Cor. 5.1.6, and Dudley [12]). The following inequality is part of Corollary 2.2.8 in van der Vaart and Wellner [38].

Gaussian moment inequality. *For some universal constant $A_4 > 0$ and all $\sigma > 0$ we have*

$$\mathbb{E}\left(\sup_{\rho(s,t)<\sigma} |\mathbb{Z}_t - \mathbb{Z}_s|\right) \le A_4 \int_{[0,\sigma]} \sqrt{\log N(\varepsilon, T, \rho)}\, d\varepsilon. \tag{4.9}$$

We shall be applying these inequalities to the Gaussian process $\mathbb{Z} = \mathbb{G}$ defined in introduction, so that $T = \mathcal{F}$ and $\rho = d_P$.

4.3. A maximal inequality

The following version of a maximal inequality due to Montgomery–Smith [30] (see also Theorem 1.1.5 in [7]) will come in handy.

A maximal inequality. *Let $X_1, \dots, X_n$, $n \ge 1$, be i.i.d. random variables taking values in a separable Banach space. Then for all $t > 0$,*

$$\mathbb{P}\left\{\max_{1\le m\le n}\left\|\sum_{i=1}^m X_i\right\| > t\right\} \le 9\mathbb{P}\left\{\left\|\sum_{i=1}^n X_i\right\| > \frac{t}{30}\right\}. \tag{4.10}$$

5. Proofs of main results

5.1. Description of construction of $(\alpha_n, \mathbb{G})$

Under (F.i), (F.ii) and either (VC) or (BR) for any $\varepsilon > 0$ we can choose a grid

$$\mathcal{H}(\varepsilon) = \{h_k : 1 \le k \le N(\varepsilon)\}$$

of measurable functions on $(\mathcal{X}, \mathcal{A})$ such that each $f \in \mathcal{F}$ is in a ball $\{f \in \mathcal{M} : d_P(h_k, f) < \varepsilon\}$ around some h_k, $1 \le k \le N(\varepsilon)$. The choice

$$N(\varepsilon) \le N(\varepsilon/2, \mathcal{F}, d_P) \tag{5.1}$$

permits us to select $h_k \in \mathcal{F}$. Set

$$\mathcal{F}(\varepsilon) = \left\{(f, f') \in \mathcal{F}^2 : d_P(f, f') < \varepsilon\right\}.$$

Fix $n \ge 1$. Let X, $X_1, \ldots, X_n$ be independent with common law $P = \mathbb{P}^X$ and $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher random variables mutually independent of $X_1, \ldots, X_n$. Write for $\varepsilon > 0$,

$$\mu_n(\varepsilon) = \mathbb{E}\left\{\sup_{(f,f')\in\mathcal{F}(\varepsilon)} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \epsilon_i (f - f')(X_i)\right|\right\}$$

and

$$\mu(\varepsilon) = \mathbb{E}\left\{\sup_{(f,f')\in\mathcal{F}(\varepsilon)} |\mathbb{G}(f) - \mathbb{G}(f')|\right\}.$$

Given $\varepsilon > 0$ and $n \ge 1$, our aim is to construct a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ on which sit $X_1, \ldots, X_n$ and a version of the Gaussian process $\mathbb{G}$ indexed by $\mathcal{F}$ such that for $\mathcal{H}(\varepsilon)$ and $\mathcal{F}(\varepsilon)$ defined as above and for all $A > 0$, $\delta > 0$ and $t > 0$,

$$\begin{aligned}
&\mathbb{P}\{\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} > A\mu_n(\varepsilon) + \mu(\varepsilon) + \delta + (A+1)t\} \\
&\quad \le \mathbb{P}\left\{\max_{h\in\mathcal{H}(\varepsilon)} |\alpha_n(h) - \mathbb{G}(h)| > \delta\right\} \\
&\qquad + \mathbb{P}\left\{\sup_{(f,f')\in\mathcal{F}(\varepsilon)} |\alpha_n(f) - \alpha_n(f')| > A\mu_n(\varepsilon) + At\right\} \\
&\qquad + \mathbb{P}\left\{\sup_{(f,f')\in\mathcal{F}(\varepsilon)} |\mathbb{G}(f) - \mathbb{G}(f')| > t + \mu(\varepsilon)\right\} \\
&\quad =: P_n(\delta) + Q_n(t, \varepsilon) + Q(t, \varepsilon),
\end{aligned} \tag{5.2}$$

with all these probabilities simultaneously small for suitably chosen $A > 0$, $\delta > 0$ and $t > 0$. Consider the n i.i.d. mean zero random vectors in $\mathbb{R}^{N(\varepsilon)}$,

$$Y_i := \frac{1}{\sqrt{n}}\left(h_1(X_i) - \mathbb{E}(h_1(X)), \ldots, h_{N(\varepsilon)}(X_i) - \mathbb{E}(h_{N(\varepsilon)}(X))\right), \ 1 \le i \le n.$$

First note that by $h_k \in \mathcal{F}$ and (F.i), we have

$$|Y_i|_{N(\varepsilon)} \le M\sqrt{\frac{N(\varepsilon)}{n}}, \ 1 \le i \le n.$$

Therefore by the coupling inequality (1.1) we can define $Y_1, \ldots, Y_n$ i.i.d.

$$Y := \left(Y^1, \ldots, Y^{N(\varepsilon)}\right)$$

and $Z_1, \ldots, Z_n$ i.i.d.

$$Z := \left(Z^1, \ldots, Z^{N(\varepsilon)}\right)$$

mean zero Gaussian vectors on the same probability space such that

$$\text{(5.3)} \quad P_n(\delta) \le \mathbb{P}\left\{\left|\sum_{i=1}^{n}(Y_i - Z_i)\right|_{N(\varepsilon)} > \delta\right\} \le C_1 N(\varepsilon)^2 \exp\left(-\frac{C_2\sqrt{n}\,\delta}{(N(\varepsilon))^{5/2} M}\right),$$

where $cov(Z^l, Z^k) = cov(Y^l, Y^k) = \langle h_l, h_k \rangle$. Moreover by Lemma A1 of Berkes and Philipp [1] (also see Vorob'ev [40]) this space can be extended to include a P-Brownian bridge $\mathbb{G}$ indexed by $\mathcal{F}$ such that

$$\mathbb{G}(h_k) = n^{-1/2} \sum_{i=1}^{n} Z_i^k.$$

The $P_n(\delta)$ in (5.2) is defined through this $\mathbb{G}$. Notice that the probability space on which $Y_1, \ldots, Y_n$, $Z_1, \ldots, Z_n$ and $\mathbb{G}$ sit depends on $n \ge 1$ and the choice of $\varepsilon > 0$ and $\delta > 0$.

Observe that the class

$$\mathcal{G}(\varepsilon) = \{f - f' : (f, f') \in \mathcal{F}(\varepsilon)\}$$

satisfies (F.i) with $M/2$ replaced by M, (F.ii) and

$$\sigma^2_{\mathcal{G}(\varepsilon)} = \sup_{(f,f')\in\mathcal{F}(\varepsilon)} Var(f(X) - f'(X)) \le \sup_{(f,f')\in\mathcal{F}(\varepsilon)} d_P^2(f, f') \le \varepsilon^2.$$

Thus with $A > 0$ as in (4.1) we get by applying Talagrand's inequality,

$$\text{(5.4)} \quad \begin{aligned} Q_n(t, \varepsilon) &= \mathbb{P}\left\{||\alpha_n||_{\mathcal{G}(\varepsilon)} > A(\mu_n(\varepsilon) + t)\right\} \\ &\le 2\exp\left(-\frac{A_1 t^2}{\varepsilon^2}\right) + 2\exp\left(-\frac{A_1\sqrt{n}\,t}{M}\right). \end{aligned}$$

Next, consider the separable centered Gaussian process $\mathbb{Z}_{(f,f')} = \mathbb{G}(f) - \mathbb{G}(f')$ indexed by $T = \mathcal{F}(\varepsilon)$. We have

$$\begin{aligned} \sigma_T^2(\mathbb{Z}) &= \sup_{(f,f')\in\mathcal{F}(\varepsilon)} \mathbb{E}\left((\mathbb{G}(f) - \mathbb{G}(f'))^2\right) = \sup_{(f,f')\in\mathcal{F}(\varepsilon)} Var(f(X) - f'(X)) \\ &\le \sup_{(f,f')\in\mathcal{F}(\varepsilon)} d_P^2(f, f') \le \varepsilon^2. \end{aligned}$$

Borell's inequality (4.7) now gives

$$\text{(5.5)} \quad Q(t, \varepsilon) = \mathbb{P}\left\{\sup_{(f,f')\in\mathcal{F}(\varepsilon)} |\mathbb{G}(f) - \mathbb{G}(f')| > t + \mu(\varepsilon)\right\} \le 2\exp\left(-\frac{t^2}{2\varepsilon^2}\right).$$

Putting (5.3), (5.4) and (5.5) together we obtain, for some positive constants A, A_1 and A_5 with $A_5 \leq 1/2$,

$$\begin{aligned}\mathbb{P}\{ \|\alpha_n - \mathbb{G}\|_{\mathcal{F}} &> A\mu_n(\varepsilon) + \mu(\varepsilon) + \delta + (A+1)\,t\} \\ &\leq C_1 N(\varepsilon)^2 \exp\left(-\frac{C_2\sqrt{n}\,\delta}{(N(\varepsilon))^{5/2} M}\right) \\ &\quad + 2\exp\left(-\frac{A_1\sqrt{n}\,t}{M}\right) + 4\exp\left(-\frac{A_5 t^2}{\varepsilon^2}\right). \end{aligned} \tag{5.6}$$

Proof of Proposition 1. Let us assume that (VC) holds with $F := M/2$, so that for some $c_0 > 0$ and $\nu_0 > 0$, with $c_1 = c_0(2\sqrt{PF^2})^{\nu_0} = c_0 M^{\nu_0}$,

$$N(\varepsilon) \leq N(\varepsilon/2, \mathcal{F}, d_P) \leq c_1 \varepsilon^{-\nu_0},\ 0 < \varepsilon < 1.$$

Notice that both

$$N(\varepsilon, \mathcal{G}(\varepsilon), d_P) \leq (N(\varepsilon/2, \mathcal{F}, d_P))^2 \leq c_1^2 \varepsilon^{-2\nu_0}$$

and

$$N(\varepsilon, \mathcal{F}(\varepsilon), d_P) \leq (N(\varepsilon/2, \mathcal{F}, d_P))^2 \leq c_1^2 \varepsilon^{-2\nu_0}.$$

Therefore we can apply the moment bound assuming (VC) given in (4.2) taken with $\mathcal{G} = \mathcal{G}(\varepsilon)$, $G := M$, $v = 2\nu_0$ and $\beta = M$, to get for any $0 < \varepsilon < 1/e$ and $n \geq 1$ so that

$$\frac{\sqrt{n}\varepsilon}{2\sqrt{1+2\nu_0}\sqrt{\log(M \vee 1/\varepsilon)}} > M \tag{5.7}$$

the bound

$$\mu_n(\varepsilon) \leq A_2 \varepsilon \sqrt{2\nu_0 \log(M \vee 1/\varepsilon)}.$$

Whereas, by the Gaussian moment bound (4.9), we have for all $0 < \varepsilon < 1/e$,

$$\mu(\varepsilon) \leq A_4 \sqrt{2\nu_0} \int_{[0,\varepsilon]} \sqrt{\log(1/x)}dx.$$

Hence, for some $D > 0$ it holds for all $0 < \varepsilon < 1/e$ and $n \geq 1$ so that (5.7) holds,

$$A\mu_n(\varepsilon) + \mu(\varepsilon) \leq D\varepsilon\sqrt{\log(1/\varepsilon)}. \tag{5.8}$$

Therefore, in view of (5.8) and (5.6) it is natural to define for suitably large positive γ_1 and γ_2,

$$\delta = \gamma_1 \varepsilon \sqrt{\log(1/\varepsilon)} \text{ and } t = \gamma_2 \varepsilon \sqrt{\log(1/\varepsilon)}.$$

We now have for all $0 < \varepsilon < 1/e$ and $n \geq 1$ so that (5.7) is satisfied on a suitable probability space depending on $n \geq 1$, ε and δ so that (5.6) holds,

$$\begin{aligned}\mathbb{P}&\left\{\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} > (D + \gamma_1 + (1+A)\gamma_2)\,\varepsilon\sqrt{\log(1/\varepsilon)}\right\} \\ &\leq \frac{C_1 c_1^2}{\varepsilon^{2\nu_0}} \exp\left(-\frac{\gamma_1 C_2 \sqrt{n}}{c_1^{5/2} M}\, \varepsilon^{1+5\nu_0/2}\sqrt{\log(1/\varepsilon)}\right) \\ &\quad + 2\exp\left(-\frac{A_1\gamma_2\sqrt{n}}{M}\, \varepsilon\sqrt{\log(1/\varepsilon)}\right) + 4\exp\left(-A_5\gamma_2^2 \log(1/\varepsilon)\right). \end{aligned}$$

By taking $\varepsilon = ((\log n)/n)^{1/(2+5\nu_0)}$, which satisfies (5.7) for all large enough n, we readily obtain from these last bounds that for every $\lambda > 1$ there exist $D > 0$, $\gamma_1 > 0$ and $\gamma_2 > 0$ such that for all $n \geq 1$, α_n and $\mathbb{G}$ can be defined on the same probability space so that

$$\mathbb{P}\left\{\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} > (D + \gamma_1 + (1+A)\,\gamma_2)\left(\frac{\log n}{n}\right)^{1/(2+5\nu_0)}\sqrt{\frac{\log n}{2+5\nu_0}}\right\} \leq n^{-\lambda}.$$

It is clear now that there exists a $\rho\,(\lambda) > 0$ such that (2.3) holds. This completes the proof of Proposition 1. □

Proof of Proposition 2. Under (BR) as defined in (2.6) we have, for some $0 < r_0 < 1$ and $b_0 > 0$,

$$N\,(\varepsilon) \leq N(\varepsilon/2, \mathcal{F}, d_P) \leq N_{[\,]}(\varepsilon/2, \mathcal{F}, d_P) \leq \exp\left(\frac{2^{2r_0} b_0^2}{\varepsilon^{2r_0}}\right),\ \ 0 < \varepsilon < 1,$$

and as above both

$$N(\varepsilon, \mathcal{G}(\varepsilon), d_P) \leq N_{[\,]}(\varepsilon, \mathcal{G}(\varepsilon), d_P) \leq \left(N_{[\,]}(\varepsilon/2, \mathcal{F}, d_P)\right)^2 \leq \exp\left(2\frac{2^{2r_0} b_0^2}{\varepsilon^{2r_0}}\right)$$

and

$$N(\varepsilon, \mathcal{F}(\varepsilon), d_P) \leq N_{[\,]}(\varepsilon, \mathcal{F}(\varepsilon), d_P) \leq \left(N_{[\,]}(\varepsilon/2, \mathcal{F}, d_P)\right)^2 \leq \exp\left(2\frac{2^{2r_0} b_0^2}{\varepsilon^{2r_0}}\right).$$

Setting $\sigma = \varepsilon$ in (4.3) and (4.4) we get

$$J\,(\varepsilon, \mathcal{G}(\varepsilon)) \leq \sqrt{2} b_0 \int_{[0,\varepsilon]} \frac{ds}{s^{r_0}} \leq \frac{\sqrt{2} b_0}{1 - r_0}\varepsilon^{1-r_0}$$

and

$$a\,(\varepsilon, \mathcal{G}(\varepsilon)) = \frac{\varepsilon}{\sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}(\varepsilon), d_P)}} \geq \frac{\varepsilon^{1+r_0}}{\sqrt{2} b_0}.$$

Hence by the moment bound assuming (BR) given in (4.5) taken with $G\,(X) = M$,

$$\mu_n\,(\varepsilon) \leq A_3\left(\frac{\sqrt{2} b_0}{1 - r_0}\varepsilon^{1-r_0} + \sqrt{n}\,\mathbb{I}_{\left\{M > \frac{\sqrt{n}\varepsilon^{1+r_0}}{\sqrt{2} b_0}\right\}}\right)$$

and, since in the same way we have

$$J\,(\varepsilon, \mathcal{F}(\varepsilon)) \leq \frac{\sqrt{2} b_0}{1 - r_0}\varepsilon^{1-r_0} \text{ and } a\,(\varepsilon, \mathcal{F}(\varepsilon)) \geq \frac{\varepsilon^{1+r_0}}{\sqrt{2} b_0},$$

we get by the Gaussian moment inequality,

$$\mu\,(\varepsilon) \leq \frac{A_4\sqrt{2} b_0}{1 - r_0}\varepsilon^{1-r_0}.$$

As a consequence, for some $D > 0$ and

$$\varepsilon > \frac{(DM)^{1/(1+r_0)}}{n^{1/(2+2r_0)}}$$

it follows that

$$A\mu_n(\varepsilon) + \mu(\varepsilon) \le D\varepsilon^{1-r_0}.$$

Thus it is natural to take in (5.6) for some $\gamma_1 > 0$ and $\gamma_2 > 0$ large enough,

$$\delta = \gamma_1 \varepsilon^{1-r_0} \text{ and } t = \gamma_2 \varepsilon^{1-r_0},$$

which gives with $\rho = D + \gamma_1 + (A+1)\gamma_2$,

$$\begin{aligned}
&\mathbb{P}\left\{\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} > \rho\varepsilon^{1-r_0}\right\} \\
&\quad \le C_1 \exp\left(\frac{2^{2r_0+1}b_0^2}{\varepsilon^{2r_0}} - \frac{\gamma_1 C_2\sqrt{n}}{M}\varepsilon^{1-r_0}\exp\left(-\frac{5\left(2^{2r_0}b_0^2\right)}{2\varepsilon^{2r_0}}\right)\right) \\
&\qquad + 2\exp\left(-\frac{A_1\gamma_2\sqrt{n}}{M}\varepsilon^{1-r_0}\right) + 4\exp\left(-\frac{A_5\gamma_2^2}{\varepsilon^{2r_0}}\right).
\end{aligned}$$

We choose

$$\varepsilon = \left(\frac{10 b_0^2 2^{2r_0}}{\log n}\right)^{1/(2r_0)},$$

which makes

$$\exp\left(-\frac{5\left(2^{2r_0}b_0^2\right)}{2\varepsilon^{2r_0}}\right) = n^{-1/4}.$$

Given any $\lambda > 0$ we clearly see now from this last probability bound that for $\rho(\lambda) > 0$ made large enough by increasing γ_1 and γ_2 we get for all $n \ge 1$,

$$\mathbb{P}\left\{\|\alpha_n - \mathbb{G}\|_{\mathcal{F}} > \rho(\lambda)(\log n)^{-(1-r_0)/2r_0}\right\} \le n^{-\lambda}.$$

The proof of Proposition 2 now follows the same lines as that of Proposition 1. □

5.2. Proofs of strong approximations

Notice that the conditions on $\mathcal{F}$ in Propositions 1 and 2 imply that there exists a constant B such that

$$\sup_{n\ge 1}\mathbb{E}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\|_{\mathcal{F}}\right) \le B \text{ and } \mathbb{E}\left(\|\mathbb{G}\|_{\mathcal{F}}\right) \le B.$$

Therefore by Talagrand's inequality (4.1) and the Montgomery–Smith inequality (4.10) for all $n \ge 1$ and $t > 0$ we have, for suitable finite constants $C > 0$ and $C_1 > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\le m\le n}\sqrt{m}\|\alpha_m\|_{\mathcal{F}} > C\sqrt{n}(B+t)\right\} \\
&\quad \le 18\exp\left(-\frac{C_1 t^2}{\sigma_{\mathcal{F}}^2}\right) + 18\exp\left(-\frac{C_1 t\sqrt{n}}{M}\right),
\end{aligned} \tag{5.9}$$

where $\sigma_{\mathcal{F}}^2 := \sup_{f\in\mathcal{F}} Var(f(X))$. Furthermore, by Borell's inequality (4.7), the Montgomery–Smith inequality (4.10) and the fact that $n^{-1/2}\sum_{i=1}^{n}\mathbb{G}_i =_d \mathbb{G}$, for i.i.d. $\mathbb{G}_i$, we get for all $n \ge 1$ and $t > 0$ that for a suitable finite constant $D > 0$,

$$\mathbb{P}\left\{\max_{1\le m\le n}\left\|\sum_{i=1}^{m}\mathbb{G}_i\right\|_{\mathcal{F}} > D\sqrt{n}(B+t)\right\} \le 18\exp\left(-\frac{t^2}{2\sigma_{\mathcal{F}}^2}\right). \tag{5.10}$$

Proof of Theorem 1. Choose any $\gamma > 0$. We shall modify the scheme described on pages 236–238 of Philipp [31] to construct a probability space on which (2.4) and (2.5) hold. Let $n_0 = 1$ and for each $k \geq 1$ set $n_k = [k^\alpha]$, where $[x]$ denotes the integer part of x and α is chosen so that

$$1/2 < \tau_1 \alpha < 1. \tag{5.11}$$

Notice that $\tau_1 < 1/2$ in Proposition 1 and thus $\alpha > 1$.

Applying Proposition 1, we see that for each $\lambda > 1$ there exists a $\rho = \rho(\lambda) > 0$ such that one can construct a sequence of independent pairs $(\alpha_{n_k}^{(k)}, \mathbb{G}^{(k)})_{k \geq 1}$ sitting on the same probability space satisfying for all $k \geq 1$,

$$\mathbb{P}\left\{\left\|\alpha_{n_k}^{(k)} - \mathbb{G}^{(k)}\right\|_{\mathcal{F}} > \rho n_k^{-\tau_1} (\log n_k)^{\tau_2}\right\} \leq n_k^{-\lambda}. \tag{5.12}$$

Set for $k \geq 1$

$$t_k = \sum_{j<k} n_j \sim \frac{1}{1+\alpha} k^{\alpha+1}.$$

Using Lemma A1 of Berkes and Philipp [1] we can assume that each $\alpha_{n_k}^{(k)}$ is formed from $X_{t_k+1}, \ldots, X_{t_{k+1}}$ i.i.d. X and that each $\mathbb{G}^{(k)}$ is formed as

$$\mathbb{G}^{(k)} = \frac{1}{\sqrt{n_k}} \sum_{t_k < j \leq t_{k+1}} \mathbb{G}_j,$$

where $\mathbb{G}_{t_k+1}, \ldots, \mathbb{G}_{t_{k+1}}$ are i.i.d. $\mathbb{G}$. Moreover we can do this in such a way that $X_1, X_2 \ldots,$ are i.i.d. X and $\mathbb{G}_1, \mathbb{G}_2, \ldots,$ are i.i.d. $\mathbb{G}$. For any integer $N \geq 2$ set $N(\beta) = \left[N^\beta\right]$, where $\beta = \alpha/(1+\alpha)$. Define

$$s(N) = \sum_{k=N(\beta)}^{N} n_k^{1/2-\tau_1} (\log n_k)^{\tau_2}.$$

Now for some constants $c_1 > 0$ and $c > 0$,

$$s(N) \sim c_1 N^{(1+\alpha)/2-(\alpha\tau_1-1/2)} (\log N)^{\tau_2} \sim c(t_N)^{1/2-\tau(\alpha)} (\log t_N)^{\tau_2}, \tag{5.13}$$

where $\tau(\alpha) = (\alpha\tau_1 - 1/2)/(1+\alpha) > 0$, by (5.11).

We have

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1 \leq m \leq t_N} \left\|\sum_{j=1}^{m} [f(X_j) - \mathbb{E}f(X) - \mathbb{G}_j(f)]\right\|_{\mathcal{F}} > \rho s(N)\right\} \\
&\leq \mathbb{P}\left\{\max_{1 \leq m \leq t_{N(\beta)}} \left\|\sum_{j=1}^{m} [f(X_j) - \mathbb{E}f(X)]\right\|_{\mathcal{F}} > \frac{\rho s(N)}{4}\right\} \\
&+\mathbb{P}\left\{\max_{1 \leq m \leq t_{N(\beta)}} \left\|\sum_{j=1}^{m} \mathbb{G}_j(f)\right\|_{\mathcal{F}} > \frac{\rho s(N)}{4}\right\} \\
&+\sum_{k=N(\beta)}^{N-1} \mathbb{P}\left\{\max_{t_k+1 \leq m \leq t_{k+1}} \left\|\sum_{j=t_k+1}^{m} [f(X_j) - \mathbb{E}f(X)]\right\|_{\mathcal{F}} > \frac{\rho s(N)}{8}\right\}
\end{aligned}$$

$$+\sum_{k=N(\beta)}^{N-1}\mathbb{P}\left\{\max_{t_k+1\le m\le t_{k+1}}\left\|\sum_{j=t_k+1}^{m}\mathbb{G}_j(f)\right\|_{\mathcal{F}}>\frac{\rho s(N)}{8}\right\}$$

$$+\mathbb{P}\left\{\max_{N(\beta)\le j<N}\left\|\sum_{k=N(\beta)}^{j}\left(\sqrt{n_k}\alpha_{n_k}^{(k)}-\sqrt{n_k}\mathbb{G}^{(k)}\right)\right\|_{\mathcal{F}}>\frac{\rho s(N)}{4}\right\}=:\sum_{i=1}^{5}P_i(\rho,N).$$

It is easy to show using inequalities (5.9) and (5.10), along with the choice of $1/2<\beta=\alpha/(1+\alpha)<1$, that for any $\gamma>0$ for all large enough ρ,

$$\sum_{i=1}^{2}P_i(\rho,N)\le t_N^{-\gamma}/4,\ \text{for all } N\ge 1. \tag{5.14}$$

For instance, consider $P_1(\rho,N)$. Observe that

$$P_1(\rho,N)\le\mathbb{P}\left\{\max_{1\le m\le t_{N(\beta)}}\sqrt{m}||\alpha_m||_{\mathcal{F}}>C\sqrt{t_{N(\beta)}}\,(B+\tau_N)\right\},$$

where

$$\tau_N=\left(\frac{\rho s(N)}{4}-B\right)/\left(C\sqrt{t_{N(\beta)}}\right).$$

Now $\sqrt{t_{N(\beta)}}\sim c_2N^{\alpha/2}$ for some $c_2>0$. Therefore by (5.13) for some $c_3>0$,

$$\tau_N\sim c_3N^{1-\tau_1\alpha}(\log N)^{\tau_2}.$$

Since by (5.11) we have $1-\tau_1\alpha>0$, we readily get from inequality (5.9) that for any $\gamma>0$ and all large enough ρ, $P_1(\rho,N)\le t_N^{-\gamma}/8$, for all $N\ge 1$. In the same way we get using inequality (5.10) that for any $\gamma>0$ and all large enough ρ, $P_2(\rho,N)\le t_N^{-\gamma}/8$, for all $N\ge 1$. Hence we have (5.14).

In a similar fashion one can verify that for any $\gamma>0$ and all large enough ρ,

$$\sum_{i=3}^{4}P_i(\rho,N)\le t_N^{-\gamma}/4,\ \text{for all } N\ge 1. \tag{5.15}$$

To see this, notice that

$$P_3(\rho,N)\le N\mathbb{P}\left\{\max_{1\le m\le n_N}\sqrt{m}||\alpha_m||_{\mathcal{F}}>\rho s(N)/8\right\}$$

and

$$P_4(\rho,N)\le N\mathbb{P}\left\{\max_{1\le m\le n_N}||\sum_{j=1}^{m}\mathbb{G}_j(f)||_{\mathcal{F}}>\rho s(N)/8\right\}.$$

Since $\sqrt{n_N}\sim N^{\alpha/2}$ and $N\sim c_3t_N^{1/(\alpha+1)}$ for some $c_3>0$, we get (5.15) by proceeding as above using inequalities (5.9) and (5.10).

Next, recalling the definition of $s(N)$, we get

$$P_5(\rho,N)\le\mathbb{P}\left\{\sum_{k=N(\beta)}^{N}\left\|\sqrt{n_k}\alpha_{n_k}^{(k)}-\sqrt{n_k}\mathbb{G}^{(k)}\right\|_{\mathcal{F}}>\frac{\rho s(N)}{4}\right\}$$

$$\le\sum_{k=N(\beta)}^{N}\mathbb{P}\left\{\left\|\sqrt{n_k}\alpha_{n_k}^{(k)}-\sqrt{n_k}\mathbb{G}^{(k)}\right\|_{\mathcal{F}}>\frac{\rho n_k^{1/2-\tau_1}(\log n_k)^{\tau_2}}{4}\right\},$$

which by (5.12) for any $\lambda > 0$ and $\rho = \rho(\alpha, \lambda) > 0$ large enough is

$$\leq N\left(\left[N^{\beta}\right]^{\alpha}\right)^{-\lambda}, \text{ for all } N \geq 1,$$

which, in turn, for large enough $\lambda > 0$ is $\leq t_N^{-\gamma}/2$. Thus for all $\gamma > 0$ there exists a $\rho > 0$ so that

$$\sum_{i=1}^{5} P_i(\rho, N) \leq t_N^{-\gamma}, \text{ for all } N \geq 1.$$

Since α can be any number satisfying $1/2 < \tau_1\alpha < 1$ and $t_{N+1}/t_N \to 1$, this implies (2.4) for $\rho = \rho(\alpha, \lambda)$ large enough. The almost sure statement (2.5) follows trivially from (2.4) using a simple blocking and the Borel–Cantelli lemma on the just constructed probability space. This proves Theorem 1. □

Proof of Theorem 2. The proof follows along the same lines as that of Theorem 1. Therefore for the sake of brevity we shall only outline the proof. Here we borrow ideas from the proof of Theorem 6.2 of Dudley and Philipp [14]. Recall that in Theorem 2 we assume that $1/2 < r_0 < 1$ in Proposition 2, which means that $0 < \kappa := (1 - r_0)/2r_0 < 1/2$. For $k \geq 1$ set

$$t_k = \left[\exp\left(k^{1-\kappa}\right)\right] \text{ and } n_k = t_k - t_{k-1}, \text{ where } t_0 = 1. \tag{5.16}$$

Now for some $b > 0$ we get $n_k \sim b^2 k^{-\kappa} t_k$,

$$\frac{\sqrt{n_k}}{(\log n_k)^{\kappa}} \sim \frac{b\sqrt{t_k}}{k^{\kappa(1-\kappa)+\kappa/2}} = \frac{b\sqrt{t_k}}{k^{\kappa+\theta}},$$

where $\theta = \kappa\left(\frac{1}{2} - \kappa\right) > 0$. Choose $0 < \beta < 1$ and set $N(\beta) = \left[N^{\beta}\right]$. Using an integral approximation we get for suitable constants $c_1 > 0$ and $c_2 > 0$, for all large N

$$\frac{c_1\sqrt{t_N}}{N^{\theta}} \leq s(N) := \sum_{k=N(\beta)}^{N} \frac{\sqrt{n_k}}{(\log n_k)^{\kappa}} \leq \frac{c_2\sqrt{t_N}}{N^{\theta}} \leq \frac{c_2\sqrt{t_N}}{(\log(t_N))^{\theta/(1-\kappa)}}. \tag{5.17}$$

Also for all large N,

$$s(N)/\sqrt{n_N} \geq \frac{c_1}{2b} N^{\kappa/2-\kappa\left(\frac{1}{2}-\kappa\right)} =: c_0 N^{\kappa^2}. \tag{5.18}$$

For later use note that for any $0 < \beta < 1$ and $\zeta > 0$

$$\frac{s(N)}{\sqrt{t_{N(\beta)}}N^{\zeta}} \to \infty, \text{ as } N \to \infty, \tag{5.19}$$

and observe that

$$t_{N+1}/t_N \to 1, \text{ as } N \to \infty. \tag{5.20}$$

Constructing a probability space and defining $P_i(\rho, N)$, $i = 1, \ldots, 5$, as in the proof of Theorem 1, but with n_k, t_k and $s(N)$ as given in (5.16) and (5.17) the proof now goes much like that of Theorem 1. In particular, using inequalities (5.9) and (5.10), and noting that $N \sim (\log(t_N))^{1/(1-\kappa)}$, one can check that for some $\nu > 0$, for all large enough N,

$$\sum_{i=1}^{4} P_i(\rho, N) \leq \exp\left(-(\log(t_N))^{\nu}\right)$$

and by arguing as in the proof of Theorem 1, but now using Proposition 2, we easily see that for every $H > 0$ there is a probability space on which sit i.i.d. $X_1, X_2 ...$, and i.i.d. $\mathbb{G}_1, \mathbb{G}_2, \ldots$, and a $\rho > 0$ such that

$$P_5(\rho, N) \leq (\log(t_N))^{-H-1}, \quad \text{for all } N \geq 1.$$

Since for all $H > 0$,

$$\log(t_N)^H \left(\exp\left(-(\log(t_N))^\nu\right) + (\log(t_N))^{-H-1}\right) \to 0, \text{ as } N \to \infty,$$

this in combination with (5.17) and (5.20) proves that (2.8) holds with $\tau = \theta/(1-\kappa)$ and $\rho(\tau, H)$ large enough. A simple blocking argument shows that (2.9) follows from (2.8). Choose $H > 1$ in (2.8). Notice that for any $k \geq 1$,

$$\begin{aligned}
&\mathbb{P}\left\{\bigcup_{2^k < n \leq 2^{k+1}} \left\{\max_{1 \leq m \leq n} \left\| \sqrt{m}\alpha_m - \sum_{i=1}^m \mathbb{G}_i \right\|_{\mathcal{F}} > \sqrt{2n}\rho(\tau, H)(\log n)^{-\tau}\right\}\right\} \\
&\quad \leq \mathbb{P}\left\{\max_{1 \leq m \leq 2^{k+1}} \left\| \sqrt{m}\alpha_m - \sum_{i=1}^m \mathbb{G}_i \right\|_{\mathcal{F}} > \sqrt{2^{k+1}}\rho(\tau, H)(\log 2^{k+1})^{-\tau}\right\} \\
&\quad \leq ((k+1)\log 2)^{-H}.
\end{aligned}$$

Hence (2.9) holds by the Borel-Cantelli lemma. □

Acknowledgement. The authors thank the referee for pointing out a number of oversights and misprints, as well as showing the way to an improvement in Theorem 2.

References

[1] Berkes, I. and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 29–54.

[2] Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216.

[3] Bovier, A. and Mason, D. M. (2001). Extreme value behavior in the Hopfield model. *Ann. Appl. Probab.* **11** 91–120.

[4] Bretagnolle, J. and Massart, P. (1989). Hungarian construction from the non asymptotic viewpoint. *Ann. Probab.* **17** 239–256.

[5] Castelle, N. (2002). Approximations fortes pour des processus bivariés. *Canad. J. Math.* **54** 533–553.

[6] Castelle, N. and Laurent-Bonvalot, F. (1998). Strong approximation of bivariate uniform empirical processes. *Ann. Inst. H. Poincaré* **34** 425–480.

[7] de la Peña, V. H. and Giné, E. (1999). *Decoupling. From Dependence to Independence. Randomly Stopped Processes. U-statistics and Processes. Martingales and Beyond.* Probability and Its Applications. Springer-Verlag, New York.

[8] Dehling, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrsch. Verw. Gebiete* **63** 393–432.

[9] Dudley, R. M. (1967). The sizes of compact subsets of an Hilbert space and continuity of Gaussian processes. *J. of Funct. Anal.* **1** 290–330.

[10] Dudley, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* **1** 66–103.
[11] Dudley, R. M. (1989). *Real Analysis and Probability.* Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
[12] Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge Univ. Press.
[13] Dudley, R. M. (2000). Notes on empirical processes. Lectures notes for a course given at Aarhus University, Denmark, August 1999, January 2000 revision.
[14] Dudley, R. M. and Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete* **62** 509–552.
[15] Einmahl, U. and Kuelbs, J. (2001). Cluster sets for a generalized law of the iterated logarithm in Banach spaces. *Ann. Probab.* **29** 1451–1475.
[16] Einmahl, U. and Kuelbs, J. (2004). Moderate deviation probabilities for open convex sets: Nonlogarithmic behavior. *Ann. Probab.* **32** 1316–1355.
[17] Einmahl, U. and Mason, D. M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Th. Rel. Fields* **107** 283–311.
[18] Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** 1–37.
[19] Gentz, B. and Löwe, M. (1999). The fluctuations of the overlap in the Hopfield model with finitely many patterns at the critical temperature. *Probab. Theory Related Fields* **115** 357–381.
[20] Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré* **37** 503–522.
[21] Koltchinskii, V. I. (1994). Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab.* **7** 73–118.
[22] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df I. *Z. Wahrsch. verw. Gebiete.* **32** 111–131.
[23] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces. Isoperimetry and Processes.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 23. Springer-Verlag, Berlin.
[24] Marcus, M. D. and Pisier. G. (1981). *Random Fourier Series with Applications to Harmonic Analysis.* Annals of Mathematics Studies, Vol. 101. Princeton University Press, Princeton, NJ.
[25] Mason, D. M. (2001). Notes on the KMT Brownian bridge approximation to the uniform empirical process. Asymptotic methods in probability and statistics with applications (St. Petersburg, 1998). *Stat. Ind. Technol.* Birkhäuser Boston, Boston, MA, pp. 351–369.
[26] Mason, D. M. (2004). A uniform functional law of the logarithm for the local empirical process. *Ann. Probab.* **32** 1391–1418.
[27] Mason, D. M. (2006). Some observations on the KMT dyadic scheme. Special issue of *Journal of Statistical Planning and Inference.* To appear.
[28] Mason, D. M. and van Zwet, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884.
[29] Massart, P. (1989). Strong approximations for multidimensional empirical and related processes, via KMT constructions. *Ann. Probab.* **17** 266–291.

[30] Montgomery-Smith, S. (1993). Comparison of sums of independent identically distributed random variables. *Prob. Math. Statist.* **14** 281–285.

[31] Philipp, W. (1986). Invariance principles for independent and weakly dependent random variables. Dependence in probability and statistics (Oberwolfach, 1985), *Progr. Probab. Statist.* **11**. Birkhäuser Boston, Boston, MA, pp. 225–268.

[32] Révesz, P. (1976). On strong approximation of the multidimensional empirical process. *Ann. Probab.* **4** 729–743.

[33] Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Th. Rel. Fields* **98** 21–45.

[34] Rio, E. (1996). Vitesses de convergence dans le principe d'invariance faible pour la fonction de répartition empirique multivariée. *C. R. Acad. Sci. Paris Série I* **322** 169–172.

[35] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76.

[36] Tusnády, G. (1977). A remark on the approximation of the sample DF in the multidimensional case. *Period. Math. Hungar.* **8** 53–55.

[37] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

[38] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

[39] Vapnik, V. N. and Červonekis, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280.

[40] Vorob'ev, N. N. (1962). Consistent families of measures and their extensions. *Theory Probab. Appl.* **7** 147–163.

[41] Yurinskii, V. V. (1977). On the error of the Gaussian approximation for convolutions. *Theor. Probab. Appl.* **22** 236–247.

[42] Zaitsev, A. Yu. (1987a). Estimates of the Lévy-Prokhorov distance in the multivariate central limit theorem for random variables with finite exponential moments. *Theory Probab. Appl.* **31** 203–220.

[43] Zaitsev, A. Yu. (1987b). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probab. Th. Rel. Fields* **74** 534–566.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 173–184

DOI: 10.1214/074921706000000833

Modified empirical CLT's under only pre-Gaussian conditions

Shahar Mendelson[1,*] **and Joel Zinn**[2,†]

ANU & Technion I.I.T and Texas A&M University

Abstract: We show that a modified Empirical process converges to the limiting Gaussian process whenever the limit is continuous. The modification depends on the properties of the limit via Talagrand's characterization of the continuity of Gaussian processes.

1. Introduction

Given a class of functions $F \subset L_2(\mu)$, a now standard method to use (iid) data to uniformly estimate or predict the mean of one of the functions in the class, is through the use of empirical processes. One has to bound the random variable

$$\sup_{f \in F} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E} f \right| \equiv \|P_n - P\|_F.$$

One possibility that comes to mind is to use the fact that there is a bounded Gaussian process indexed by F to bound $\sqrt{n}\|P_n - P\|_F$. To illustrate the difficulty one encounters, note that to use the classical Central Limit Theorem in finite dimensions, one only needs finite variance or, equivalently, the existence of the (Gaussian) limit. However, when working in the infinite dimensional situation there are sometimes non-trivial side conditions other than just the set being pregaussian. Those are connected to the random geometry of the set F that are needed to ensure the existence of a useful bound. For example, one such situation is when the class is a collection of indicators of sets. If such a class does not satisfy the VC (Vapnik-Červonenkis) condition, then, in addition to knowing that the limiting Gaussian process is continuous, one has to check, for example, the following:

$$\frac{\log \#\{C \cap \{X_1, \ldots, X_n\} : C \in \mathcal{C}\}}{\sqrt{n}} \to 0 \text{ in outer probability .}$$

In this note we try to get around this problem by looking at a variant of the standard empirical process for which only the existence of the limiting Gaussian process is required to obtain both tail estimates and the Central Limit Theorem for the modified process.

The motivation for our study were the articles [7, 8], which focus on the following problem. Consider a density $p(x)$ on $\mathbb{R}$ which has a support that is a finite union

[1]The Australian National University, Canberra, ACT 0200, Australia and, Technion - Israel Institute of Technology, Haifa 32000, Israel, e-mail: shahar.mendelson@anu.edu.au

[2]Department of Mathematics, Texas A&M University, College Station, TX 77843-3386, e-mail: jzinn@math.tamu.edu

*Research partially supported by an Australian Research Council Discovery Grant DP0343616.

†Research partially supported by NSA Grant H98230-04-1-0108.

AMS 2000 subject classifications: primary 60F05; secondary 60F17.

Keywords and phrases: central limit theorems, empirical processes.

of intervals, and on this support p satisfies that $c^{-1} \le p(x) \le c$ and $|p(x) - p(y)| \le c|x - y|$. It was shown in [7, 8] that under this assumption there is a histogram rule $\tilde{P}_n$ for which the following holds. If F is a P-pregaussian class of indicator functions, or if F is a P-pregaussian class of functions bounded by 1 which satisfy a certain metric entropy condition, then $\sqrt{n}(\tilde{P}_n - P)$ converges weakly to the limiting Gaussian process.

It seems that the method used in [7, 8] can not be extended to densities in $\mathbb{R}^2$, and even in the one-dimensional case the convergence result holds for a very restricted set of densities. Thus, our aim is to find some other empirical estimator for which the existence of the limiting Gaussian would imply convergence as above. Our estimator is based on Theorem 1 in [9] (see also Theorem 1.3 [4] and Theorem 14.8 [6]).

We begin with several facts and notation we will use throughout this note. If G is a (centered) Gaussian process indexed by the set T, then for every $s, t \in T$, $\rho_2(s,t) = \left(\mathbb{E}(G_s - G_t)^2\right)^{1/2}$.

To prove that a process (and in particular, the modified empirical process we define) converges to the Gaussian process indexed by F, we require the notion of asymptotic equicontinuity.

Definition 1.1. A net $X_\alpha : \Omega_\alpha \longrightarrow \ell_\infty(T)$ is asymptotically uniformly ρ-equicontinuous in probability, if for every $\epsilon, \eta > 0$, there exists a $\delta > 0$ such that

$$\text{limsup}_\alpha \Pr{}^*\Big(\sup_{\rho(s,t)<\delta} |X_\alpha(s) - X_\alpha(t)| > \epsilon\Big) < \eta.$$

Theorem 1.2 ([13], 1.5.10). *Let G be a Gaussian process and let X_α be a net of random variables with values in $\ell_\infty(T)$. Then there exists a version of G which is a tight Borel measurable map into $\ell_\infty(T)$, and X_α converges weakly to G if and only if*

(i) *the finite dimensional distributions of X_α converge weakly to the corresponding finite dimensional distributions of G,*
(ii) *X_α is asymptotically uniformly equicontinuous in probability with respect to ρ_2,*
(iii) *(T, ρ_2) is totally bounded.*

The main technical tool we require is *generic chaining* which was developed by Talagrand (see [11] for the most recent survey on this topic).

Definition 1.3. For a metric space (T, d), an *admissible sequence* of T is a collection of subsets of T, $\{T_s : s \ge 0\}$, such that for every $s \ge 1$, $|T_s| = 2^{2^s}$ and $|T_0| = 1$. For $p \ge 1$, we define the γ_p functional by

$$\gamma_p(T, d) = \inf \sup_{t \in T} \sum_{s=0}^{\infty} 2^{s/p} d(t, T_s),$$

where the infimum is taken with respect to all admissible sequences of T.

For every integer s define the function $\pi_s : T \to T_s$, which maps every $t \in T$ to a nearest element to t in T_s.

Using the γ_p functionals one can bound the supremum of a process which satisfies an increment condition. As an example, recall the well known Bernstein inequality.

Theorem 1.4. *There exists an absolute constant c for which the following holds. Let $(\Omega, \mathcal{S}, P)$ be a probability space, let $f \in L_\infty(P)$ and let $X_1, \ldots, X_n$ be independent random variables distributed according to P. Then*

$$Pr\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}f\right| \geq t\right\}\right) \leq 2\exp\left(-cn\min\left\{\frac{t^2}{\|f\|_{L_2}^2}, \frac{t}{\|f\|_{L_\infty}}\right\}\right).$$

When one combines Bernstein's inequality with the generic chaining method, the following corollary is evident:

Corollary 1.5 ([11]). *There exists an absolute constant c for which the following holds. Let F be a class of functions on a probability space (Ω, P). Then, for every integer n,*

$$\mathbb{E}\|P_n - P\|_F \leq c\left(\frac{\gamma_2(F, \|\cdot\|_2)}{\sqrt{n}} + \frac{\gamma_1(F, \|\cdot\|_\infty)}{n}\right).$$

A key result, which follows from the generic chaining method, is that the expectation of the supremum of the Gaussian process indexed by F and has a covariance structure endowed by $L_2(P)$ is finite if and only if $\gamma_2(F, L_2(P))$ is finite. Moreover, the result is quantitative in nature.

Theorem 1.6. *There exist absolute constants c_1 and c_2 for which the following holds. Let F be a class of functions on (Ω, P), and let G be the Gaussian process indexed by F. Then,*

$$c_1\gamma_2(F, L_2(P)) \leq \mathbb{E}\sup_{f\in F}|G_f| \leq c_2\gamma_2(F, L_2(P)).$$

The upper bound is due to Fernique [3], while the lower bound in due to Talagrand [10]. A proof of both parts can be found in [11]. From here on we denote $\mathbb{E}\sup_f |G_f|$ by $\mathbb{E}\|G\|_F$.

In a similar fashion one can formulate a continuity condition for the Gaussian process indexed by F using the generic chaining machinery.

Theorem 1.7 ([11] Theorem 1.4.1). *Consider a Gaussian process, $\{G_f : f \in F\}$, where F is countable. Then the following are equivalent:*

(1) *The map $f \longrightarrow G_f(\omega)$ is uniformly continuous on $(F, \|\cdot\|_2)$ with probability one.*
(2) *We have*
$$\lim_{\epsilon\longrightarrow 0}\mathbb{E}\sup_{\|G_f - G_{f'}\|_2\leq\epsilon}|G_f - G_{f'}| = 0.$$
(3) *There exist an admissible sequence of partitions of F such that*
$$\lim_{s_0\longrightarrow\infty}\sup_{f\in F}\sum_{s\geq s_0} 2^{s/2}d(f, F_s) = 0.$$

Note that the admissible sequence in (3) can be taken as an almost optimal admissible sequence in the definition of $\gamma_2(F, \|\cdot\|_2)$, at a price of an absolute constant. Indeed, by combining the two admissible sequences (T_s) and (T_s') - the first - an almost optimal one from the definition of γ_2 and the second from (1.7), we obtain a new admissible sequence for which $T_s \cup T_s' \subset T_{s+1}''$. Thus, we may assume that the almost optimal sequence in (3) satisfies that $\sup_{f\in F}\sum_{s=0}^{\infty} 2^{s/2}d_{\|\cdot\|_2}(t, T_s) \sim \mathbb{E}\|G\|_F$.

Finally, a notational convention. Throughout, all absolute constants will be denoted by c, C or K. Their values may change from line to line. We write $a \sim b$ if there are absolute constants c and C such that $ca \leq b \leq Ca$.

2. The main theorems

Let $F \subset L_2(P)$ and assume that there exists a Gaussian process indexed by F such that $\mathbb{E}\|G\|_F < \infty$. By Theorem 1.6, $\gamma_2(F, L_2(P)) \sim \mathbb{E}\|G\|_F$, and let $(F_s)_{s=0}^{\infty}$, $F_s \subset F$ be an almost optimal admissible sequence with respect to the $L_2(P)$ norm as described above, that is,

$$\sup_{f \in F} \sum_{s=0}^{\infty} 2^{s/2} \|f - \pi_s(f)\|_{L_2(P)} \leq c\gamma_2(F, L_2(P)),$$

and

$$\lim_{s_0 \to \infty} \sup_{f \in F} \sum_{s \geq s_0} 2^{s/2} \|f - \pi_s(f)\|_{L_2(P)} = 0.$$

Set $\Delta_s(f) = \pi_s(f) - \pi_{s-1}(f)$. For every s, f and $\lambda \geq 0$ consider a truncated part of $\Delta_s(f)$, defined by $\Delta_s'(f,\lambda) = \Delta_s(f)\mathbb{1}_{\{|\Delta_s(f)| \leq \lambda\}}$. As will be made clear shortly, the truncation level λ depends both on the specific $f \in F$, as well as on the size of the sample n and the index s.

Lemma 2.1. *There exist absolute constants c_1, c_2 and c_3 for which the following holds. Let F be a class of functions on the probability space (Ω, P), and set $X_1, \ldots, X_n$ to be independent, distributed according to P. Let $(F_s)_{s=0}^{\infty}$ be an admissible sequence in F with respect to $L_2(P)$, and for every $f \in F$ set $\lambda = c_1\sqrt{n}\|\Delta_s(f)\|_2/2^{s/2}$. Then, for every $u > 1/2$ and every integer s, with probability at least $1 - 2\exp(-c_2 2^s \min\{u^2, u\})$, for every $f \in F$.*

$$\left| \frac{1}{n} \sum_{i=1}^{n} \left(\Delta_s'(f,\lambda)\right)(X_i) - \mathbb{E}\Delta_s'(f,\lambda) \right| \leq c_3 u \frac{2^{s/2}\|\Delta_s(f)\|_2}{\sqrt{n}}. \tag{2.1}$$

Proof. Let c_1 and c_2 be constants to be named later. By Bernstein's inequality [2], for $t = c_1 2^{s/2} u \|\Delta_s(f)\|_2/\sqrt{n}$, it is evident that for any $\lambda > 0$,

$$\begin{aligned} &Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} \left(\Delta_s'(f,\lambda)\right)(X_i) - \mathbb{E}\Delta_s'(f,\lambda) \right| \geq t \right) \\ \leq &2\exp\left(-cn \min\left\{ \frac{t^2}{\|\Delta_s'(f,\lambda)\|_2^2}, \frac{t}{\lambda} \right\} \right). \end{aligned}$$

Since $\|\Delta_s'(f,\lambda)\|_2 \leq \|\Delta_s(f)\|_2$ and $\lambda = c_2\sqrt{n}\|\Delta_s(f)\|_2/2^{s/2}$, then for the choice of t, it follows that with probability at least $1 - 2\exp(-c_3 2^s \min\{u^2, u\})$,

$$\left| \frac{1}{n} \sum_{i=1}^{n} \left(\Delta_s'(f,\lambda)\right)(X_i) - \mathbb{E}\Delta_s'(f,\lambda) \right| \leq c_4 u \frac{2^{s/2}\|\Delta_s(f)\|_2}{\sqrt{n}},$$

where c_3 and c_4 depend only on c_1 and c_2. Thus, for an appropriate choice of these constants and since $|\{\Delta_s(f) : f \in F\}| \leq |F_s| \cdot |F_{s-1}| \leq 2^{2^{s+1}}$ the claim follows. □

Note that there is nothing magical with the lower bound of $1/2$ on u. Any other absolute constant would do, and would lead to changed absolute constants c_1, c_2 and c_3.

Using the Lemma we can define a process Φ_n for which $\|P_n(\Phi_n) - P\|_F \leq c\mathbb{E}\|G\|_F/\sqrt{n}$, and thus, the fact that the limiting Gaussian process exists suffices to yield a useful bound on the way in which the empirical estimator approximates the mean.

Definition 2.2. Let $\lambda(f,n,s) \equiv \lambda = c_0\sqrt{n}\|\Delta_s(f)\|_2/2^{s/2}$, where c_0 was determined in Lemma 2.1, and for each s_0 let

$$\Phi_{n,s_0}(f) = \sum_{s=s_0+1}^{\infty} \Delta'_s(f,\lambda)$$

and

$$\Phi_n(f) = \sum_{s=1}^{\infty} \Delta'_s(f,\lambda)$$

Theorem 2.3. *There exist absolute constants c_1 and c_2 for which the following holds. Let F and $X_1,\ldots,X_n$ be as above. Then, the mapping $\Phi_n : F \to L_1(P)$, and for every $u > 1/2$, with probability at least*

$$1 - 2\exp(-c_1\min\{u^2,u\}),$$

for every $f \in F$

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\Phi_n(f)\right)(X_i) - \mathbb{E}f\right| \le c_2(u+1)\frac{\mathbb{E}\|G\|_F}{\sqrt{n}}. \tag{2.2}$$

and also with probability at least $1 - 2\exp(-c_1 2^{s_0}\min\{u^2,u\})$

$$\sup_{f\in F}\left|\frac{1}{n}\sum_{i=1}^{n}\Big(\Phi_{n,s_0}(f)(X_i) - \mathbb{E}\Phi_{n,s_0}(f)\Big)\right| \le \frac{c_2 u}{\sqrt{n}}\sup_{f\in F}\sum_{s=s_0+1}^{\infty} 2^{s/2}\|\Delta_s(f)\|_2. \tag{2.3}$$

Proof. Without loss of generality, assume that $0 \in F$ and that $\pi_0 = 0$. Let $(F_s)_{s=1}^{\infty}$ be an almost optimal admissible sequence, and in particular, by Theorem 1.6

$$\sup_{f\in F}\sum_{s=1}^{\infty} 2^{s/2}\|\Delta_s(f)\|_2 \le K\mathbb{E}\|G\|_F$$

for a suitable absolute constant K.

Note that as $\pi_0(f) = 0$ for every $f \in F$ one can write $f = \sum_{s=1}^{\infty}\Delta_s(f)$. Let us show that Φ_n, and therefore Φ_{n,s_0}, are well defined and maps F into $L_1(P)$. Indeed, since $\sum_{s=1}^{\infty}\Delta_s(f)$ converges in $L_2(P)$, it suffices to prove that $\sum_{s=1}^{\infty}\big(\Delta_s(f) - \Delta'_s(f,\lambda)\big) \equiv \sum_{s=1}^{\infty}\Delta''_s(f)$ converges in $L_1(P)$. Observe that for every $f \in F$,

$$\begin{aligned}
\mathbb{E}|\Delta''_s(f,\lambda)| &= \mathbb{E}|\Delta_s(f)|\mathbb{1}_{\{|\Delta_s(f)|>\lambda\}} \\
&\le \|\Delta_s(f)\|_2\left(Pr\left(|\Delta_s(f)|>\lambda\right)\right)^{1/2} \le \frac{\|\Delta_s(f)\|_2^2}{\lambda} \\
&\le \frac{2^{s/2}\|\Delta_s(f)\|_2}{c_0\sqrt{n}}.
\end{aligned} \tag{2.4}$$

Since $\sum_{s=1}^{\infty} 2^{s/2}\|\Delta_s(f)\|_2$ converges for every f, it implies that Φ_n is well defined and takes values in L_1.

By Lemma 2.1, with probability at least

$$1 - 2\sum_{s=s_0+1}^{\infty}\exp(-c_1 2^s\min\{u^2,u\}) \ge 1 - 2\exp(-c_2 2^{s_0}\min\{u^2,u\}), \tag{2.5}$$

$$\sup_{f\in F}\sum_{s=s_0+1}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\Delta_s'(f,\lambda)\right)(X_i)-\mathbb{E}\Delta_s'(f,\lambda)\right|$$
$$\leq \frac{c_3u}{\sqrt{n}}\sup_{f\in F}\sum_{s=s_0+1}^{\infty}2^{s/2}\|\Delta_s(f)\|_2,$$

and when $s_0=0$ we'll use that by Theorem 1.6 this last quantity is

$$\leq c_4u\frac{\mathbb{E}\|G\|_F}{\sqrt{n}}.$$

Hence, with that probability, for every $f\in F$,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\Phi_n(f)(X_i)-\mathbb{E}f\right)\right|\leq\left|\frac{1}{n}\sum_{i=1}^{n}\left(\Phi_n(f)(X_i)-\mathbb{E}\Phi_n(f)\right)\right|+|\mathbb{E}f-\mathbb{E}\Phi_n(f)|$$
$$\leq c_5u\frac{\mathbb{E}\|G\|_F}{\sqrt{n}}+\left|\mathbb{E}\sum_{s=1}^{\infty}\Delta_s''(f)\right|\leq c_6(u+1)\frac{\mathbb{E}\|G\|_F}{\sqrt{n}},$$

where the last term is estimated using the same argument as in (2.4) and the inequality (2.2) in Theorem 2.3.

We also have that with the promised lower bound on the probability,

$$\sup_{f\in F}\left|\frac{1}{n}\sum_{i=1}^{n}\Big(\Phi_{n,s_0}(f)(X_i)-\mathbb{E}\Phi_{n,s_0}(f)\Big)\right|\leq\frac{c_3u}{\sqrt{n}}\sup_{f\in F}\sum_{s=s_0+1}^{\infty}2^{s/2}\|\Delta_s(f)\|_2.$$
□

Next, we prove a limit theorem for $\{\sqrt{n}(P_n-P)(\Phi_n)(f):f\in F\}$ and show that we can replace $\mathbb{E}\Phi_n(f)$ with $\mathbb{E}f$ and still obtain a limit theorem. For this we need to prove an inequality for the oscillation of the process $\sqrt{n}(P_n-P)(\Phi_n(f))$. To that end, define $Q_n:=\sqrt{n}\,(P_n-P)$.

Proposition 2.4. *Let F be a class of functions on (Ω,P), such that the Gaussian process indexed by F exists and is continuous. If Φ_n is as above, then for any $\eta>0$,*

$$\lim_{\delta\to 0}\lim_{n\to\infty}Pr\left(\sup_{\|f-\tilde{f}\|_2<\delta}\left|Q_n(\Phi_n(f)-\Phi_n(\tilde{f}))\right|>\eta\right)=0.$$

Proof. By the definition of Φ_n which uses an almost optimal admissible sequence, for every $\delta>0$ there is some s_0 such that

$$\sup_{f\in F}\sum_{s=s_0}^{\infty}2^{s/2}\|f-\pi_s(f)\|_2\leq\delta,$$

hence for any $f,\tilde{f}\in F$, $\|\pi_{s_0}(f)-\pi_{s_0}(\tilde{f})\|_2<2\delta+\|f-\tilde{f}\|_2$. Using the notation of

Theorem 2.3, put $\Phi_{n,s_0}(f) := \sum_{s=s_0+1}^{\infty} \Delta'_s(f,\lambda)$ and $\Psi_{n,s_0} = \Phi_n - \Phi_{n,s_0}$. Therefore,

$$
\begin{aligned}
I :&= Pr\left(\sup_{\|f-\tilde f\|_2<\delta} \left| Q_n\left(\Phi_n(f)-\Phi_n(\tilde f)\right)\right| > \eta \right) \\
&= Pr\Bigg(\sup_{\|f-\tilde f\|_2<\delta} \Big| Q_n\left(\Psi_{n,s_0}(f)-\Psi_{n,s_0}(\tilde f)\right) \\
&\qquad + Q_n(\Phi_{n,s_0}(f)) - Q_n(\Phi_{n,s_0}(\tilde f))\Big| > \eta \Bigg) \\
&\leq Pr\left(\sup_{\|\pi_{s_0} f-\pi_{s_0}\tilde f\|_2<3\delta} \left| Q_n\left(\Psi_{n,s_0}(f)-\Psi_{n,s_0}(\tilde f)\right)\right| > \frac{\eta}{3} \right) \\
&\qquad + 2Pr\left(\sup_f |Q_n(\Phi_{n,s_0}(f))| > \frac{\eta}{3} \right) \\
&:= (II) + (III)
\end{aligned}
$$

From the proof of Theorem 2.3 by integrating tail probabilities

$$
\mathbb{E} \sup_{f\in F} |Q_n\left(\Phi_{n,s_0}(f)\right)| \leq c \sup_{f\in F} \sum_{s>s_0} 2^{s/2}\|f-\pi_s(f)\|_2
$$

which by Theorem 1.7(3) and our choice of the admissible sequence converges to 0 as $s_0 \to \infty$. Furthermore, by the finite dimensional Central Limit Theorem $\lim_{\delta\to 0}\lim_{n\to\infty}(II) = 0$, which completes the proof. □

Hence, we know that Q_n is asymptotically uniformly equicontinuous. We'll now prove the other necessary ingredients needed to show that Q_n converges to the original Gaussian process.

Proposition 2.5. *Let F and Φ_n be as in Proposition 2.4. Then the following holds:*

(i) $\lim_{n\to\infty} \sqrt{n} \sup_{f\in F} |\mathbb{E}\Phi_n(f) - \mathbb{E}f| = 0$,
(ii) *For every $f \in F$, $\lim_{n\to\infty} \|\Phi_n(f) - f\|_2 = 0$,*
(iii) *For every $f \in F$, $\lim_{n\to\infty} \mathbb{E}\max_{j\leq n} \dfrac{|\Phi_n(f)(X_j)|^2}{n} = 0$.*

Proof.

1. Let s_0 be an integer to be named later and set

$$
\lambda = c_0\sqrt{n}\|\Delta_s(f)\|_2/2^{s/2}
$$

as in Lemma 2.1. In particular, the set $\{\Delta_s(f) : s \leq s_0, f \in F\}$ is finite and for every $f \in F$,

$$
\sqrt{n}\mathbb{E}|\Delta_s(f)|\mathbb{1}_{\{|\Delta_s(f)|>\lambda\}} \leq \frac{2^{s/2}}{c_0\|\Delta_s(f)\|_2}\mathbb{E}|\Delta_s(f)|^2\mathbb{1}_{\{|\Delta_s(f)|>\lambda\}}
$$

which, by the definition of λ tends to 0 as n tends to infinity. Hence, for every fixed s_0,

$$
\lim_{n\to\infty} \sum_{s=1}^{s_0} \sqrt{n}\mathbb{E}|\Delta_s(f)|\mathbb{1}_{\{|\Delta_s(f)|>\lambda\}} = 0.
$$

Therefore, for every s_0,

$$\begin{aligned}
&\lim_{n\to\infty} \sqrt{n} \sup_{f\in F} |\mathbb{E}\Phi_n(f) - \mathbb{E}f| \\
&\quad \le \lim_{n\to\infty} \sqrt{n} \sup_{f\in F} \left(\sum_{s=1}^{s_0} \mathbb{E}|\Delta_s(f)| \mathbb{1}_{\{|\Delta_s(f)|>\lambda\}} + \sum_{s>s_0} \mathbb{E}|\Delta_s(f)| \mathbb{1}_{\{|\Delta_s(f)|>\lambda\}} \right) \\
&\quad \le c_2 \sup_{f\in F} \sum_{s>s_0} 2^{s/2} \|\Delta_s(f)\|_2,
\end{aligned}$$

where the last inequality is evident from (2.4) and the choice of a suitable absolute constant c_2.

2. Again, we shall use the fact that for every fixed f and s, λ depends on n and tends to 0 as n tends to infinity. Clearly, for every fixed s_0,

$$\begin{aligned}
\|f - \Phi_n(f)\|_2 &\le \sum_{s\le s_0} \|\Delta_s(f) \mathbb{1}_{\{|\Delta_s(f)|>\lambda\}}\|_2 + \sum_{s>s_0} \|\Delta_s(f)\|_2 \\
&\le \sum_{s\le s_0} \|\Delta_s(f) \mathbb{1}_{\{|\Delta_s(f)>\lambda\}}\|_2 + c_3\gamma_2(F, L_2) \sum_{s>s_0} 2^{-s/2}
\end{aligned}$$

For an absolute constant c_3. Indeed, this follows from the fact that for every s,

$$2^{s/2}\|\Delta_s(f)\|_2 \le \sum_{s=0}^{\infty} 2^{s/2}\|\Delta_s(f)\|_2 \le c_3\gamma_2(F, L_2),$$

and of course the constant c_3 does not depend on s.
Hence, for every fixed $f \in F$,

$$\limsup_{n\to\infty} \|f - \Phi_n(f)\|_2 \le c_3\gamma_2(F, L_2) \sum_{s>s_0} 2^{-s/2}$$

for every s_0, and this last quantity goes to zero as $s_0 \to \infty$.

3. If $f(X)$ is square integrable then for any $b > 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{E} \max_{j\le n} \frac{|f(X_j)|^2}{n} &\le \limsup_{n\to\infty} \mathbb{E}\big(\frac{b^2}{n} + \frac{1}{n} \sum_{j\le n} f^2(X_j) \mathbb{1}_{\{|f(X_j)|>b\}} \big) \\
&= \mathbb{E}f^2(X)\mathbb{1}_{\{|f(X)|>b\}}.
\end{aligned}$$

Since the left hand side does not depend on b and the right hand side converges to zero as b tends to ∞, $\lim_{n\to\infty} \mathbb{E}\max_{j\le n} \frac{|f(X_j)|^2}{n} = 0$. Therefore, to complete the proof it suffices to show that

$$\mathbb{E} \max_{j\le n} \frac{|f(X_j) - \Phi_n(f)(X_j)|^2}{n} \to 0.$$

But, using (2),

$$\begin{aligned}
\mathbb{E} \max_{j\le n} \frac{|f(X_j) - \Phi_n(f)(X_j)|^2}{n} &\le \frac{1}{n} \sum_{j\le n} \mathbb{E}|f(X_j) - \Phi_n(f)(X_j)|^2 \\
&= \mathbb{E}|(f - \Phi_n(f))(X)|^2 \to 0. \qquad \square
\end{aligned}$$

The final ingredient we require is the following result on triangular arrays.

Lemma 2.6. *For each n, let $\{\xi_{n,j}\}_{j=1}^n$ by nonnegative, square integrable, independent random variables for which* $\lim_{n\to\infty}\mathbb{E}\max_{j\le n}\xi_{n,j}^2 = 0$*. Then, for every $\delta > 0$,* $\lim_{n\to\infty}\sum_{j=1}^n \mathbb{E}\xi_{n,j}^2 \mathbb{1}_{\{\xi_{n,j}\ge\delta\}} = 0$.

Proof. Consider the stopping times

$$\tau = \tau_n = \begin{cases} \inf\{k \le n : \xi_{n,k} > \delta\} & \text{if } \max_{r\le n}\xi_{n,r} > \delta \\ \infty & \text{if } \max_{r\le n}\xi_{n,r} \le \delta. \end{cases}$$

Then, (see [12]) for every n

$$\begin{aligned} \mathbb{E}\max_{j\le n}\xi_{n,j}^2 &\ge \mathbb{E}\xi_{n,\tau_n}^2\mathbb{1}_{\{\tau_n<\infty\}} = \sum_{l=1}^n \mathbb{E}\xi_{n,l}^2\mathbb{1}_{\{\xi_{n,l}>\delta, \max_{i\le l-1}\xi_{n,i}\le\delta\}} \\ &= \sum_{l=1}^n \mathbb{E}\xi_{n,l}^2\mathbb{1}_{\{\xi_{n,l}>\delta\}}\Pr(\max_{i\le l-1}\xi_{n,i}\le\delta) \\ &\ge \sum_{l=1}^n \mathbb{E}\xi_{n,l}^2\mathbb{1}_{\{\xi_{n,l}>\delta\}}\Pr(\max_{i\le n}\xi_{n,i}\le\delta) \end{aligned}$$

The result now follows, since the hypothesis implies that this last probability converges to one as n tends to infinity. □

We now can conclude

Theorem 2.7. *If the Gaussian process indexed by F is continuous then* $\{\sqrt{n}(P_n(\Phi_n(f)) - Pf) : f \in F\}$ *converges to* $\{G_f : f \in F\}$.

Proof. By Theorem 1.2 and Proposition 2.4 we only need to show that

(i) the finite dimensional distributions of Q_n converge to those of G and
(ii) (F, ρ_2) is totally bounded.

For (i) we need to check that for any $\{f_i\}_{i=1}^k \subseteq F$,

$$(Q_n(\Phi_n(f_1)), \ldots, Q_n(\Phi_n(f_k)))$$

converges in distribution to $(G_{f_1}, \ldots, G_{f_k})$. To see this we apply the Cramer-Wold device, that is, by noting that to show the convergence in distribution, we only have to check that the characteristic function (on $\mathbb{R}^k$) converges, and hence it suffices to show that any finite linear combination of $\{Q_n(\Phi_n(f_i))\}_{i=1}^k$, say, $\sum_{i=1}^k a_i Q_n(\Phi_n(f_i))$ converges in distribution to $\sum_{i=1}^k a_i G_{f_i}$. To verify this, recall the classical Central Limit Theorem for triangular arrays (see, e.g, [5] or [1] Theorem 3.5). Namely, it suffices to prove that

(a) for any $\eta > 0$, $\lim_{n\to\infty}\Pr(\max_{j\le n}|\sum_{i=1}^k a_i\Phi_n(f_i(X_j))| > \eta) = 0$ and
(b) $\lim_{n\to\infty}\mathrm{Var}((\sum_{i=1}^k a_i\Phi_n(f_i))\mathbb{1}_{\{|\sum_{i=1}^k a_i\Phi_n(f_i)|>\eta\}}) = \mathrm{Var}(\sum_{i=1}^k a_i f_i)$.

(a) follows from Proposition 2.5(iii) and (ii) and (b) follows from 2.5(ii) and Lemma 2.6.
(ii) follows from the assumed continuity of $\{G_f : f \in F\}$ with respect to ρ_2 (see p. 41 [13]). □

3. Changing the level of truncation

The question we wish to tackle here is whether it is possible to find different "universal" truncation levels instead of $\sqrt{n}$, and still have a process Ψ which is tight, and satisfies that $n^{-1}\sum_{i=1}^{n}(\psi_n(f))(X_i)$ uniformly approximates $\mathbb{E}f$ (that is, one can replace $\mathbb{E}\Psi_n(f)$ with $\mathbb{E}f$). We show that such a uniform level of truncation has to be asymptotically larger than $\sqrt{n}$.

Definition 3.1. Given a class of functions F and a non-decreasing sequence of positive numbers, $b=\{b_n\}_{n=1}^{\infty}$, let

$$\Phi_{n,b}=\sum_{s=1}^{\infty}\Delta_s(f)\mathbb{1}_{\{|\Delta_s(f)|\le b_n\|\Delta_s(f)\|_2/2^{s/2}\}}.$$

Definition 3.2. A sequence of processes $\{U_n(f):f\in F\}$ is said to be stochastically bounded if for every $\epsilon>0$ there is a constant $C<\infty$ such that

$$\Pr(\sup_{f\in\mathcal{F}}|U_n(f)|>C)<\epsilon.$$

Theorem 3.3. *Assume that $\{b_n\}_n$ is an increasing sequence of positive numbers and that the probability space, $(\Omega,\mathcal{S},P)$, is continuous. Assume also that for every pregaussian class of functions on $(\Omega,\mathcal{S},P)$, the process*

$$\{\sup_{f\in F}\sqrt{n}|P_n(\Phi_{n,b}(f))-\mathbb{E}f|\}_n$$

is stochastically bounded. Then, there exists $\delta>0$ such that $\inf_n\dfrac{b_n}{\sqrt{n}}>\delta$.

Proof. Clearly, if $\{\sqrt{n}\,(P_n'(\Phi_{n,b}(f))-\mathbb{E}f):f\in F\}$ is based on an independent copy, $\{X_j'\}$, then $\{\sup_{f\in F}\sqrt{n}|P_n'(\Phi_{n,b}(f))-\mathbb{E}f|\}_{n=1}^{\infty}$ is also stochastically bounded. Hence, the difference is stochastically bounded, and thus,

$$\{\sup_{f\in F}\sqrt{n}|P_n(\Phi_{n,b}(f))-\mathbb{E}\Phi_{n,b}(f)|\}_n$$

is stochastically bounded, implying that $\sqrt{n}\sup_{f\in F}|\mathbb{E}f-\mathbb{E}\Phi_{n,b}(f)|$ is bounded.

In particular, for every nonnegative $f\in L_2(P)$, if we let $F=\{f,0\}$, then the sequence $\{\sqrt{n}|\mathbb{E}f-\mathbb{E}\Phi_{n,b}(f)|\}_{n=1}^{\infty}$ is bounded. Note that in this case we may assume that $\pi_s(f)=f$ for $s\ge 1$ and $\pi_0(f)=0$, implying that $\sqrt{n}\mathbb{E}f\mathbb{1}_{\{f>b_n\|f\|_2/\sqrt{2}\}}$ is bounded.

Observe that this implies that $\sqrt{n}\mathbb{E}f\mathbb{1}_{\{f>b_n\}}$ is bounded. Indeed, choose b_{k_0} such that $\|f\mathbb{1}_{\{f>b_{k_0}\}}\|_2\le\sqrt{2}$. Applying the above to the function $h=f\mathbb{1}_{\{f>b_{k_0}\}}$, it follows that $\|h\|_2\le\sqrt{2}$ and

$$\sqrt{n}\mathbb{E}h\mathbb{1}_{\{h>b_n\}}=\sqrt{n}\mathbb{E}f\mathbb{1}_{\{f>b_{k_0}\}}\mathbb{1}_{\{f\mathbb{1}_{\{f>b_{k_0}\}}>b_n\}}=\sqrt{n}\mathbb{E}f\mathbb{1}_{\{f>b_{\max(k_0,n)}\}}.$$

Hence, $\sqrt{n}\mathbb{E}f\mathbb{1}_{\{f>b_n\}}$ is bounded, as claimed.

For every sequence $\{a_k\}_k$ for which $\sum_k|a_k|/k<\infty$, consider a function f with $\Pr(f=b_k)=b_1^2\frac{|a_k|}{kb_k^2}$. Such a function exists by the continuity of the probability space $(\Omega,\mathcal{S},P)$. Then,

$$\mathbb{E}f^2=\sum_k b_k^2b_1^2\frac{|a_k|}{kb_k^2}<\infty.$$

Therefore, $\mathbb{E}f\mathbb{1}_{\{f>b_k\}} = \sum_{l>k} b_l b_1^2 \frac{|a_l|}{lb_l^2}$, implying that for every sequence $\{a_k\}_{k=1}^{\infty}$ as above, $\sup_k \sqrt{k}\sum_{l>k}\frac{|a_l|}{l} < \infty$.

Consider the Banach spaces B_1 and B_2, endowed with the norms $\|\{a_k\}\|_1 := \sum_{k=1}^{\infty}\frac{|a_k|}{k}$ and $\|\{a_k\}\|_2 := \sup_{k\geq 1}\sqrt{k}\sum_{l>k}\frac{|a_l|}{lb_l}$. Note that the identity map $\mathbb{I} : B_1 \longrightarrow B_2$ is bounded using the Closed Graph Theorem. Indeed, for $A_n := \{a_{n,k}\}_{k=1}^{\infty}, B := \{b_k\}_{k=1}^{\infty}$ and $C := \{c_k\}_{k=1}^{\infty}$ assume that $\|A_n - B\|_1 \to 0$ and $\|A_n - C\|_2 \to 0$. These conditions respectively imply convergence coordinate-wise, that is, for every r, $\lim_{n\to\infty} a_{n,r} = b_r$ and $\lim_{n\to\infty} a_{n,r} = c_r$. Thus, $B = C$, and the graph is closed, implying that the map is bounded.

Therefore, there exists a constant, C, such that

$$\sup_{k\geq 1}\sqrt{k}\sum_{l>k}\frac{|a_l|}{lb_l} \leq C\sum_{k=1}^{\infty}\frac{|a_k|}{k}. \tag{3.1}$$

Applying (3.1) to the sequence for which the n^{th} term is one and others zero shows that for $n > 1$:

$$\frac{\sqrt{n-1}}{nb_n} \leq C\frac{1}{n},$$

from which the claim follows. □

References

[1] Araujo, A. and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-Chichester-Brisbane. MR576407 (83e:60003)

[2] Bennet, G. (1962). Probability inequalities for sums of independent random variables. *JASA* **57** 33–45.

[3] Fernique, X. (1975). *Regularité des trajectoires des fonctions aléatoires gaussiennes.* École d'Été de Probabilités de Saint-Flour, IV-1974. *Lectures Notes in Math.*, Vol. 480. Springer, Berlin, pp. 1–96. MR0413238 (54 #1355)

[4] Giné, E. and Zinn, J. (1986). Lectures on the central limit theorem for empirical processes. Probability and Banach spaces (Zaragoza, 1985). *Lectures Notes in Math.*, Vol. 1221. Springer, Berlin, pp. 50–113. MR88i:60063

[5] Gnedenko, B. V. and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables.* Translated from the Russian, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition, Addison-Wesley Publishing Col., Reading, MA–London–Don Mills., Ont. MR0233400 (38 #1722)

[6] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3) Isoperimetry and processes], Vol. 23. Springer-Verlag, Berlin. MR1102015 (93c:60001)

[7] Radulović, D. and Wegkamp, M. (2000). Weak convergence of smoothed empirical processes: beyond Donsker classes. *High Dimensional Probability*, II (Seattle, WA, 1999), *Progr. Probab.*, Vol. 47. Birkhäuser Boston, Boston, MA, pp. 89–105. MR1857317 (2002h:60043)

[8] Radulović, D. and Wegkamp, M. (2003). Necessary and sufficient conditions for weak convergence of smoothed empirical processes. *Statist. Probab. Lett.* **61** (3) 321–336. MR2003i:60041

[9] Talagrand, M. (1987). Donsker classes and random geometry. *Ann. Probab.* **15** (4) 1327–1338. MR89b:60090

[10] Talagrand, M. (1987). Regularity of Gaussian processes. *Acta Math.* **159** (1–2) 99–149. MR906527 (89b:60106)

[11] Talagrand, M. (2005). *The Generic Chaining. Upper and Lower Bounds of Stochastic Processes.* Springer Monographs in Mathematics. Springer-Verlag, Berlin. MR2133757

[12] Vakhania, N. N., Tarieladze, V. I. and Chobanyan, S. A. (1987). *Probability Distributions on Banach spaces.* Mathematics and Its Applications (Soviet Series), Vol. 14. D. Reidel Publishing Co., Dordrecht. Translated from the Russian and with a preface by Wojbor A. Woyczynski. MR1435288 (97k:60007)

[13] van der Vaart, Aad W. and Wellner, Jon A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer Series in Statistics. Springer-Verlag, New York. MR1385671 (97g:60035)

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 185–195

DOI: 10.1214/074921706000000842

Empirical and Gaussian processes on Besov classes

Richard Nickl

Department of Statistics, University of Vienna

Abstract: We give several conditions for pregaussianity of norm balls of Besov spaces defined over $\mathbb{R}^d$ by exploiting results in Haroske and Triebel (2005). Furthermore, complementing sufficient conditions in Nickl and Pötscher (2005), we give necessary conditions on the parameters of the Besov space to obtain the Donsker property of such balls. For certain parameter combinations Besov balls are shown to be pregaussian but not Donsker.

1. Introduction

Bounds for the size (measured, e.g., by metric entropy) of a subset $\mathcal{F}$ of the space $\mathcal{L}^2(\mathbb{P})$ of functions square-integrable w.r.t. some probability measure $\mathbb{P}$ allow one to derive limit theorems for the empirical process (indexed by $\mathcal{F}$) as well as continuity properties of the (limiting) Gaussian process (indexed by $\mathcal{F}$). These bounds are often derived from smoothness conditions on the functions contained in $\mathcal{F}$. Function classes that satisfy differentiability or Hölder conditions were among the first examples for pregaussian and Donsker classes, cf. Strassen and Dudley [14], Giné [7], Stute [15], Marcus [11], Giné and Zinn [8], Arcones [1] and van der Vaart [18]. In recent years, interest in spaces of functions with 'generalized smoothness', e.g., spaces of Besov- and Triebel- type, has grown. These spaces contain the spaces defined by more classical smoothness conditions (such as Hölder(-Zygmund), Lipschitz and Sobolev spaces) as special cases and serve as a unified theoretical framework. Besov and Triebel spaces play an increasing role in nonparametric statistics, information theory and data compression, see, e.g., Donoho and Johnstone [3], Donoho, Vetterli, DeVore and Daubechies [4] and Birgé and Massart [2]. Relatively little was known until recently about empirical and Gaussian processes on such function classes, in particular with focus on spaces defined over the whole Euclidean space $\mathbb{R}^d$. Building on Haroske and Triebel [10], sufficient conditions for the parameters of the Besov space were given in [12] implying that the corresponding norm balls are Donsker classes. In the present paper, we extend and complement these results. We give necessary and sufficient conditions for the pregaussian/Donsker property of balls in Besov spaces. In certain 'critical' cases, Besov balls are shown to be pregaussian but *not* Donsker.

2. Besov spaces

For h a real-valued Borel-measurable function defined on $\mathbb{R}^d$ ($d \in \mathbb{N}$) and μ a (nonnegative) Borel measure on $\mathbb{R}^d$, we set $\mu f := \int_{\mathbb{R}^d} f d\mu$ as well as $\|h\|_{r,\mu} := (\int_{\mathbb{R}^d} |h|^r d\mu)^{1/r}$ for $1 \leq r \leq \infty$ (where $\|h\|_{\infty,\mu}$ denotes the μ-essential supremum

AMS 2000 subject classifications: primary 60F17; secondary 46E35.
Keywords and phrases: Besov space, Donsker class, pregaussian class.

of $|h|$). As usual, we denote by $\mathcal{L}^r(\mathbb{R}^d, \mu)$ the vector space of all Borel-measurable functions $h : \mathbb{R}^d \to \mathbb{R}$ that satisfy $\|h\|_{r,\mu} < \infty$. In accordance, $L^r(\mathbb{R}^d, \mu)$ denotes the corresponding Banach spaces of equivalence classes $[h]_\mu$, $h \in \mathcal{L}^r(\mathbb{R}^d, \mu)$, modulo equality μ-a.e. The symbol λ will be used to denote Lebesgue-measure on $\mathbb{R}^d$.

We follow Edmunds and Triebel ([6], 2.2.1) in defining Besov spaces: Let φ_0 be a complex-valued C^∞-function on $\mathbb{R}^d$ with $\varphi_0(x) = 1$ if $\|x\| \leq 1$ and $\varphi_0(x) = 0$ if $\|x\| \geq 3/2$. Define $\varphi_1(x) = \varphi_0(x/2) - \varphi_0(x)$ and $\varphi_k(x) = \varphi_1(2^{-k+1}x)$ for $k \in \mathbb{N}$. Then the functions φ_k form a dyadic resolution of unity. Let $\mathcal{S}(\mathbb{R}^d)$ denote the Schwartz space of rapidly decreasing infinitely differentiable complex-valued functions and let $\mathcal{S}'(\mathbb{R}^d)$ denote the (dual) space of complex tempered distributions on $\mathbb{R}^d$.. In this paper we shall restrict attention to real-valued tempered distributions T (i.e., $T = \bar{T}$, where $\bar{T}$ is defined via $\bar{T}(\phi) = \overline{T(\bar{\phi})}$ for $\phi \in \mathcal{S}(\mathbb{R}^d)$). Let F denote the Fourier transform acting on $\mathcal{S}'(\mathbb{R}^d)$ (see, e.g., Chapter 7.6 in [13]). Then $F^{-1}(\varphi_k FT)$ is an entire analytic function on $\mathbb{R}^d$ for any $T \in \mathcal{S}'(\mathbb{R}^d)$ and any k by the Paley-Wiener-Schwartz theorem (see, e.g., p. 272 in [13]).

Definition 1 (Besov spaces). Let $-\infty < s < \infty$, $1 \leq p \leq \infty$, and $1 \leq q \leq \infty$. For $T \in \mathcal{S}'(\mathbb{R}^d)$ define

$$\|T\|_{s,p,q,\lambda} := \left(\sum_{k=0}^{\infty} 2^{ksq} \left\| F^{-1}(\varphi_k FT) \right\|_{p,\lambda}^q \right)^{1/q}$$

with the modification in case $q = \infty$

$$\|T\|_{s,p,\infty,\lambda} := \sup_{0 \leq k < \infty} 2^{ks} \left\| F^{-1}(\varphi_k FT) \right\|_{p,\lambda}.$$

Define further (real) Besov spaces as

$$B^s_{pq}(\mathbb{R}^d) := \{T \in \mathcal{S}'(\mathbb{R}^d) : T = \bar{T}, \;\; \|T\|_{s,p,q,\lambda} < \infty\}.$$

$B^s_{pq}(\mathbb{R}^d)$ is a Banach space and the norm is independent of the choice of φ_0, and, in particular, different φ_0 result in equivalent norms, cf. Edmunds and Triebel [6], 2.2.1.

Remark 2. (i) The focus in the present paper will be on $s > 0$, in which case it follows (e.g., from 2.3.2 in [17]) that $B^s_{pq}(\mathbb{R}^d)$ consists of (equivalence classes of) p-fold integrable functions. In fact, for these parameters, we could alternatively have defined the spaces $B^s_{pq}(\mathbb{R}^d)$ as $\{[f]_\lambda \in L^p(\mathbb{R}^d, \lambda), \, \|f\|_{s,p,q,\lambda} < \infty\}$.

(ii) We note that $\|T\|_{s,p,q,\lambda} < \infty$ if and only if $\left\|\bar{T}\right\|_{s,p,q,\lambda} < \infty$ for any $T \in \mathcal{S}'(\mathbb{R}^d)$. In fact,

$$\|T\|_{s,p,q,\lambda} \leq \|\mathrm{Re}T\|_{s,p,q,\lambda} + \|\mathrm{Im}T\|_{s,p,q,\lambda} \leq c \, \|T\|_{s,p,q,\lambda}$$

holds for some $1 \leq c < \infty$ and for every $T \in \mathcal{S}'(\mathbb{R}^d)$. As a consequence, one can easily carry over results for complex Besov spaces to real ones and vice versa.

(iii) At least for positive s, there are many equivalent norms on $B^s_{pq}(\mathbb{R}^d)$, some of them possibly more common than the one used in Definition 1; see, e.g., Remark 2 in [12]. In particular, the Hölder-Zygmund Spaces are identical (up to an equivalent norm) to the spaces $B^s_{\infty\infty}(\mathbb{R}^d)$ if $s > 0$.

(iv) Triebel spaces $F^s_{pq}(\mathbb{R}^d)$ are defined in 2.2.1/7 in [6]. We have the chain of continuous imbeddings $B^s_{pu}(\mathbb{R}^d) \hookrightarrow F^s_{pq}(\mathbb{R}^d) \hookrightarrow B^s_{pv}(\mathbb{R}^d)$ for $0 < u \leq \min(p, q)$ and $\max(p, q) \leq v \leq \infty$. By using these imbeddings, the results of the present paper can also be applied to Triebel spaces. Note that $F^0_{p2}(\mathbb{R}^d) = L^p(\mathbb{R}^d, \lambda)$ holds, and for positive s, we have that $F^s_{p2}(\mathbb{R}^d)$ is equal to the classical Sobolev spaces. See 2.2.2 in [6] for further details.

Let $\mathsf{C}(\mathbb{R}^d)$ be the vector space of bounded continuous real-valued functions on $\mathbb{R}^d$ normed by the sup-norm $\|\cdot\|_\infty$. If either $s > d/p$ or $s = d/p$ and $q = 1$, it is well-known (see, e.g., Proposition 3 in [12]) that each equivalence class $[f]_\lambda \in B^s_{pq}(\mathbb{R}^d)$, contains a (unique) continuous representative. [In fact, the Banach space $B^s_{pq}(\mathbb{R}^d)$ is imbedded (up to a section map) into the space $\mathsf{C}(\mathbb{R}^d)$.] Hence, if either $s > d/p$ or $s = d/p$ and $q = 1$, we can define the (closely related) Banach space

$$\mathsf{B}^s_{pq}(\mathbb{R}^d) = \{f \in \mathsf{C}(\mathbb{R}^d) : [f]_\lambda \in L^p(\mathbb{R}^d, \lambda), \|f\|_{s,p,q,\lambda} < \infty\}$$

(again normed by $\|\cdot\|_{s,p,q,\lambda}$) by collecting the continuous representatives.

Throughout the paper we shall use the following notational agreements: We define the function $\langle x\rangle^\gamma = (1 + \|x\|^2)^{\gamma/2}$ parameterized by $\gamma \in \mathbb{R}$, where x is an element of $\mathbb{R}^d$ and where $\|\cdot\|$ denotes the Euclidean norm. Also, for two real-valued functions $a(\cdot)$ and $b(\cdot)$, we write $a(\varepsilon) \lesssim b(\varepsilon)$ if there exists a positive (finite) constant c not depending on ε such that $a(\varepsilon) \le cb(\varepsilon)$ holds for all $\varepsilon > 0$. If $a(\varepsilon) \lesssim b(\varepsilon)$ and $b(\varepsilon) \lesssim a(\varepsilon)$ both hold we write $a(\varepsilon) \sim b(\varepsilon)$. [In abuse of notation, we shall also use this notation for sequences a_k and b_k, $k \in \mathbb{N}$ as well as for two (semi)norms $\|\cdot\|_{X,1}$ and $\|\cdot\|_{X,2}$ on a vector space X.]

3. Main results

Let $(S, \mathcal{A}, \mu)$ be some probability space and let $\mathbb{P}$ be a (Borel) probability measure on $\mathbb{R}^d$. Let $\emptyset \ne \mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$. A Gaussian process $\mathbb{G} : (S, \mathcal{A}, \mu) \times \mathcal{F} \to \mathbb{R}$ with mean zero and covariance $E\mathbb{G}(f)\mathbb{G}(g) = \mathbb{P}[(f - \mathbb{P}f)(g - \mathbb{P}g)]$ for $f, g \in \mathcal{F}$ is called a (generalized) *Brownian bridge* process on $\mathcal{F}$. The covariance induces a semimetric $\rho^2(f, g) = E[\mathbb{G}(f) - \mathbb{G}(g)]^2$ for $f, g \in \mathcal{F}$. A function class $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$ will be called $\mathbb{P}$-*pregaussian* if such a Gaussian process $\mathbb{G}$ can be defined such that for every $s \in S$, the map $f \longmapsto \mathbb{G}(f, s)$ is bounded and uniformly continuous w.r.t. the semimetric ρ from $\mathcal{F}$ into $\mathbb{R}$. For further details see p.92-93 in [5].

Let $\mathbb{P}_n = 1/n \sum_{i=1}^n \delta_{X_i}$ denote the empirical measure of n independent $\mathbb{R}^d$-valued random variables $X_1, \ldots, X_n$ identically distributed according to some law $\mathbb{P}$. [We assume here the standard (canonical) model as on p.91 in [5].] For $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$, the $\mathcal{F}$-indexed *empirical process* ν_n is given by $f \longmapsto \nu_n(f) = \sqrt{n}\,(\mathbb{P}_n - \mathbb{P})\,f$. The class $\mathcal{F}$ is said to be $\mathbb{P}$-*Donsker* if it is $\mathbb{P}$-pregaussian and if ν_n converges in law in the space $\ell^\infty(\mathcal{F})$ to a (generalized) Brownian bridge process over $\mathcal{F}$, cf. p.94 in [5]. Here $\ell^\infty(\mathcal{F})$ denotes the Banach space of all bounded real-valued functions on $\mathcal{F}$. If $\mathcal{F}$ is $\mathbb{P}$-Donsker for all probability measures $\mathbb{P}$ on $\mathbb{R}^d$, it is called *universally* Donsker.

In [12], Corollary 5, Proposition 1 and Theorem 2, the following results were proved. [Clearly, one may replace $\mathcal{U}$ by and bounded subset of $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ in the proposition.]

Proposition 3. *Let $\mathcal{U}$ be the closed unit ball of $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ where $1 \le p \le \infty$, $1 \le q \le \infty$. Let $\mathbb{P}$ be a probability measure on $\mathbb{R}^d$.*

1. *Let $1 \le p \le 2$ and $s > d/p$. Then $\mathcal{U}$ is $\mathbb{P}$-Donsker, and hence also $\mathbb{P}$-pregaussian.*

2. *Let* $2 < p \leq \infty$ *and* $s > d/2$. *Assume that* $\int_{\mathbb{R}^d} \|x\|^{2\gamma} d\mathbb{P} < \infty$ *holds for some* $\gamma > d/2 - d/p$. *Then* $\mathcal{U}$ *is* $\mathbb{P}$*-Donsker, and hence also* $\mathbb{P}$*-pregaussian.*
3. *Let* $d = q = 1$, $1 \leq p < 2$ *and* $s = 1/p$. *Then* $\mathcal{U}$ *is* $\mathbb{P}$*-Donsker, and hence also* $\mathbb{P}$*-pregaussian.*

In the present paper we show on the one hand that, if one is interested in the pregaussian property only, the conditions of Proposition 3 can be substantially weakened. On the other hand, we show that Proposition 3 is (essentially) best possible w.r.t. the Donsker property: It turns out that $s \geq \max(d/p, d/2)$ always has to be satisfied for $\mathcal{U}$ to be $\mathbb{P}$-Donsker and that the moment condition in Part 2 of Proposition 3 cannot be improved upon. We also give a rather definite picture of the limiting case $s = d/p$ (where only the cases $q = 1$ and $d > 1$, as well as $p = 2$ and $q = 1$, will remain undecided).

3.1. The pregaussian property

We first discuss the pregaussian property in the 'nice' case $s > \max(d/p, d/2)$: If $s > d/p$ and $p \leq 2$, Proposition 3 implies that the unit ball of the Besov space is pregaussian for every probability measure. On the other hand, maybe not surprisingly, if the integrability parameter p of the Besov space is larger than 2, Proposition 1 requires an additional moment condition on the probability measure to obtain the pregaussian property. The following theorem shows that this additional moment condition is also necessary (for most probability measures possessing Lebesgue-densities). [Note that $s > d/2$ ensures also that $s > d/p$ holds, so the condition $s > \max(d/p, d/2)$ is always satisfied.]

Theorem 4. *Let* $\mathcal{U}$ *be the closed unit ball of* $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ *with* $2 < p \leq \infty$, $1 \leq q \leq \infty$ *and* $s > d/2$. *Let* δ *be arbitrary subject to* $0 < \delta \leq d/2 - d/p$. *Define the probability measure* $\mathbb{P}$ *by* $d\mathbb{P}(x) = \varphi(x) \langle x \rangle^{-d-2\delta} d\lambda(x)$ *where* $0 < c \leq \varphi(x)$ *holds for some constant* c *and all* $x \in \mathbb{R}^d$ *(and where* $\|\varphi \langle x \rangle^{-d-\delta}\|_{1,\lambda} = 1$*). Then the set* $\mathcal{U}$ *is not* $\mathbb{P}$*-pregaussian.*

Proof. Note first that $\mathcal{U}$ is a bounded subset of $\mathsf{C}(\mathbb{R}^d)$ (see, e.g., Proposition 3 in [12]) and hence also of $\mathcal{L}^r(\mathbb{R}^d, \mathbb{P})$ for every $1 \leq r \leq \infty$. Observe that

$$\|f - g\|^2_{2,\mathbb{P}} = \int_{\mathbb{R}^d} [f - g]^2 d\mathbb{P} = \int_{\mathbb{R}^d} [(f - g) \langle x \rangle^{-(d-2\delta)/2}]^2 \varphi d\lambda$$
$$\geq c \left\| (f - g) \langle x \rangle^{-(d-2\delta)/2} \right\|^2_{2,\lambda}$$

holds for $f, g \in \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$. Hence we have for the metric entropy (see Definition 9 in the Appendix) that

$$H(\varepsilon, U, \|\cdot\|_{2,\mathbb{P}}) \geq H(\varepsilon/c, U, \left\| (\cdot) \langle x \rangle^{-(d-2\delta)/2} \right\|_{2,\lambda})$$

holds. We obtain a lower bound of order $\varepsilon^{-\alpha}$ for the r.h.s. of the above display from Corollary 12 in the Appendix upon setting $\gamma = (d - 2\delta)/2$ in that corollary. Since $s - d/p > d/2 - d/p > 0$ and $\delta \leq d/2 - d/p$, it follows that $\gamma < s - d/p + d/2$ and we obtain $\alpha = (\delta/d + 1/p)^{-1}$. Clearly $\alpha > 2$ holds since $\delta \leq d/2 - d/p$. Define the Gaussian process $\mathbb{L}(f) = \mathbb{G}(f) + Z \cdot \mathbb{P}f$ for $f \in \mathcal{U}$ where Z is a standard normal variable independent of $\mathbb{G}$. It is easily seen that this process has covariance $E\mathbb{L}(f)\mathbb{L}(g) = \mathbb{P}fg$. Since $a > 2$ and since $\mathbb{P}$ possesses a Lebesgue-density, we can apply the Sudakov-Chevet minoration (Theorem 2.3.5 in [5]) which implies that the process $\mathbb{L}$ is μ-a.s. unbounded on $\mathcal{U}$. Since $\sup_{f\in\mathcal{U}}|\mathbb{P}f| < \infty$ holds, we have that $\sup_{f\in\mathcal{U}}|\mathbb{L}(f)| = \infty$ μ-a.s. implies $\sup_{f\in\mathcal{U}}|\mathbb{G}(f)| = \infty$ μ-a.s. This proves that $\mathcal{U}$ is not $\mathbb{P}$-pregaussian. □

The set $\mathcal{U}$ is uniformly bounded (in fact, for $p < \infty$, any $f \in \mathcal{U}$ satisfies $\lim_{\|x\|\to\infty} f(x) = 0$, see Proposition 3 in [12]), but nevertheless one needs a moment condition on the probability measure to obtain the pregaussian property. [The reason is, not surprisingly, that the degree of compactness in $L^2(\mathbb{R}^d, \mathbb{P})$ measured in terms of metric entropy is driven *both* by smoothness of the function class *and* by its rate of decay at infinity.]

In the remainder of this section we shed light on the critical cases $s \le d/p$ and/or $s \le d/2$. The following proposition shows that in case $s \le d/p$ but $s > d/2$ (and hence $1 \le p < 2$), Besov balls are again pregaussian for a large class of probability measures:

Theorem 5. *Let U be the closed unit ball of $B^s_{pq}(\mathbb{R}^d)$ with $1 \le p < 2$, $1 \le q \le \infty$ and $s > d/2$, and let $\mathcal{U}$ be any set constructed by selection of one arbitrary representative out of every $[f]_\lambda \in U$. Let $\mathbb{P}$ be a probability measure on $\mathbb{R}^d$ that possesses a density φ w.r.t. Lebesgue measure on $\mathbb{R}^d$ such that $\|\varphi \langle x\rangle^d\|_{\infty,\lambda} < \infty$. Then $\mathcal{U}$ is $\mathbb{P}$-pregaussian.*

Proof. Note first that U is a bounded subset of $L^2(\mathbb{R}^d, \lambda)$ (by Proposition 11 and (4) in the Appendix) and hence also of $L^2(\mathbb{R}^d, \mathbb{P})$ since $[\varphi]_\lambda \in L^\infty(\mathbb{R}^d, \lambda)$. Observe next that

$$\begin{aligned}\|f-g\|^2_{2,\mathbb{P}} &= \int_{\mathbb{R}^d} [f-g]^2 \varphi d\lambda = \int_{\mathbb{R}^d} [(f-g)\langle x\rangle^{-d/2}]^2 \varphi \langle x\rangle^d d\lambda \\ &\le \left\|(f-g)\langle x\rangle^{-d/2}\right\|^2_{2,\lambda} \left\|\varphi\langle x\rangle^d\right\|_{\infty,\lambda}\end{aligned}$$

holds for $f, g \in \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$ by Hölder's inequality. Hence we apply Corollary 12 in the Appendix with $\gamma = d/2$ to obtain

$$H(\varepsilon, \mathcal{U}\ \|\cdot\|_{2,\mathbb{P}}) \le H(\varepsilon \left\|\varphi\langle x\rangle^d\right\|_{\infty,\lambda}, U, \left\|(\cdot)\langle x\rangle^{-d/2}\right\|_{2,\lambda}) \lesssim \varepsilon^{-\alpha}$$

where $\alpha = d/s$ if $s - d/p < 0$ and $\alpha = p$ if $s - d/p > 0$ and where we have used that $\mathbb{P}$ is absolutely continuous w.r.t. Lebesgue measure λ. In both cases we have $\alpha < 2$. Hence we can apply Theorem 2.6.1 in [5] to obtain (a.s.) sample-boundedness and -continuity of the process $\mathbb{L}$ (defined in the proof of Theorem 4 above) on $\mathcal{U}$ w.r.t. the $\mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$-seminorm. If $\pi(f) = f - \mathbb{P}f$, then $\mathbb{L}(\pi(f)) = \mathbb{G}(f)$ is also (a.s.) sample-bounded and -continuous on $\mathcal{U}$ and hence we obtain the $\mathbb{P}$-pregaussian property for $\mathcal{U}$ by the same reasoning as on p.93 in [5]. If $s = d/p$, view U as a bounded subset of $B^{d/p-\varepsilon}_{pq}(\mathbb{R}^d)$ where ε can be chosen small enough such that $d/p - \varepsilon > d/2$ holds (note that $d/p > d/2$ since $p < 2$) and hence the pregaussian property follows from the case $s - d/p < 0$ just established. This finishes the proof. □

Note that any probability measure $\mathbb{P}$ that possesses a bounded density which is eventually monotone, or a bounded density with polynomial or exponential tails, satisfies the condition of the theorem. [We note that at least for the special case $d = q = 1$, $s = 1/p$, $p < 2$, the condition on $\mathbb{P}$ can be removed by Proposition 3.]

The following theorem deals with the remaining cases and shows that $s \ge d/2$ always has to be satisfied (irrespective of p) to obtain the pregaussian property.

Theorem 6. *Let U be the closed unit ball of $B^s_{pq}(\mathbb{R}^d)$ with $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $0 < s < d/2$, and let $\mathcal{U}$ be any set constructed by selection of one arbitrary representative out of every $[f]_\lambda \in U$. Let $\mathbb{P}$ be a probability measure that possesses a bounded density φ w.r.t. Lebesgue measure on $\mathbb{R}^d$ which satisfies $0 < c \leq \varphi(x)$ for some constant c and all x in some open subset V of $\mathbb{R}^d$. Then the set $\mathcal{U}$ is not $\mathbb{P}$-pregaussian.*

Proof. Since V is open, it contains an open Euclidean ball Ω, which is a bounded C^∞-domain in the sense of Triebel [17], 3.2.1. Denote by $\lambda\,|\Omega$ Lebesgue measure on Ω and by $L^2(\Omega,\lambda)$ the usual Banach space normed by the usual L^2-norm $\|\cdot\|_{2,\lambda|\Omega}$ on Ω. Let $U\,|\Omega$ be the set of restrictions $[f\,|\Omega]_{\lambda|\Omega}$ of elements $[f]_\lambda \in U$ to the set Ω. Note that $U\,|\Omega$ is the unit ball of the factor Besov space $B^s_{pq}(\mathbb{R}^d)\,|\Omega$ over Ω obtained by restricting the elements of $B^s_{pq}(\mathbb{R}^d)$ to Ω with the restricted Besov norm

$$\|f\|_{s,p,q,|\Omega} := \inf\left\{\|g\|_{s,p,q,\lambda} : \ [g]_\lambda \in B^s_{pq}(\mathbb{R}^d), \quad [g\,|\Omega]_{\lambda|\Omega} = f\right\}.$$

We first handle the case $p = 1$. In view of 2.5.1/7 and 2.2.2/1 in [6], we have that $B^{d/2}_{1\infty}(\mathbb{R}^d)\,|\Omega \nsubseteq L^2(\Omega,\lambda)$. But by $s < d/2$ and 3.3.1/7 of Triebel (1983) we also have $B^{d/2}_{1\infty}(\mathbb{R}^d)\,|\Omega \subseteq B^s_{1q}(\mathbb{R}^d)\,|\Omega$ hence we conclude that $B^s_{1q}(\mathbb{R}^d)\,|\Omega \nsubseteq L^2(\Omega,\lambda)$. Since $\varphi \geq c$ on Ω, this implies $U \nsubseteq L^2(\mathbb{R}^d,\mathbb{P})$, so $\mathcal{U}$ cannot be $\mathbb{P}$-pregaussian.

We now turn to $p > 1$. We first treat the case $s = d/2-\varepsilon$ where $\varepsilon > 0$ is arbitrary subject to $\varepsilon < d - d/p$. Then U is a bounded subset of $L^2(\mathbb{R}^d,\lambda)$ by Proposition 11 and (4) in the Appendix and hence also of $L^2(\mathbb{R}^d,\mathbb{P})$ since φ is bounded. We now obtain a metric entropy lower bound for U in $L^2(\mathbb{R}^d,\mathbb{P})$. Observe that

$$\|f-g\|^2_{2,\mathbb{P}} = \int_{\mathbb{R}^d} [f-g]^2 d\mathbb{P} \geq c \int_\Omega [f-g]^2 d\lambda\,|\Omega$$

holds for $f, g \in \mathcal{L}^2(\mathbb{R}^d,\mathbb{P})$ and hence

$$H(\varepsilon, U, \|\cdot\|_{2,\mathbb{P}}) \geq H(\varepsilon/c, U\,|\Omega\,, \|\cdot\|_{2,\lambda|\Omega}) \tag{1}$$

holds. By 3.3.3/1 in [6], we obtain the entropy number (see Definition 8 in the Appendix)

$$e\left(k, id(U\,|\Omega), \|\cdot\|_{0,2,\infty,|\Omega}\right) \sim k^{-s/d}.$$

Now by Lemma 1 as well as expression (4) in the Appendix we obtain

$$H(\varepsilon, U\,|\Omega\,, \|\cdot\|_{2,\lambda|\Omega}) \gtrsim \varepsilon^{-d/s}.$$

But since $s < d/2$ holds by assumption, this (together with (1)) implies that $\sup_{f\in\mathcal{U}} |\mathbb{G}(f)| = \infty$ μ-a.s. by the same application of the Sudakov-Chevet minoration as in the proof of Theorem 4 above, noting that $\sup_{f\in\mathcal{U}} |\mathbb{P}f| < \infty$ holds since U is bounded in $L^2(\mathbb{R}^d,\mathbb{P})$. Hence $\mathcal{U}$ is not $\mathbb{P}$-pregaussian in this case. The remaining cases $s-\varepsilon$ with $\varepsilon \geq d - d/p$ now follow from the continuous imbedding $B^s_{pq}(\mathbb{R}^d) \hookrightarrow B^t_{pq}(\mathbb{R}^d)$ for $s > t$, cf. 2.3.2/7 in [17]. □

Observe that $\mathbb{P}$ in the above theorem could be compactly supported, so the pregaussian property cannot be restored by a moment condition. Inspection of the proof shows that a similar negative result can be proved for the unit ball of a Besov space over any subdomain of $\mathbb{R}^d$ (that possesses a suitably regular boundary).

The limiting case $p = 2$, $1 \le q \le \infty$, $s = d/2$ remains open: Here, one would have to go to the logarithmic scale of metric entropy rates, in which case it is known that metric entropy conditions are not sharp in terms of proving the pregaussian property, see p.54 in [5]. At least for $q \ge 2$ we conjecture that the unit ball of $B_{2q}^{d/2}(\mathbb{R}^d)$ is not $\mathbb{P}$-pregaussian for absolutely continuous probability measures possessing a bounded density.

3.2. The Donsker property

In this section we show that Proposition 3 is (essentially) best possible in terms of the Donsker property for norm balls in Besov spaces. We first discuss the 'nice' case $s > \max(d/p, d/2)$. If $p \le 2$, Part of Proposition 3 is certainly best possible (since then Besov balls are universally Donsker). Since $\mathbb{P}$-Donsker classes must be $\mathbb{P}$-pregaussian, the moment condition in Part 2 ($p > 2$) of Proposition 3 is (essentially) necessary in view of Theorem 4 above. For the case $p = q = \infty$, these findings imply known results for Hölder and Lipschitz classes due to Giné and Zinn [8], Arcones [1] and van der Vaart [18]; cf. also the discussion in Remark 5 in [12].

We now turn to the critical cases $s \le d/p$ and/or $s \le d/2$. Since Donsker classes need to be pregaussian, Theorem 6 implies that $s \ge d/2$ *always has to be satisfied* (at least for the class of probability measures defined in that theorem).

On the other hand, for $1 \le p < 2$ we showed in Theorem 5 that norm balls of $B^s_{pq}(\mathbb{R}^d)$ with $d/2 < s \le d/p$ (and hence $1 \le p < 2$) are pregaussian for a large class of probability measures. So the question arises whether these classes are also Donsker classes for these probability measures. In the special case $q = d = 1$ and $s = 1/p$, these classes are in fact universally Donsker in view of Part 3 of Proposition 3. We do not know whether this can be generalized to the case $d > 1$, that is, whether the unit ball of $\mathsf{B}^{d/p}_{p1}(\mathbb{R}^d)$ with $1 \le p < 2$ is a (universal) Donsker class. [The proof in case $d = 1$ uses spaces of functions of bounded p-variation, a concept which is not straightforwardly available for $d > 1$.]

On the other hand, the following theorem shows that the function classes that were shown to be pregaussian in Theorem 5 are in fact *not* $\mathbb{P}$-Donsker for probability measures $\mathbb{P}$ possessing a bounded density if $s < d/p$, or if $s = d/p$ but $q > 1$ hold. The proof strategy partially follows the proof of Theorem 2.3 in [11].

Theorem 7. *Let U be the closed unit ball of $B^s_{pq}(\mathbb{R}^d)$ with $1 \le p \le \infty$, $1 \le q \le \infty$, $s > 0$ and let $\mathcal{U}$ be any set constructed by selection of one arbitrary representative out of every $[f]_\lambda \in U$. Assume that $\mathbb{P}$ possesses a bounded density w.r.t. Lebesgue measure. Suppose that either $s < d/p$ or that $s = d/p$ but $q > 1$ holds. Then $\mathcal{U}$ is not a $\mathbb{P}$-Donsker class.*

Proof. We first consider the case $s = d/p$ but $q > 1$. By Theorem 2.6.2/1 in [16], p. 135, $B^{d/p}_{pq}(\mathbb{R}^d)$ contains a function $\psi \in \mathcal{L}^1(\mathbb{R}^d, \lambda)$ that satisfies

$$|\psi(x)| \ge C \log |\log |x||$$

for $|x| \in (0, \varepsilon]$ and some $0 < \varepsilon < 1$. We may assume w.l.o.g. $\|\psi\|_{s,p,q,\lambda} \le 1$. Since $(F\psi(\cdot - y))(u) = e^{-iyu} F\psi(u)$ holds, inspection of Definition 1 shows that $\|\psi(\cdot - y)\|_{s,p,q,\lambda} = \|\psi\|_{s,p,q,\lambda} \le 1$ for every $y \in \mathbb{R}^d$. Let $(z_i)_{i=1}^\infty$ denote all points in $\mathbb{R}^d$ with rational coordinates and define $\psi_i = \psi(\cdot - z_i)$ which satisfies $\|\psi_i\|_{s,p,q,\lambda} \le 1$ for every i. Consequently, we have $\{\tilde{\psi}_i\}_{i=1}^\infty \subseteq \mathcal{U}$ where $\tilde{\psi}_i$ is obtained by modifying each ψ_i on a set N_i of Lebesgue-measure zero if necessary. Clearly $\cup_{i=1}^\infty N_i$ is again a set of Lebesgue measure zero. Let now $x \in \mathbb{R}^d \setminus \cup_{i=1}^{\infty} N_i$ be arbitrary and let the index set I_x consist of all $i \in \mathbb{N}$ s.t. $|x - z_i| < \varepsilon$ holds. Clearly $(z_i)_{i \in I_x}$ is dense in a neighborhood of x. Consequently

$$\sup_{f \in \mathcal{U}} |f(x)| \geq \sup_{i \in \mathbb{N}} \left|\tilde{\psi}_i(x)\right| = \sup_{i \in \mathbb{N}} |\psi_i(x)| = \sup_{i \in \mathbb{N}} |\psi(x - z_i)|$$
$$\geq \sup_{i \in I_x} C \log |\log |x - z_i|| = \infty$$

holds for every $x \in \mathbb{R}^d \setminus \cup_{i=1}^{\infty} N_i$ and hence Lebesgue almost everywhere. Note furthermore that U is bounded in $L^2(\mathbb{R}^d, \lambda)$ (by Proposition 11 and (4) in the Appendix). Furthermore, $\mathbb{P}$ possesses a density $[\phi]_\lambda \in L^\infty(\mathbb{R}^d, \lambda) \cap L^1(\mathbb{R}^d, \lambda) \subseteq L^2(\mathbb{R}^d, \lambda)$, so we have $\sup_{f \in \mathcal{U}} |\mathbb{P}f| < \infty$ by using the Cauchy-Schwarz inequality. Conclude that

$$M_{\mathbb{P}}(x) = \sup_{f \in \mathcal{U}} |f(x) - \mathbb{P}f| \geq \sup_{f \in \mathcal{U}} |f(x)| - \sup_{f \in \mathcal{U}} |\mathbb{P}f| = \infty$$

holds λ-a.e. Since $\mathbb{P}$ is absolutely continuous, we have that $\mathcal{U}$ is not a $\mathbb{P}$-Donsker class since $t^2 \mathbb{P}(M_{\mathbb{P}} > t) \to 0$ is necessary for the $\mathbb{P}$-Donsker property to hold for $\mathcal{U}$ (see, e.g., Proposition 2.7 in [9]). The remaining cases follow from the continuous imbedding $B^{d/p}_{pu}(\mathbb{R}^d) \hookrightarrow B^s_{pv}(\mathbb{R}^d)$ for $s < d/p$ and $u, v \in [1, \infty]$ (cf. 2.3.2/7 in [17]). $\square$

At least on the sample space $\mathbb{R}^d$ we are not aware of any other ('constructive') examples for pregaussian classes that are not Donsker: The above theorem shows that the empirical process does *not* converge in law in $\ell^\infty(\mathcal{U})$ if $\mathcal{U}$ is the unit ball of $B^s_{pq}(\mathbb{R}^d)$ (with $s < d/p$ or $s = d/p$ but $q > 1$). However, if $p < 2$ and $s > d/2$ a sample-bounded and -continuous Brownian bridge process *can* be defined on $\mathcal{U}$ by Theorem 5 above.

Inspection of the proof shows that a similar negative result can be proved for the unit ball of a Besov space defined over any (non-empty) subset Ω of $\mathbb{R}^d$ (at least if Ω has regular boundary). Note that the above theorem also implies for the case $p = 2$, $s = d/2$ (not covered in Section 3.1) that the unit ball of $B^{d/2}_{2q}(\mathbb{R}^d)$ is not Donsker if $q > 1$. [The special case $q = 1$ remains open.]

Appendix A: Technical results

Definition 8. Let $\mathcal{J}$ be a subset of the normed space $(Y, \|\cdot\|_Y)$, and let $U_Y = \{y \in Y : \|y\|_Y \leq 1\}$ be the closed unit ball in Y. Then, for all natural numbers k, the k-th entropy number of $\mathcal{J}$ is defined as

$$e(k, \mathcal{J}, \|\cdot\|_Y) = \inf \left\{ \varepsilon > 0 : \mathcal{J} \subseteq \bigcup_{j=1}^{2^{k-1}} (y_j + \varepsilon U_Y) \quad \text{for some } y_1, \ldots, y_{2^{k-1}} \in Y \right\},$$

with the convention that the infimum equals $+\infty$ if the set over which it is taken is empty.

Suppose $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are normed spaces such that X is a linear subspace of Y. Let U_X the closed unit ball in X. Then, $e(k, id(U_X), \|\cdot\|_Y)$ is called the k-th entropy number of the operator $id : X \to Y$. Clearly, $e(k, id(U_X), \|\cdot\|_Y)$ is finite for all k if and only if id is continuous from X to Y (in which case we shall write $(X, \|\cdot\|_X) \hookrightarrow (Y, \|\cdot\|_Y)$) and the entropy numbers converge to zero as $k \to \infty$ if and only if the operator id is compact (has totally bounded image in Y.)

Definition 9. For a (non-empty) subset $\mathcal{J}$ of a normed space $(Y, \|\cdot\|_Y)$, denote by $N(\varepsilon, \mathcal{J}, \|\cdot\|_Y)$ the *minimal covering number*, i.e., the minimal number of closed balls of radius ε, $0 < \varepsilon < \infty$, (w.r.t. $\|\cdot\|_Y$) needed to cover $\mathcal{J}$. In accordance, let $H(\varepsilon, \mathcal{J}, \|\cdot\|_Y) = \log N(\varepsilon, \mathcal{J}, \|\cdot\|_Y)$ be the *metric entropy* of the set $\mathcal{J}$, where log denotes the natural logarithm.

The following lemma gives a relationship between metric entropy and entropy numbers:

Lemma 10. *Let $0 < \alpha < \infty$ and let $\mathcal{J}$ be a totally bounded (non-empty) subset of a normed space $(Y, \|\cdot\|_Y)$ satisfying*

$$e(k, \mathcal{J}, \|\cdot\|_Y) \sim k^{-1/\alpha}.$$

We then have for the metric entropy

$$H(\varepsilon, \mathcal{J}, \|\cdot\|_Y) \sim \varepsilon^{-\alpha}.$$

Proof. The inequality $H(\varepsilon) \leq C_1 \varepsilon^{-\alpha}$ is part of the proof of Theorem 1 in [12]. The lower bound follows from an obvious inversion of the argument. □

We next state a special case of more general results due to Haroske and Triebel [10]. Here we use weighted Besov spaces $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^{-\gamma})$ defined in Section 4.2 in [6], see also Definition 2 in [12]. Note that $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^{-0}) = B^s_{pq}(\mathbb{R}^d)$.

Proposition 11 (Haroske and Triebel). *Suppose $p, q_1, q_2 \in [1, \infty]$, $s - d/p + d/2 > 0$. Then $B^s_{pq_1}(\mathbb{R}^d)$ is imbedded into $B^0_{2q_2}(\mathbb{R}^d, \langle x\rangle^{-\gamma})$ for every $\gamma \geq 0$. If $\gamma > 0$, the imbedding is even compact, in which case the entropy numbers of this imbedding satisfy*

$$e\left(k, id(U_{B^s_{pq_1}(\mathbb{R}^d)}), \left\|(\cdot)\langle x\rangle^{-\gamma}\right\|_{0,2,q_2,\lambda}\right) \sim k^{-1/\alpha}$$

for all $k \in \mathbb{N}$ where $\alpha = d/s$ if $\gamma > s - d/p + d/2$ and $\alpha = (\gamma/d + 1/p - 1/2)^{-1}$ if $\gamma < s - d/p + d/2$.

Proof. The first imbedding follows from Theorem 4.2.3 in [6]. The remaining claims of the proposition are proved in Theorem 4.1 in [10] for complex Besov spaces noting that the norms used in that reference are equivalent to the weighted norm $\|(\cdot)\langle x\rangle^{-\gamma}\|_{0,2,q_2,\lambda}$ used here; cf. Theorem 4.2.2 in [6]. The proposition for real Besov spaces follows from Lemma 1 in [12], see also the proof of Proposition 2 in the latter paper. □

Finally, we obtain the following corollary. [Here, and in other proofs of the paper, we use the obvious fact that metric entropy is not increased under Lipschitz-transformations between normed spaces (e.g., linear and continuous mappings); cf. also Lemma 2 in [12]].

Corollary 12. *Suppose $p, q \in [1, \infty]$, $s - d/p + d/2 > 0$ and $\gamma > 0$. We then have that*

$$H(\varepsilon, U_{B^s_{pq}(\mathbb{R}^d)}, \left\|(\cdot)\langle x\rangle^{-\gamma}\right\|_{2,\lambda}) \sim \varepsilon^{-\alpha} \tag{2}$$

where $\alpha = d/s$ if $\gamma > s - d/p + d/2$ and $\alpha = (\gamma/d + 1/p - 1/2)^{-1}$ if $\gamma < s - d/p + d/2$.

Proof. We have

$$H(\varepsilon, U_{B^s_{pq}(\mathbb{R}^d)}, \left\| (\cdot) \langle x \rangle^{-\gamma} \right\|_{0,2,q',\lambda}) \sim \varepsilon^{-\alpha} \tag{3}$$

for every $1 \leq q' \leq \infty$ by Proposition 11 and Lemma 1 above. Since

$$\|f\|_{0,2,\infty,\lambda} \lesssim \|f\|_{2,\lambda} \lesssim \|f\|_{0,2,1,\lambda} \tag{4}$$

holds for all $f \in [f] \in B^0_{21}(\mathbb{R}^d) \supseteq B^s_{pq}(\mathbb{R}^d)$ by 2.5.7/1 in [17], we have (2) by using (3) to construct upper ($q' = 1$) and lower ($q' = \infty$) bounds for $H(\varepsilon, U_{B^s_{pq}(\mathbb{R}^d)}, \|(\cdot) \langle x \rangle^{-\gamma} \|_{2,\lambda})$. □

Acknowledgement

The author wishes to thank Evarist Giné and Benedikt M. Pötscher for very helpful discussions (in particular about Theorem 7) and comments on a preliminary version of the paper.

References

[1] Arcones, M. A. (1994). The central limit theorem for U-processes indexed by Hölder's functions. *Statist. Probab. Lett.* **20** 57–62.

[2] Birgé, L. and Massart, P. (2000). An adaptive compression algorithm in Besov spaces. *Constr. Approx.* **16** 1–36.

[3] Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Stat.* **26** 879–921.

[4] Donoho, D. L., Vetterli, M., DeVore, R. A. and Daubechies, I. (1998). Data compression and harmonic analysis. *IEEE Trans. Inf. Theory* **44** 2435–2476.

[5] Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge, England.

[6] Edmunds, D. E. and Triebel, H. (1996). *Function Spaces, Entropy Numbers and Differential Operators.* Cambridge University Press, Cambridge, England.

[7] Giné, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev-norms. *Ann. Stat.* **3** 1243–1266.

[8] Giné, E. and Zinn, J. (1986a). Empirical processes indexed by Lipschitz functions. *Ann. Prob.* **14** 1329–1338.

[9] Giné, E. and Zinn, J. (1986b). Lectures on the central limit theorem for empirical processes. In *Probability and Banach Spaces. Lecture Notes in Mathematics* **1221** pp. 50–113.

[10] Haroske, D. and Triebel, H. (2005). Wavelet bases and entropy numbers in weighted function spaces. *Math. Nachr.* **278** 108–132.

[11] Marcus, D. J. (1985). Relationships between Donsker classes and Sobolev spaces. *Z. Wahrsch. Verw. Gebiete* **69** 323–330.

[12] Nickl, R. and Pötscher, B. M. (2005). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov and Sobolev-type. *J. Theoret. Probab.,* forthcoming.

[13] Schwartz, L. (1966). *Théorie des distributions.* Hermann, Paris.

[14] Strassen, V. and Dudley, R. M. (1969). The central limit theorem and epsilon-entropy. Probability and information theory. *Lecture Notes in Math* **1247**, 224–231.

[15] STUTE, W. (1983). Empirical processes indexed by smooth functions. *Stochastic Process. Appl.* **14** 55–66.
[16] TRIEBEL, H. (1978). *Spaces of Besov-Hardy-Sobolev type.* Teubner, Leipzig.
[17] TRIEBEL, H. (1983). *Theory of Function Spaces.* Birkhäuser, Basel.
[18] VAN DER VAART, A. W. (1994). Bracketing smooth functions. *Stochastic Process. Appl.* **52** 93–105.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 196–206

DOI: 10.1214/074921706000000851

On the Bahadur slope of the Lilliefors and the Cramér–von Mises tests of normality

Miguel A. Arcones[1]

Binghamton University

Abstract: We find the Bahadur slope of the Lilliefors and Cramér–von Mises tests of normality.

1. Introduction

The simplest goodness of fit testing problem is to test whether a random sample $X_1, \dots, X_n$ is from a particular c.d.f. F_0. The testing problem is:

$$H_0 : F = F_0, \quad \text{versus } H_1 : F \not\equiv F_0.$$

A common goodness of fit test is the Kolmogorov–Smirnov test (see Chapter 6 in [8]; and Section 5.1 in [13]). The Kolmogorov–Smirnov test rejects the null hypothesis for large values of the statistic

$$\sup_{t \in \mathbb{R}} |F_n(t) - F_0(t)|\,, \tag{1.1}$$

where $F_n(t) = n^{-1} \sum_{j=1}^{n} I(X_j \le t)$, $t \in \mathbb{R}$, is the empirical c.d.f. Another possible test is the Cramér–von Mises test, which is significative for large values of the statistic:

$$\int_{-\infty}^{\infty} [F_n(t) - F_0(t)]^2 dF_0(t). \tag{1.2}$$

Anderson and Darling [1] generalize the previous test by adding a weight function and considering:

$$\int_{-\infty}^{\infty} [F_n(t) - F_0(t)]^2 \psi(F_0(t)) dF_0(t), \tag{1.3}$$

where ψ is a (nonnegative) weight function. The asymptotic distribution of the statistics in (1.1)–(1.3) can be found in [20].

A natural definition of efficiency of tests was given by Bahadur [5, 6]. Let $\{f(\cdot, \theta) : \theta \in \Theta\}$ be a family of p.d.f.'s on a measurable space $(S, \mathcal{S})$ with respect to a measure μ, where Θ is a Borel subset of $\mathbb{R}^d$. Let $X_1, \dots, X_n$ be i.i.d.r.v.'s with values in $(S, \mathcal{S})$ and p.d.f. $f(\cdot, \theta)$, for some unknown value of $\theta \in \Theta$. Let $\Theta_0 \subset \Theta$ and let $\Theta_1 := \Theta - \Theta_0$. Consider the hypothesis testing problem $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$. The level (or significance level) of the test is

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta\{\text{reject } H_0\}.$$

[1]Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902, USA, e-mail: arcones@binghamton.edu

AMS 2000 subject classifications: primary 62F05; secondary 60F10.
Keywords and phrases: Bahadur slope, Lilliefors test of normality, large deviations.

The p–value of a test is the smallest significance level at which the null hypothesis can be rejected. Suppose that a test rejects H_0 if $T_n \geq c$, where $T_n := T_n(X_1, \ldots, X_n)$ is a statistic and c is a constant. Then, the significance level of the test is

$$H_n(c) := \sup_{\theta \in \Theta_0} P_\theta(T_n \geq c), \tag{1.4}$$

where P_θ denotes the probability measure for which the data has p.d.f. $f(\cdot, \theta)$. The p–value of the test is

$$H_n(T_n). \tag{1.5}$$

Notice that the p–value is a r.v. whose distribution depends on n and on the specified value of the alternative hypothesis. Given two different tests, the one with smallest p–value under alternatives is preferred. Since the distribution of a p–value is difficult to calculate, Bahadur (1967, 1971) proposed to compare tests using the quantity

$$c(\theta_1) := -2 \liminf_{n\to\infty} n^{-1} \ln H_n(T_n) \ a.s. \tag{1.6}$$

where the limit is found assuming that $X_1, \ldots, X_n$ are i.i.d.r.v.'s from the p.d.f. $f(\cdot, \theta_1)$, $\theta_1 \in \Theta_1$. The quantity $c(\theta_1)$ is called the Bahadur slope of the test. Given two tests, the one with the biggest Bahadur slope is preferred. For a review on Bahadur asymptotic optimality see and Nikitin [16]. The Bahadur slopes of the tests in (1.1) and (1.2) can be found in Chapter 2 in [16].

For the statistic in (1.1), it is known (see [6]) that if F_0 is a continuous c.d.f., then

$$\lim_{n\to\infty} n^{-1} \ln H_n(T_n) = -G(\sup_{t\in\mathbb{R}} |F(t) - F_0(t)|) \ \ a.s. \tag{1.7}$$

when the data comes from the c.d.f. F, $F \not\equiv F_0$, where

$$G(a) = \inf_{0\leq t\leq 1-a} \left((a+t)\ln(t^{-1}(a+t)) + (1-a-t)\ln((1-t)^{-1}(1-a-t))\right). \tag{1.8}$$

In this paper, we will consider the Bahadur slopes of some tests of normality, i.e. given a random sample $X_1, \ldots, X_n$ from a c.d.f. F we would like to test:

$$H_0 : F \text{ has a normal distribution, versus } H_1 : F \text{ does not,} \tag{1.9}$$

We would like to obtain results similar to the one in (1.7) for several tests of normality. Reviews of normality tests are [12, 21] and [15].

Lilliefors [14] proposed the normality test which rejects the null hypothesis for large values of the statistic

$$\sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|, \tag{1.10}$$

where Φ is the c.d.f. of the standard normal distribution, $\bar{X}_n := n^{-1}\sum_{j=1}^n X_j$ and $s_n^2 := (n-1)^{-1}\sum_{j=1}^n (X_j - \bar{X}_n)^2$. This test can be used because the distribution of (1.10) is location and scale invariant. We will consider the test of normality which rejects the null hypothesis if

$$\sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|\psi(t), \tag{1.11}$$

where $\psi : \mathbb{R} \to [0, \infty)$ be a bounded function.

We also consider the test of normality which rejects normality if

$$\int_{-\infty}^{\infty} [F_n(\bar{X}_n + s_n t) - \Phi(t)]^2 \psi(\Phi(t)) d\Phi(t), \tag{1.12}$$

where $\psi : \mathbb{R} \to [0, \infty)$ satisfies $\int_{-\infty}^{\infty} \psi(F_0(t)) dF_0(t) < \infty$. Notice that the statistics in (1.11) and (1.12) are location and scale invariant.

In Section 2, we present bounds in the Bahadur slope for the statistics in (1.11) and (1.12). Our techniques are based on the (LDP) large deviation principle for empirical processes in [2–4]. We refer to the LDP to [10] and [9]. The proofs are in Section 4.

In Section 3, we present some simulations of the mean of the p–value for several test under different alternatives. The simulations show that Lilliefors test has a high p-value. However, the p-value of the Anderson–Darling is competitive with other test of normality such as the Shapiro–Wilk test ([19]) and the BHEP test ([11] and [7]).

2. Main results

In this section we review some results on the LDP for empirical processes. We determine the rate function of the LDP of empirical processes using Orlicz spaces theory. A reference in Orlicz spaces is [18]. A function $\Upsilon : \mathbb{R} \to \bar{\mathbb{R}}$ is said to be a Young function if it is convex, $\Upsilon(0) = 0$; $\Upsilon(x) = \Upsilon(-x)$ for each $x > 0$; and $\lim_{x\to\infty} \Upsilon(x) = \infty$. Let X be a r.v. with values in a measurable space $(S, \mathcal{S})$. The Orlicz space $\mathcal{L}^{\Upsilon}(S, \mathcal{S})$ (abbreviated to $\mathcal{L}^{\Upsilon}$) associated with the Young function Υ is the class of measurable functions $f : (S, \mathcal{S}) \to \mathbb{R}$ such that $E[\Upsilon(\lambda f(X))] < \infty$ for some $\lambda > 0$. The Minkowski (or gauge) norm of the Orlicz space $\mathcal{L}^{\Upsilon}(S, \mathcal{S})$ is defined as

$$N_{\Upsilon}(f) = \inf\{\lambda > 0 : E[\Upsilon(f(X)/\lambda)] \le 1\}.$$

It is well known that the vector space $\mathcal{L}^{\Upsilon}$ with the norm N_{Υ} is a Banach space. Define

$$\mathcal{L}^{\Upsilon_0} := \{f : S \to \mathbb{R} : E[\Upsilon_0(\lambda|f(X)|)] < \infty \text{ for some } \lambda > 0\},$$

where $\Upsilon_0(x) = e^{|x|} - |x| - 1$. Let $(\mathcal{L}^{\Upsilon_0})^*$ be the dual of $(\mathcal{L}^{\Upsilon_0}, N_{\Upsilon_0})$. The function $f \in \mathcal{L}^{\Upsilon_0} \mapsto \ln\left(E[e^{f(X)}]\right) \in \mathbb{R}$ is a convex lower semicontinuous function. The Fenchel–Legendre conjugate of the previous function is:

$$J(l) := \sup_{f \in \mathcal{L}^{\Upsilon_0}} \left(l(f) - \ln\left(E[e^{f(X)}]\right)\right),\ l \in (\mathcal{L}^{\Upsilon_0})^*. \tag{2.1}$$

J is a function with values in $[0, \infty]$. Since J is a Fenchel–Legendre conjugate, it is a nonnegative convex lower semicontinuous function. If $J(l) < \infty$, then:

(i) $l(\mathbf{1}) = 1$, where $\mathbf{1}$ denotes the function constantly 1.

(ii) l is a nonnegative definite functional: if $f(X) \ge 0$ a.s., then $l(f) \ge 0$.

Since the double Fenchel–Legendre transform of a convex lower semicontinuous function coincides with the original function (see e.g. Lemma 4.5.8 in [9]), we have that

$$\sup_{l \in \mathcal{L}^{\Upsilon_0}} (l(f) - J(l)) = \ln E[e^{f(X)}]. \tag{2.2}$$

The previous function J can be used to determine the rate function in the large deviations of statistics. Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with the distribution of X. If $f_1, \dots, f_m \in \mathcal{L}^{\Upsilon_0}$, then

$$\{(n^{-1}\sum_{j=1}^{n} f_1(X_j), \dots, n^{-1}\sum_{j=1}^{n} f_m(X_j))\}$$

satisfies the LDP in $\mathbb{R}^m$ with speed n and rate function

$$I(u_1, \dots, u_m) := \sup_{\lambda_1,\dots,\lambda_m \in \mathbb{R}} \left(\sum_{j=1}^{m} \lambda_j u_j - \ln E[\exp(\sum_{j=1}^{m} \lambda_j f_j(X))] \right)$$

(see for example Corollary 6.1.16 in [9]). This rate function can be written as

$$\inf \left\{ J(l) : l \in (\mathcal{L}^{\Upsilon_0})^*, l(f_j) = u_j \text{ for each } 1 \le j \le m \right\},$$

(see Lemma 2.2 in [4]).

To deal with empirical processes, we will use the following theorem:

Theorem 2.1 (Theorem 2.8 in [3]). *Suppose that* $\sup_{t \in T} |f(X,t)| < \infty$ *a.s. Then, the following sets of conditions* ((a) *and* (b)) *are equivalent:*

(a.1) (T, d) *is totally bounded, where* $d(s,t) = E[|f(X,s) - f(X,t)|]$.

(a.2) *There exists a* $\lambda > 0$ *such that*

$$E[\exp(\lambda F(X))] < \infty,$$

where $F(x) = \sup_{t \in T} |f(x,t)|$.

(a.3) *For each* $\lambda > 0$, *there exists a* $\eta > 0$ *such that* $E[\exp(\lambda F^{(\eta)}(X))] < \infty$, *where* $F^{(\eta)}(x) = \sup_{d(s,t) \le \eta} |f(x,s) - f(x,t)|$.

(a.4) $\sup_{t \in T} |n^{-1} \sum_{j=1}^{n} (f(X_j,t) - E[f(X_j,t)])| \xrightarrow{\Pr} 0$.

(b) $\{n^{-1} \sum_{j=1}^{n} f(X_j,t) : t \in T\}$ *satisfies the large deviation principle in* $l_\infty(T)$ *with speed* n *and a good rate.*

Besides, the rate function is

$$I(z) = \inf\{J(l) : l \in (\mathcal{L}^{\Upsilon_0})^*, l(f(\cdot,t)) = z(t), \text{ for each } t \in T\}, z \in l_\infty(T).$$

We will consider large deviations when the r.v.'s have a standard normal distribution. We denote $(\mathcal{L}_\Phi^{\Upsilon_0}, N_{\Upsilon_0})$ to the Orlicz space, when the distribution of X is a standard normal one. Similarly,

$$J_\Phi(l) := \sup_{f \in \mathcal{L}_\Phi^{\Upsilon_0}} \left(l(f) - \ln \left(E_\Phi[e^{f(X)}] \right) \right), \ l \in (\mathcal{L}^{\Upsilon_0})^*. \tag{2.3}$$

First, we consider the Bahadur efficiency of the test in (1.11). Next lemma considers the large deviations of the test statistic in (1.11) under the null hypothesis.

Lemma 2.1. *Let $\psi : \mathbb{R} \to \mathbb{R}$ be a bounded function. Then, for each $u \geq 0$,*

$$
\begin{aligned}
&- \inf \Big\{ J_\Phi(l) : l \in (\mathcal{L}_\Phi^{\Upsilon_0})^*, \sup_{t\in\mathbb{R}} |x(a + (b-a^2)^{1/2} t) - \Phi(t)|\psi(t) > u, \\
&\qquad l(f_t) = x(t), t \in \mathbb{R}, l(g) = a, l(g^2) = b, \\
&\qquad \text{where } f_t(s) = I(s \leq t), g(s) = s, s \in \mathbb{R} \Big\} \\
&\quad \leq \liminf_{n\to\infty} n^{-1} \ln P_\Phi \Big\{ \sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|\psi(t) > u \Big\} \\
&\quad \leq \limsup_{n\to\infty} n^{-1} \ln P_\Phi \Big\{ \sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|\psi(t) \geq u \Big\} \\
&\quad \leq - \inf \Big\{ J_\Phi(l) : l \in (\mathcal{L}_\Phi^{\Upsilon_0})^*, \sup_{t\in\mathbb{R}} |x(a + (b-a^2)^{1/2} t) - \Phi(t)|\psi(t) \geq u, \\
&\qquad\qquad l(f_t) = x(t), t \in \mathbb{R}, l(g) = a, l(g^2) = b, \\
&\qquad\qquad \text{where } f_t(s) = I(s \leq t), g(s) = s, s \in \mathbb{R} \Big\}
\end{aligned}
\tag{2.4}
$$

Theorem 2.2. *Let $\psi : \mathbb{R} \to \mathbb{R}$ be a bounded function, let*

$$
H_n^{\mathrm{Li}}(u) := P_\Phi \Big\{ \sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|\psi(t) \geq u \Big\}, u \geq 0,
\tag{2.5}
$$

and let

$$
\begin{aligned}
G^{\mathrm{Li}}(u) := \inf \Big\{ & J_\Phi(l) : l \in (\mathcal{L}_\Phi^{\Upsilon_0})^*, \sup_{t\in\mathbb{R}} |x(a + (b-a^2)^{1/2} t) - \Phi(t)|\psi(t) \geq u, \\
& l(f_t) = x(t), t \in \mathbb{R},\ l(g) = a,\ l(g^2) = b, \\
& \text{where } f_t(s) = I(s \leq t), g(s) = s, s \in \mathbb{R} \}
\end{aligned}
$$

Let $\{X_j\}_{j=1}^\infty$ be a sequence of i.i.d.r.v.'s with c.d.f. F. Then,

$$
\begin{aligned}
&- \lim_{\delta\to 0+} G^{\mathrm{Li}} \Big(\sup_{t\in\mathbb{R}} |F(\mu_F + \sigma_F t) - \Phi(t)|\psi(t) + \delta \Big) \\
&\quad \leq \liminf_{n\to\infty} n^{-1} \ln H_n^{\mathrm{Li}} \Big(\sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|\psi(t) \Big) \\
&\quad \leq \limsup_{n\to\infty} n^{-1} \ln H_n^{\mathrm{Li}} \Big(\sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|\psi(t) \Big) \\
&\quad \leq - \lim_{\delta\to 0+} G^{\mathrm{Li}} \Big(\sup_{t\in\mathbb{R}} |F(\mu_F + \sigma_F t) - \Phi(t)|\psi(t) - \delta \Big) \ \text{a.s.}
\end{aligned}
\tag{2.6}
$$

where $\mu_F = E_F[X]$ and $\sigma_F^2 = \mathrm{Var}_F(X)$.

For the statistic in (1.12), we have similar results:

Lemma 2.2. *Let $\psi : \mathbb{R} \to [0, \infty)$ be a function such that $\int_{-\infty}^{\infty} \psi(F_0(t)) dF_0(t) < \infty$.*

Then, for each $u \geq 0$,

$$-\inf\Big\{J_\Phi(l) : l \in (\mathcal{L}_\Phi^{\Upsilon_0})^*, \\ \int_{-\infty}^{\infty} [x(a + (b - a^2)^{1/2} t) - \Phi(t)]^2 \psi(F_0(t)) dF_0(t) > u, \\ l(f_t) = x(t), t \in \mathbb{R},\ l(g) = a,\ l(g^2) = b, \\ \text{where } f_t(s) = I(s \leq t), g(s) = s, s \in \mathbb{R}\Big\}$$

$$\begin{aligned} &\leq \liminf_{n\to\infty} n^{-1} \ln P_\Phi \left\{\int_{-\infty}^{\infty} [F_n(\bar{X}_n + s_n t) - \Phi(t)]^2 \psi(F_0(t)) dF_0(t) > u\right\} \\ &\leq \limsup_{n\to\infty} n^{-1} \ln P_\Phi \left\{\int_{-\infty}^{\infty} [F_n(\bar{X}_n + s_n t) - \Phi(t)]^2 \psi(F_0(t)) dF_0(t) \geq u\right\} \\ &\leq -\inf\Big\{J_\Phi(l) : l \in (\mathcal{L}_\Phi^{\Upsilon_0})^*, \\ &\qquad \int_{-\infty}^{\infty} [x(a + (b - a^2)^{1/2} t - \Phi(t))]^2 \psi(F_0(t)) dF_0(t) > u, \\ &\qquad l(f_t) = x(t), t \in \mathbb{R},\ l(g) = a,\ l(g^2) = b, \\ &\qquad \text{where } f_t(s) = I(s \leq t), g(s) = s, s \in \mathbb{R}\Big\} \end{aligned} \tag{2.7}$$

Theorem 2.3. *Let $\psi : \mathbb{R} \to [0, \infty)$ be a function such that $\int_{\mathbb{R}} \psi(x)\, dx < \infty$, let*

$$H_n^{\mathrm{AD}}(u) := P_\Phi \left\{\int_{-\infty}^{\infty} (F_n(\bar{X}_n + s_n t) - \Phi(t))^2 \psi(\Phi(t)) d\Phi(t) \geq u\right\}, u \geq 0, \tag{2.8}$$

and let

$$\begin{aligned} G^{\mathrm{AD}}(u) := \inf\Big\{&J_\Phi(l) : l \in (\mathcal{L}_\Phi^{\Upsilon_0})^*, \\ &\int_{-\infty}^{\infty} [x(a + (b - a^2)^{1/2} t) - \Phi(t)]^2 \psi(F_0(t)) dF_0(t) \geq u, \\ &l(f_t) = x(t), t \in \mathbb{R},\ l(g) = a,\ l(g^2) = b, \\ &\text{where } f_t(s) = I(s \leq t), g(s) = s, s \in \mathbb{R}\Big\}. \end{aligned} \tag{2.9}$$

Let $\{X_j\}_{j=1}^\infty$ be a sequence of i.i.d.r.v.'s with a continuous c.d.f. F. Then,

$$\begin{aligned} &-\lim_{\delta\to 0+} G^{\mathrm{AD}} \left(\int_{-\infty}^{\infty} [F(\mu_F + \sigma_F t) - \Phi(t)]^2 \psi(\Phi(t)) d\Phi(t) + \delta\right) \\ &\leq \liminf_{n\to\infty} n^{-1} \ln H_n^{\mathrm{AD}} \left(\int_{-\infty}^{\infty} [F_n(\bar{X}_n + s_n t)) - \Phi(t)]^2 \psi(\Phi(t)) d\Phi(t) > u\right) \\ &\leq \limsup_{n\to\infty} n^{-1} \ln H_n^{\mathrm{AD}} \left(\int_{-\infty}^{\infty} [F_n(\bar{X}_n + s_n t) - \Phi(t)]^2 \psi(\Phi(t)) d\Phi(t) \geq u\right) \\ &\leq -\lim_{\delta\to 0+} G^{\mathrm{AD}} \left(\int_{-\infty}^{\infty} [F(\mu_F + \sigma_F t) - \Phi(t)]^2 \psi(\Phi(t)) d\Phi(t) - \delta\right) \text{ a.s.} \end{aligned} \tag{2.10}$$

3. Simulations

We present simulations of the mean of the p-value of several alternatives. As before, suppose that a test rejects H_0 if $T_n \geq c$, where $T_n := T_n(X_1, \ldots, X_n)$ is a statistic and c is a constant. The significance level of the test is

$$H_n(c) := \sup_{\theta \in \Theta_0} P_\theta(T_n \geq c), \tag{3.1}$$

where P_θ denotes the probability measure for which the data has p.d.f. $f(\cdot, \theta)$. The p–value of the test is $H_n(T_n)$. We do simulations estimating $E[H_n(T_n)]$.

Let $T_n^1, \ldots, T_n^N$ be N simulations of the test under the null hypothesis using a sample size n. Let $T_n^{1,\mathrm{alt}}, \ldots, T_n^{1,\mathrm{alt}}$ be N simulations of the test under a certain alternative hypothesis. Then, $N^{-2} \sum_{j,k=1}^{N} I(T_n^j \geq T_n^{k,\mathrm{alt}})$ estimates $E[H_n(T_n)]$, where the expectation is taken assuming that T_n is obtained using n i.i.d.r.v.s from the alternative hypothesis. In Table 1, $N = 10000$ is used for the Lilliefors, the Cramer–von Mises, the Anderson–Darling, the Shapiro-Wilk, and the BHEP test ([11] and [7]).

Table 1

n	L	CM	AD	SW	BHEP
Alternative: exponential distribution					
10	0.248325	0.2065738	0.1878327	0.1621557	0.1813481
15	0.1601991	0.1178508	0.0946044	0.07611985	0.09569895
20	0.1043946	0.06510648	0.05291726	0.03304067	0.05206452
30	0.044566	0.02152872	0.0129459	0.00750638	0.01409681
50	0.00818707	0.00203949	0.0009082	0.00241882	0.00121646
Alternative: double exponential distribution					
10	0.3992631	0.3939724	0.3983314	0.4148731	0.4109459
15	0.3499758	0.339891	0.3389608	0.3677549	0.3640368
20	0.3169975	0.2979569	0.3009778	0.3294354	0.3244178
30	0.2616672	0.2341172	0.2397223	0.2807123	0.2555247
50	0.1796492	0.1417312	0.1434135	0.2837385	0.1564005
Alternative: Cauchy distribution					
10	0.1566254	0.1493284	0.1500601	0.1705618	0.1682588
15	0.08474307	0.07479607	0.07505173	0.09128725	0.08964657
20	0.04651569	0.03862244	0.03767999	0.04857044	0.04194876
30	0.01420118	0.0100633	0.00974881	0.01496182	0.01179398
50	0.0017361	0.00066862	0.00087474	0.00405	0.00048095
Alternative: Beta(2,1) distribution					
10	0.41099	0.3884071	0.3686801	0.3358273	0.3565215
15	0.3608343	0.321224	0.2986805	0.2631936	0.2861669
20	0.3147082	0.272089	0.2446055	0.1861695	0.2330953
30	0.2411935	0.1840438	0.1534217	0.09638084	0.1428481
50	0.1355312	0.08707787	0.05502258	0.0835284	0.0572854
Alternative: Beta(3,3) distribution					
10	0.5084849	0.5040435	0.51387	0.4928259	0.4947155
15	0.5063525	0.5029377	0.5033103	0.4872658	0.4864098
20	0.5072011	0.5037995	0.4993457	0.4762998	0.4797991
30	0.4899722	0.4745532	0.4780857	0.4285843	0.4510982
50	0.4590308	0.4447785	0.4382911	0.4189339	0.4064414

TABLE 1 (Continued)

Alternative: Logistic(1) distribution					
10	0.4736219	0.4725762	0.4685902	0.4749748	0.4677617
15	0.4560905	0.4468502	0.4335687	0.4624648	0.46532
20	0.4493409	0.4488339	0.4450634	0.4426041	0.4510982
30	0.4410423	0.422886	0.4233825	0.4348024	0.4153006
50	0.4204326	0.3978938	0.3770672	0.4458524	0.3819914
Alternative: uniform distribution					
10	0.4438842	0.4241476	0.404153	0.3681328	0.3922898
15	0.4059967	0.3716177	0.3468994	0.3023148	0.338187
20	0.3739766	0.3308353	0.2951247	0.2270906	0.27552
30	0.3050368	0.2429758	0.2024329	0.117917	0.1862224
50	0.2066687	0.1359771	0.0889871	0.1007488	0.08932786

We should expect the average p–value is small than 0.5. However, for the Beta$(3, 3)$ distribution the average p-value is bigger than 0.5. Between the tree test which use the empirical c.d.f., the Anderson–Darling test is the one with smallest average p–value. For almost half of the considered distributions, this is the test with the smallest average p–value. The Shapiro–Wilk test also performs very well overall.

4. Proofs

We will need the following lemmas:

Lemma 4.1 (Lemma 5.1 (i), [4]). *For each $k \geq 0$ and each function $f \in \mathcal{L}^{\Upsilon_0}$,*

$$|l(f)| \leq (J(l) + 1 + 2^{1/2})N_{\Upsilon_0}(f).$$

Lemma 4.2. *Let $l \in (\mathcal{L}_\Phi^{\Upsilon_0})^*$ with $J_\Phi(l) < \infty$. Then, $x(t) = l(I(\cdot \leq t))$, $t \in \mathbb{R}$, is a continuous function with $\lim_{t\to-\infty} x(t) = 0$ and $\lim_{t\to\infty} x(t) = 1$.*

Proof. Let $\alpha : (0, \infty) \to (0, \infty)$ be defined as $\alpha(x) = \exp(1/x) - 1/x$. It is easy to see that α is one–to–one function. We claim that for a Borel set $A \subset \mathbb{R}$,

$$N_{\Upsilon_0}(I(X \in A)) = \alpha^{-1}(1 + (P_\Phi(X \in A))^{-1}), \tag{4.1}$$

where α^{-1} denotes the inverse function of α. We have that

$$\begin{aligned} E_\Phi[\Upsilon_0(\lambda^{-1}I(X \in A))] &= E[\exp(\lambda^{-1}I(X \in A)) - 1 - \lambda^{-1}I(X \in A)] \\ &= p\exp(\lambda^{-1}) + 1 - p - 1 - \lambda^{-1}p, \end{aligned}$$

where $p := P_\Phi(X \in A)$. So, $1 \geq E_\Phi[\Upsilon_0(\lambda^{-1}I(X \in A))]$ is equivalent to $\alpha(\lambda) \leq 1 + p^{-1}$. So, (4.1) follows.

By Lemma 4.1 and (4.1), for each $s, t \in \mathbb{R}$ with $s < t$,

$$|x(t)| = |l(I(X \leq t)| \leq (J_\Phi(l) + 1 + 2^{1/2})\alpha^{-1}(1 + (\Phi(t))^{-1}),$$

$$|1 - x(t)| = |l((X > t))| \leq (J_\Phi(l) + 1 + 2^{1/2})\alpha^{-1}(1 + (1 - \Phi(t))^{-1}),$$

and

$$|x(t) - x(s)| \leq (J_\Phi(l) + 1 + 2^{1/2})\alpha^{-1}(1 + (\Phi(t) - \Phi(s))^{-1}),$$

which implies the claim. □

Proof of Lemma 2.1. By Theorem 2.1, $\{U_n(t) : t \in \mathbb{R}\}$ satisfies the LDP in $l_\infty(\mathbb{R})$ with speed n, where $U_n(t) = F_n(t)$. Let ω_1 and ω_2 be two numbers, which are not in $\mathbb{R}$. Let $U_n(\omega_1) = n^{-1}\sum_{j=1}^n X_j$ and let $U_n(\omega_2) = n^{-1}\sum_{j=1}^n X_j^2$. By the LDP for sums of i.i.d. $\mathbb{R}^d$–valued r.v.'s, the finite dimensional distributions of $\{U_n(t) : t \in \mathbb{R} \cup \{\omega_1, \omega_2\}\}$ satisfy the LDP with speed n. Since $\{U_n(t) : t \in \mathbb{R}\}$ satisfies the LDP in $l_\infty(\mathbb{R})$, it satisfies an exponential asymptotic equicontinuity condition (see Theorem 3.1 in [2]). This implies that $\{U_n(t) : t \in \mathbb{R} \cup \{\omega_1, \omega_2\}\}$ satisfies an exponential asymptotic equicontinuity condition. So, $\{U_n(t) : t \in \mathbb{R} \cup \{\omega_1, \omega_2\}\}$ satisfies the LDP in $l_\infty(\mathbb{R}\cup\{\omega_1, \omega_2\})$ with speed n (see Theorem 3.1 in [2]). Besides, the rate of function is

$$\begin{aligned} I(x) = \inf\{&J_\Phi(l) : l \in (\mathcal{L}^{\Upsilon_0})^*, l(I(X \le t) = x(t), t \in \mathbb{R}, l(X) = x(\omega_1),\\ &l(X^2) = x(\omega_2), x \in l_\infty(\mathbb{R} \cup \{\omega_1, \omega_2\})\}\end{aligned}$$

Let $\Gamma_n : l_\infty(\mathbb{R} \cup \{\omega_1, \omega_2\}) \to \mathbb{R}$ be defined by

$$\Gamma_n(x) = \sup_{t\in\mathbb{R}} |x(x(\omega_1) + n^{1/2}(n-1)^{-1/2}(x(\omega_2) - (x(\omega_1))^2)^{1/2}t) - \Phi(t)|\psi(t),$$

for $x \in l_\infty(\mathbb{R} \cup \{\omega_1, \omega_2\})$. Next, we prove using Theorem 2.1 in Arcones (2003a) that

$$\Gamma_n(\{U_n(t) : t \in \mathbb{R} \cup \{\omega_1, \omega_2\}\}) = \sup_{t\in\mathbb{R}} |F_n(\bar{X}_n + s_n t) - \Phi(t)|\psi(t)$$

satisfies the LDP in $\mathbb{R}$ with speed n and rate function

$$\begin{aligned} q^{\mathrm{Li}}(u) := \inf\{&J_\Phi(l) : l \in (\mathcal{L}^{\Upsilon_0})^*, \sup_{t\in\mathbb{R}} |x(a + (b-a^2)^{1/2}t) - \Phi(t)|\psi(t) = u,\\ &l(I(X \le t)) = x(t), t \in \mathbb{R},\ l(X) = a,\ l(X^2) = b\}.\end{aligned} \tag{4.2}$$

To apply Theorem 2.1 in [2], we need to prove that if $x_n \to x$, in $l_\infty(\mathbb{R} \cup \{\omega_1, \omega_2\})$ and $I(x) < \infty$, then $\Gamma_n(x_n) \to \Gamma(x)$, where

$$\Gamma(x) = \sup_{t\in\mathbb{R}} |x(x(\omega_1) + (x(\omega_2) - (x(\omega_1))^2)^{1/2}t) - \Phi(t)|\psi(t), x \in l_\infty(\mathbb{R} \cup \{\omega_1, \omega_2\}).$$

We have that

$$\begin{aligned} &|\Gamma_n(x_n) - \Gamma(x)|\\ &\quad\le \sup_{t\in\mathbb{R}} |x_n(x_n(\omega_1) + n^{1/2}(n-1)^{-1/2}(x_n(\omega_2) - (x_n(\omega_1))^2)^{1/2}t)\psi(t)\\ &\qquad\qquad - x(x_n(\omega_1) + n^{1/2}(n-1)^{-1/2}(x_n(\omega_2) - (x_n(\omega_1))^2)^{1/2}t)\psi(t)|\\ &\qquad + \sup_{t\in\mathbb{R}} |x(x_n(\omega_1) + n^{1/2}(n-1)^{-1/2}(x_n(\omega_2) - (x_n(\omega_1))^2)^{1/2}t)\psi(t)\\ &\qquad\qquad - x(x(\omega_1) + (x(\omega_2) - (x(\omega_1))^2)^{1/2}t)\psi(t)|\\ &\quad =: I + II.\end{aligned}$$

Since $x_n \to x$, in $l_\infty(\mathbb{R} \cup \{\omega_1, \omega_2\})$,

$$I = \sup_{t\in\mathbb{R}} |x_n(t) - x(t)|\psi(t) \to 0.$$

By Lemma 4.2, x is a continuous function with $\lim_{t\to-\infty} x(t) = 0$ and $\lim_{t\to\infty} x(t) = 1$. So,

$$II \to 0.$$

From the previous computations, we get that $\Gamma_n(x_n) \to \Gamma(x)$. Hence, $\sup_{t\in\mathbb{R}} |F_n \times (\bar{X}_n + s_n t) - \Phi(t)|$ satisfies the LDP in $\mathbb{R}$ with speed n and rate function $q^{\mathrm{Li}}(u)$. This implies (2.4). □

Proof of Theorem 2.2. We have that

$$\begin{aligned}
&|\sup_{t\in\mathbb{R}}|F_n(\bar{X}_n+s_nt))-\Phi(t)|\psi(t)-\sup_{t\in\mathbb{R}}|F(\mu_F+\sigma_Ft)-\Phi(t)|\psi(t)| \\
&\quad\le \sup_{t\in\mathbb{R}}|F_n(\bar{X}_n+s_nt)-F(\mu_F+\sigma_Ft)| \\
&\quad\le \sup_{t\in\mathbb{R}}|F_n(t)-F(t)|\psi(t)+\sup_{t\in\mathbb{R}}|F(\bar{X}_n+s_nt))-F(\mu_F+\sigma_Ft)|\psi(t)|.
\end{aligned}$$

By the Glivenko–Cantelli theorem,

$$\sup_{t\in R}|F_n(t)-F(t)|\psi(t)\to 0 \quad a.s.$$

Using that

$$\bar{X}_n\to\mu_F \quad a.s., s_n\to\sigma_F \quad a.s.$$

and F is a continuous function with $\lim_{t\to-\infty}F(t)=0$ and $\lim_{t\to\infty}F(t)=1$, we get that

$$\sup_{t\in\mathbb{R}}|F(\bar{X}_n+s_nt)-F(\mu_F+\sigma_Ft)|\psi(t)\to 0 \quad a.s.$$

Hence,

$$\sup_{t\in\mathbb{R}}|F_n(\bar{X}_n+s_nt)-\Phi(t)|\psi(t)\to\sup_{t\in\mathbb{R}}|F(\mu_F+\sigma_Ft)-\Phi(t)|\psi(t)| \quad a.s. \tag{4.3}$$

The claim in this theorem follows from (4.3) and Lemma 2.1. □

The proofs of Lemma 2.2 and Theorem 2.3 are similar to those of Lemma 2.1 and Theorem 2.2 and they are omitted.

References

[1] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212. MR0050238

[2] Arcones, M. A. (2003a). The large deviation principle for stochastic processes I. 2003. *Theor. Probab. Applic.* **47** 567–583. MR2001788

[3] Arcones, M. A. (2003b). The large deviation principle for empirical processes. In *High Dimensional Probability III* (eds. J. Hoffmann-Jorgensen, M. B. Marcus and J. A. Wellner). Birkhäuser, Boston, pp. 205–223. MR2033890

[4] Arcones, M. A. (2006). Large deviations for M–estimators. *Ann. Inst. Statist. Mathem.* **58**, 21–52.

[5] Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Mathem. Statist.* **38** 303–324. MR0207085

[6] Bahadur, R. R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia, PA. MR0315820

[7] Baringhaus, L. and Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* **35** 339–348. MR0980849

[8] Conover, W. J. (1999). *Practical Nonparametric Statistics*, 3rd edition. John Wiley, New York.

[9] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications.* Springer, New York. MR1619036

[10] Deuschel, J. D. and Stroock, D. W. (1989). *Large Deviations.* Academic Press, Inc., Boston, MA. MR0997938

[11] Epps, T. W. and Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* **70** 723–726. MR0725389

[12] Henze, N. (2002). Invariant tests for multivariate normality: a critical review. *Statist. Pap.* **43** 467–506. MR 1932769

[13] Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods*, 2nd edition. John Wiley, New York. MR 1666064

[14] Lilliefors, H. (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Amer. Statist. Assoc.* **62** 399–402.

[15] Mecklin, C. J. and Mundfrom, D. J. (2004). An appraisal and bibliography of tests for multivariate normality. *Internat. Statist. Rev.* **72** 123–138.

[16] Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests.* Cambridge University Press, Cambridge, UK. MR1335235

[17] Raghavachari, M. (1970). On a theorem of Bahadur on the rate of convergence of test statistics. *Ann. Mathem. Statist.* **41** 1695–1699. MR0266361

[18] Rao, M. M. and Ren, Z. D. (1991). *Theory of Orlicz Spaces.* Marcel Dekker, New York. MR1113700

[19] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality: complete samples. *Biomet.* **52** 591–611. MR0205384

[20] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley, New York. MR0838963

[21] Thode, H. C., Jr. (2002). *Testing for Normality.* Marcel Dekker, New York. MR1989476

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 207–219

DOI: 10.1214/074921706000000860

Some facts about functionals of location and scatter

R. M. Dudley[1,*]

Massachusetts Institute of Technology

Abstract: Assumptions on a likelihood function, including a local Glivenko-Cantelli condition, imply the existence of M-estimators converging to an M-functional. Scatter matrix-valued estimators, defined on all empirical measures on $\mathbb{R}^d$ for $d \geq 2$, and equivariant under all, including singular, affine transformations, are shown to be constants times the sample covariance matrix. So, if weakly continuous, they must be identically 0. Results are stated on existence and differentiability of location and scatter functionals, defined on a weakly dense, weakly open set of laws, via elliptically symmetric t distributions on $\mathbb{R}^d$, following up on work of Kent, Tyler, and Dümbgen.

1. Introduction

In this paper a *law* will be a Borel probability measure on $\mathbb{R}^d$. Let $\mathcal{N}_d$ be the set of all $d \times d$ nonnegative definite symmetric matrices and $\mathcal{P}_d \subset \mathcal{N}_d$ the subset of strictly positive definite symmetric matrices. For $(\mu, \Sigma) \in \Theta = \mathbb{R}^d \times \mathcal{N}_d$, μ will be viewed as a location parameter and Σ as a scatter parameter, extending the notions of mean vector and covariance matrix to arbitrarily heavy-tailed distributions. For $d \geq 2$, Θ may be taken to be $\mathcal{P}_d$ or $\mathbb{R}^d \times \mathcal{P}_d$.

For a law P on $\mathbb{R}^d$, let $X_1, X_2, \dots$ be i.i.d. (P) and let P_n be the empirical measure $n^{-1}\sum_{j=1}^n \delta_{X_j}$ where $\delta_x(A) := 1_A(x)$ for any point x and set A. A class $\mathcal{F} \subset \mathcal{L}^1(\mathbb{R}^d, P)$ is called a *Glivenko-Cantelli* class for P if

$$\sup\{|\int f d(P_n - P)| : \ f \in \mathcal{F}\} \to 0 \tag{1}$$

almost surely as $n \to \infty$ (if the supremum is measurable, as it will be in all cases considered in this paper). Talagrand [20, 21] characterized such classes. A class $\mathcal{F}$ of Borel measurable functions on $\mathbb{R}^d$ is called a *universal* Glivenko-Cantelli class if it is a Glivenko-Cantelli class for all laws P on $\mathbb{R}^d$, and a *uniform* Glivenko-Cantelli class if the convergence in (1) is uniform over all laws P. Rather general sufficient conditions for the universal Glivenko-Cantelli property and a characterization up to measurability of the uniform property have been given [7].

Let $\rho : \ (x, \theta) \mapsto \rho(x, \theta) \in \mathbb{R}$ defined for $x \in \mathbb{R}^d$ and $\theta \in \Theta$, Borel measurable in x and lower semicontinuous in θ, i.e. $\rho(x, \theta) \leq \liminf_{\phi \to \theta} \rho(x, \phi)$ for all x and θ. For a law Q, let $Q\rho(\phi) := \int \rho(x, \phi) dQ(x)$ if the integral is defined (not $\infty - \infty$), as it always will be if $Q = P_n$. An *M-estimate* of θ for a given n and P_n will be a $\hat{\theta}_n$ such that $P_n\rho(\theta)$ is minimized at $\theta = \hat{\theta}_n$, if it exists and is unique. A measurable

[1]Room 2-245, MIT, Cambridge, MA 02139-4307, USA, e-mail: rmd@math.mit.edu
[*]Partially supported by NSF Grants DMS-0103821 and DMS-0504859.
AMS 2000 subject classifications: primary 62G05, 62GH20; secondary 62G35.
Keywords and phrases: equivariance, t distributions.

function, not necessarily defined a.s., whose values are M-estimates is called an M-estimator. An *M-limit* $\theta_0 = \theta_0(P) = \theta_0(P, \rho)$ (with respect to ρ) will mean a point of Θ such for every open neighborhood U of θ_0, as $n \to \infty$,

$$\Pr\Big\{ \inf\{P_n\rho(\theta):\ \theta \notin U\} \le \inf\{P_n\rho(\phi):\ \phi \in U\}\Big\} \to 0, \tag{2}$$

where the given probabilities are assumed to be defined. Then if M-estimators exist (with probability $\to 1$ as $n \to \infty$), they must converge in probability to $\theta_0(P)$. An M-limit $\theta_0 = \theta_0(P)$ with respect to ρ will be called *definite* iff for every neighborhood U of θ_0 there is an $\varepsilon > 0$ such that the outer probability

$$(P^n)^* \left\{\inf\{P_n\rho(\theta):\ \theta \notin U\} \le \varepsilon + \inf\{P_n\rho(\phi):\ \phi \in U\}\right\} \to 0 \tag{3}$$

as $n \to \infty$.

For a law P on $\mathbb{R}^d$ and a given $\rho(\cdot,\cdot)$, a $\theta_1 = \theta_1(P)$ is called the *M-functional* of P for ρ if and only if there exists a measurable function $a(x)$, called an *adjustment function*, such that for $h(x,\theta) = \rho(x,\theta) - a(x)$, $Ph(\theta)$ is defined and satisfies $-\infty < Ph(\theta) \le +\infty$ for all $\theta \in \Theta$, and is minimized uniquely at $\theta = \theta_1(P)$, e.g. Huber [13]. As Huber showed, $\theta_1(P)$ doesn't depend on the choice of $a(\cdot)$. Clearly, an M-estimate $\hat\theta_n$ is the M-functional $\theta_1(P_n)$ if either exists.

A lower semicontinuous function f from Θ into $(-\infty, +\infty]$ will be called *uniminimal* iff it has a unique relative minimum at a point θ_0 and for all $t \in \mathbb{R}$, $\{\theta \in \Theta :\ f(\theta) \le t\}$ is connected. For a differentiable function f, recall that a *critical point* of f is a point where the gradient of f is 0.

Examples. On $\Theta = \mathbb{R}$ let $f(x) = -(1-x^2)^2$. Then f has a unique relative minimum at $x = 0$, but no absolute minimum. It has two other critical points which are relative maxima. For $t < 0$ the set where $f \le t$ is not connected.

If f is a strictly convex function on $\mathbb{R}^d$ attaining its minimum, then f is uniminimal, as is $\theta \mapsto f(x-\theta)$ for any x. So is $\theta \mapsto \int f(x-\theta) - f(x)dP(x)$ if it's defined and finite and attains its minimum for a law P, as will be true e.g. if $f(x) = |x|^2$ and $\int |x|dP(x) < \infty$, or for all P if f is also Lipschitz, e.g. $f(x) = \sqrt{1+|x|^2}$.

I have not found the notion here called "uniminimal" in the literature. Similar but more complex assumptions occur in some work on sufficient conditions for minimaxity in game theory, e.g. [11]. Thus, I claim no originality for the following easily proved fact.

Proposition 1. *Let (Θ, d) be a locally compact metric space. If f is uniminimal on (Θ, d), then (a) f attains its absolute minimum at its unique relative minimum θ_0, and (b) For every neighborhood U of θ_0 there is an $\varepsilon > 0$ such that $f(\theta) \ge f(\theta_0) + \varepsilon$ for all $\theta \notin U$.*

Proof. Clearly (b) implies (a). To prove (b), suppose that for some or equivalently all small enough $\delta > 0$ and all $n = 1, 2, \ldots$, there are $\theta_n \in \Theta$ with $d(\theta_n, \theta_0) \ge \delta$ and $f(\theta_n) \le f(\theta_0) + 1/n$. By connectedness, we can take $d(\theta_n, \theta_0) = \delta$ for all n. Then for $\delta > 0$ small enough, $\{\theta :\ d(\theta, \theta_0) \le \delta\}$ is compact and there is a converging subsequence $\theta_{n(k)} \to \theta_\delta$ with $d(\theta_\delta, \theta_0) = \delta$ and $f(\theta_\delta) \le f(\theta_0)$ by lower semicontinuity. Letting $\delta \downarrow 0$ we get a contradiction to the fact that θ_0 is a unique relative minimum. □

Theorem 2. *Let (Θ, d) be a connected locally compact metric space and $(X, \mathcal{B}, P)$ a probability space. Let $h :\ X \times \Theta \mapsto \mathbb{R}$ where for each $\theta \in \Theta$, $h(\cdot, \theta)$ is measurable. Assume that:*

(i) $\theta \mapsto Ph(\theta) \in (-\infty, +\infty]$ *is well-defined and uniminimal on* Θ*, with minimum at* θ_0*;*
(ii) *Outside an event* A_n *whose probability converges to 0 as* $n \to \infty$*,* $P_n h(\cdot)$ *is uniminimal on* Θ*;*
(iii) *For some neighborhood* U *of* θ_0*,* $\{h(\cdot, \theta) : \theta \in U\}$ *is a Glivenko-Cantelli class for* P*.*

Then θ_0 *is the definite M-limit for* P *and the M-functional* $\theta_1(P)$*.*

Remark. Glivenko-Cantelli conditions on log likelihoods (and their partial derivatives through order 2) for parameters in bounded neighborhoods have been assumed in other work, e.g. [17] and [8].

Proof. That θ_0 is an M-functional for P follows from (i) and Proposition 1. By (iii), take $\delta > 0$ small enough so that $\{h(\cdot, \theta) : d(\theta, \theta_0) < \delta\}$ is a Glivenko-Cantelli class for P. By (i) and Proposition 1, take $\varepsilon > 0$ such that $Ph(\theta) > Ph(\theta_0) + 3\varepsilon$ whenever $d(\theta, \theta_0) > \delta/2$. Outside some events A_n whose probability converges to 0 as $n \to \infty$, we have $P_n h(\theta_0) < Ph(\theta_0) + \varepsilon$ and $P_n h(\theta) > Ph(\theta_0) + 2\varepsilon$ for all θ with $\delta/2 < d(\theta, \theta_0) < \delta$. Then by (ii), also with probability converging to 1, $P_n h(\theta) > P_n h(\theta_0) + \varepsilon$ for all θ with $d(\theta, \theta_0) > \delta/2$, proving (3) and the theorem. ☐

A class $\mathcal{C}$ of subsets of a set X is called a *VC (Vapnik-Chervonenkis)* class if for some $k < \infty$, for every subset A of X with k elements, there is some $B \subset A$ with $B \neq C \cap A$ for all $C \in \mathcal{C}$, e.g. [4, Chapter 4]. A class $\mathcal{F}$ of real-valued functions on X is called a *VC major* class iff $\{\{x \in X : f(x) > t\} : f \in \mathcal{F}, t \in \mathbb{R}\}$ is a VC class of sets (e.g. [4, Section 4.7]). In the following, local compactness is stronger than needed but holds for the parameter spaces being considered.

Theorem 3. *Let* $h(x, \theta)$ *be continuous in* $\theta \in \Theta$ *for each* x *and measurable in* x *for each* θ *where* Θ *is a locally compact separable metric space. Let* $h(\cdot, \cdot)$ *be uniformly bounded and let* $\mathcal{F} := \{h(\cdot, \theta) : \theta \in \Theta\}$ *be a VC major class of functions. Then* $\mathcal{F}$ *is a uniform, thus universal, Glivenko-Cantelli class.*

Proof. Theorem 6 of [7] applies: sufficient bounds for the Koltchinskii-Pollard entropy of uniformly bounded VC major classes of functions are given in [3, Theorem 2.1(a), Corollary 5.8], and sufficient measurability of the class $\mathcal{F}$ follows from the continuity in θ and the assumptions on Θ. ☐

For the t location-scatter functionals in Sections 4 and 5, the notions of VC major class, and local Glivenko-Cantelli class as in Theorem 2(iii), will be applicable. But as shown by Kent, Tyler and Vardi [16], to be recalled after Theorem 12(iii), some parts of the development work only for t functionals, rather than for functions ρ satisfying general properties such as convexity.

2. Equivariance for location and scatter

Notions of "location" and "scale" or multidimensional "scatter" functional will be defined along with equivariance, as follows.

Definitions. Let $Q \mapsto \mu(Q) \in \mathbb{R}^d$, resp. $\Sigma(Q) \in \mathcal{N}_d$, be a functional defined on a set $\mathcal{D}$ of laws Q on $\mathbb{R}^d$. Then μ (resp. Σ) is called an *affinely equivariant location* (resp. *scatter*) *functional* iff for any nonsingular $d \times d$ matrix A and $v \in \mathbb{R}^d$, with $f(x) := Ax + v$, and any law $Q \in \mathcal{D}$, the image measure $P := Q \circ f^{-1} \in \mathcal{D}$ also, with $\mu(P) = A\mu(Q) + v$ or, respectively, $\Sigma(P) = A\Sigma(Q)A'$. For $d = 1$, $\sigma(\cdot)$ with

$0 \le \sigma < \infty$ will be called an *affinely equivariant scale functional* iff σ^2 satisfies the definition of affinely equivariant scatter functional.

Well-known examples of affinely equivariant location and scale functionals (for $d = 1$), defined for all laws, are the median and MAD (median absolute deviation), where for a real random variable X with median m, the MAD of X or its distribution is defined as the median of $|X - m|$.

Call a location functional $\mu(\cdot)$ or a scatter functional $\Sigma(\cdot)$ *singularly affine equivariant* if in the definition of affine equivariance A can be any matrix, possibly singular. If a functional is defined on all laws, affinely equivariant, and weakly continuous, then it must be singularly affine equivariant. The classical sample mean and covariance are defined for all P_n and singularly affine equivariant. It turns out that in dimension $d \ge 2$, there are essentially no other singularly affine equivariant location or scatter functionals defined for all P_n, and so weak continuity at all laws is not possible. First the known fact for location will be recalled, then an at least partially known fact for scatter will be stated and proved.

Let X be a $d \times n$ data matrix whose jth column is $X_j \in \mathbb{R}^d$. Let X^i be the ith row of X. Let $\mathbf{1}_n$ be the $n \times 1$ vector with all components 1. Let $\overline{X} = \int x dP_n$ be the sample mean vector in $\mathbb{R}^d$, so that $X - \overline{X}\mathbf{1}_n'$ is the centered data matrix. Note that P_n, and thus $\overline{X}$ and $\Sigma(X)$, are preserved by any permutation of the columns of X.

Theorem 4. (a) *If $\mu(\cdot)$ is a singularly affine equivariant location functional defined for all P_n on $\mathbb{R}^d$ for $d \ge 2$ and a fixed n, then $\mu(P_n) \equiv \overline{X}$.*
(b) *If in addition $\mu(\cdot)$ is defined for all n and all P_n on $\mathbb{R}^d$, then as n varies, $\mu(\cdot)$ is not weakly continuous. Thus, there is no affinely equivariant, weakly continuous location functional defined on all laws on $\mathbb{R}^d$ for $d \ge 2$.*

Proof. Part (a) follows from work of Obenchain [18, Lemma 1] and permutation invariance, as noted e.g. by Rousseeuw [19]. Then (b) follows directly, for $x_1 = n$, $x_2 = \cdots = x_n = 0$, $n \to \infty$. $\square$

Next is a related fact about scatter functionals. Davies [1, p. 1879] made a statement closely related to part (b), strong but not quite in the same generality, and very briefly suggested a proof. I don't know a reference for part (a), or an explicit one for (b), so a proof will be given.

Theorem 5. (a) *Let $\Sigma(\cdot)$ be a singularly affine equivariant scatter functional defined on all empirical measures P_n on $\mathbb{R}^d$ for $d \ge 2$ and some fixed $n \ge 2$. Write $\Sigma(X) := \Sigma(P_n)$. Then there is a constant $c_n \ge 0$, depending on $\Sigma(\cdot)$, such that for any X, $\Sigma(X - \overline{X}\mathbf{1}_n') = c_n(X - \overline{X}\mathbf{1}_n')(X - \overline{X}\mathbf{1}_n')'$. In other words, applied to centered data matrices, Σ is proportional to the sample covariance matrix.*
(b) *If $\Sigma(\cdot)$ is an affinely equivariant scatter functional defined for all n and P_n on $\mathbb{R}^d$ for $d \ge 2$, weakly continuous as a function of P_n, then $\Sigma \equiv 0$.*

Proof. (a) We have $\Sigma(BX) = B\Sigma(X)B'$ for any $d \times d$ matrix B. For any $U, V \in \mathbb{R}^n$ let $X^1 = U'$, $X^2 = V'$, and $(U, V) := \Sigma_{12}(X)$. Then $(\cdot,\cdot)$ is well-defined, letting $B_{11} = B_{22} = 1$ and $B_{ij} = 0$ otherwise. It will be shown that $(\cdot,\cdot)$ is a semi-inner product. We have $(U, V) \equiv (V, U)$ via B with $B_{12} = B_{21} = 1$ and $B_{ij} = 0$ otherwise, since Σ is symmetric. For $B_{11} = B_{21} = 1$ and $B_{ij} = 0$ otherwise we get for any $U \in \mathbb{R}^n$ that

$$(U, U) = \Sigma_{12}(BX) = (B\Sigma(X)B')_{12} = \Sigma_{11}(X) \ge 0. \tag{4}$$

For constants a and b, $(aU, bV) \equiv ab(U,V)$ follows for $B_{11} = a$, $B_{22} = b$, and $B_{ij} = 0$ otherwise. It remains to prove biadditivity $(U, V + W) \equiv (U,V) + (U,W)$. For $d \geq 3$ this is easy, letting $X^3 = W$, $B_{11} = B_{22} = B_{23} = 1$, and $B_{ij} = 0$ otherwise. For $d = 2$, we first get $(U + V, V) = (U,V) + (V,V)$ from $B = \binom{1\ 1}{0\ 1}$. Symmetrically, $(U, U+V) = (U,U) + (U,V)$. Then from $B = \binom{1\ 1}{1\ 1}$ we get

$$(U+V, U+V) \ = \ (U,U) + 2(U,V) + (V,V). \tag{5}$$

Letting $\|W\|^2 := (W,W)$ for any $W \in \mathbb{R}^n$ we get the parallelogram law $\|U + V\|^2 + \|U - V\|^2 \equiv 2\|U\|^2 + 2\|V\|^2$. (But $\|\cdot\|$ has not yet been shown to be a norm.) Applying this repeatedly we get for any W, Y, and $Z \in \mathbb{R}^n$ that

$$\|W+Y+Z\|^2 - \|W-Y-Z\|^2 = \|W+Y\|^2 - \|W-Y\|^2 + \|W+Z\|^2 - \|W-Z\|^2,$$

letting first $U = W+Y$, $V = Z$, then $U = W-Z$, $V = Y$, then $U = W$, $V = Z$, and lastly $U = W$, $V = Y$. Applying (5) gives $(W, Y+Z) \equiv (W,Y)+(W,Z)$, the desired biadditivity. So $(\cdot,\cdot)$ is indeed a semi-inner product, i.e. there is a $C(n) \in \mathcal{N}_n$ such that $(U,V) \equiv U'C(n)V$. By permutation invariance, there are numbers $a_n \geq 0$ and b_n such that $C(n)_{ii} = a_n$ for all $i = 1, \ldots, n$ and $C(n)_{ij} = b_n$ for all $i \neq j$.

Let $c_n := a_n - b_n$ and let $e_i \in \mathbb{R}^n$ be the ith standard unit vector. For each $y \in \mathbb{R}^n$ let $y = \sum_{i=1}^n y_i e_i$ and $\overline{y} := \frac{1}{n}\sum_{i=1}^n y_i$. Then for any $z \in \mathbb{R}^n$,

$$(y - \overline{y}\mathbf{1}_n, z - \overline{z}\mathbf{1}_n) \ = \ \sum_{i,j=1}^n C(n)_{ij}(y_i - \overline{y})(z_j - \overline{z}) \ = \ c_n(y - \overline{y}\mathbf{1}_n)'(z - \overline{z}\mathbf{1}_n).$$

For $1 \leq j \leq k \leq d$, let $B_{ir} := \delta_{r\pi(i)}$ for a function π from $\{1, 2, \ldots, d\}$ into itself with $\pi(1) = j$ and $\pi(2) = k$. Then $(BX)^1 = X^j$ and $(BX)^2 = X^k$. Thus $(X^j, X^k) = \Sigma_{12}(BX) = \Sigma_{jk}(X)$, recalling (4) if $j = k$.

Let $\overline{X} \in \mathbb{R}^d$ have ith component $\overline{X}^i$. Then

$$\Sigma_{jk}(X - \overline{X}\mathbf{1}_n') \ = \ (X^j - \overline{X}^j\mathbf{1}_n, X^k - \overline{X}^k\mathbf{1}_n) \ = \ c_n(X^j - \overline{X}^j\mathbf{1}_n)'(X^k - \overline{X}^k\mathbf{1}_n),$$

where $c_n \geq 0$ is seen when $j = k$ and the coefficient of c_n is strictly positive, as it can be since $n \geq 2$. Thus part (a) is proved.

For part (b), consider empirical measures $P_n = P_{mn}$, so that each X_j in P_n is repeated m times in P_{mn}. Since the $\overline{X}$'s and Σs for P_n and P_{mn} must be the same, we get that $c_{mn} = c_n/m$ which likewise equals c_m/n. Thus there is a constant c_1 such that $c_n = c_1/n$ for all n.

Let $X_{11} := -X_{12} := \sqrt{n}$, let $X_{ij} = 0$ for all other i, j and let $n \to \infty$. Then $\overline{X} \equiv 0$, $P_n \to \delta_0$ weakly, and $\Sigma(\delta_0)$ is the 0 matrix by singular affine equivariance with $B = 0$, but $\Sigma(P_n)$ don't converge to 0 unless $c_1 = 0$ and so $c_n = 0$ for all n, proving (b). $\square$

So, for $d \geq 2$, affinely equivariant location and non-zero scatter functionals, weakly continuous on their domains, can't be defined on all laws. They can be defined on weakly dense and open domains, as will be seen in Theorem 12, on which they can have good differentiability properties, as seen in Section 5.

3. Multivariate scatter

This section treats pure scatter in $\mathbb{R}^d$, with $\Theta = \mathcal{P}_d$. Results of Kent and Tyler [15] for finite samples, to be recalled, are extended to general laws on $\mathbb{R}^d$ in [6, Section 3].

For $A \in \mathcal{P}_d$ and a function ρ from $[0, \infty)$ into itself, consider the function

$$L(y, A) \;:=\; \frac{1}{2}\log\det A + \rho(y'A^{-1}y), \quad y \in \mathbb{R}^d. \tag{6}$$

For adjustment, let

$$h(y, A) \;:=\; L(y, A) - L(y, I) \tag{7}$$

where I is the identity matrix. Then

$$Qh(A) \;=\; \frac{1}{2}\log\det A + \int \rho(y'A^{-1}y) - \rho(y'y)\, dQ(y) \tag{8}$$

if the integral is defined. We have the following, shown for $Q = Q_n$ an empirical measure in [15, (1.3)] and for general Q in [6, Section 3]. Here (9) is a redescending condition. A symmetric $d \times d$ matrix A will be parameterized by the entries A_{ij} for $1 \le i \le j \le d$. Thus in taking a partial derivative of a function $f(A)$ with respect to an entry A_{ij}, $A_{ji} \equiv A_{ij}$ will vary while A_{kl} will remain fixed except for $(k, l) = (i, j)$ or (j, i).

Proposition 6. *Let ρ be continuous from $[0, \infty)$ into itself and have a bounded continuous derivative, where $\rho'(0) \;:=\; \rho'(0+) \;:=\; \lim_{x\downarrow 0}[\rho(x) - \rho(0)]/x$. Let $0 \le u(x) \;:=\; 2\rho'(x)$ for $x \ge 0$. Assume that*

$$\sup_{0 \le x < \infty} xu(x) < \infty. \tag{9}$$

Then for each law Q on $\mathbb{R}^d$, Qh in (8) is well defined and is a C^1 function of the entries of A. Here Qh has a critical point at $A = B$ if and only if

$$B \;=\; \int u(y'B^{-1}y)yy'dQ(y). \tag{10}$$

The following, proved in [6, Section 3], extends to any law Q the uniqueness part of [15, Theorem 2.2].

Proposition 7. *Under the hypotheses of Proposition 6, if in addition $u(\cdot)$ is non-increasing and $s \mapsto su(s)$ is strictly increasing on $[0, \infty)$, then for any law Q on $\mathbb{R}^d$, Qh has at most one critical point $A \in \mathcal{P}_d$.*

A sufficient condition for existence of a pure scatter M-functional $A(Q)$ will include the following assumption from [15, (2.4)]. Given a function $u(\cdot)$ as in Proposition 7, let $a_0 \;:=\; a_0(u(\cdot)) \;:=\; \sup_{s>0} su(s)$. Since $s \mapsto su(s)$ is increasing, it follows that

$$su(s) \uparrow a_0 \quad \text{as} \quad s \uparrow +\infty. \tag{11}$$

Kent and Tyler [15] gave the following condition for empirical measures.

Definition. Given $a_0 \;:=\; a(0) > 0$, let $\mathcal{U}_{d,a(0)}$ denote the set of all laws Q on $\mathbb{R}^d$ such that for every proper linear subspace H of $\mathbb{R}^d$, of dimension $q \le d-1$, we have $Q(H) < 1 - (d - q)/a_0$.

Note that $\mathcal{U}_{d,a(0)}$ is weakly open and dense and contains all laws with densities. If $Q \in \mathcal{U}_{d,a(0)}$, then $Q(\{0\}) < 1 - (d/a_0)$, which is impossible if $a_0 \le d$. So in the next theorem we assume $a_0 > d$. In part (b), the existence of a unique $B(Q_n)$ minimizing $Q_n h$ for an empirical $Q_n \in \mathcal{U}_{d,a(0)}$ was proved in [15, Theorems 2.1 and 2.2]. For a general $Q \in \mathcal{U}_{d,a(0)}$ it's proved in [6, Section 3]; one lemma useful in the proof is proved here.

Theorem 8. *Under the assumptions of Propositions* 6 *and* 7*, for* $a(0) = a_0$ *as in* (11)*,*
(a) *If* $Q \notin \mathcal{U}_{d,a(0)}$*, then* Qh *has no critical points.*
(b) *If* $a_0 > d$ *and* $Q \in \mathcal{U}_{d,a(0)}$*, then* Qh *attains its minimum at a unique* $B = B(Q) \in \mathcal{P}_d$ *and has no other critical points.*

A proof of the theorem uses a fact about probabilities of proper subspaces or hyperplanes. A related statement is Lemma 5.1 of Dümbgen and Tyler [10].

Lemma 9. *Let* V *be a real vector space with a* σ*-algebra* $\mathcal{B}$ *for which all finite-dimensional hyperplanes* $H = x + T := \{x + u : u \in T\}$ *for finite-dimensional vector subspaces* T *are measurable. Let* Q *be a probability measure on* $\mathcal{B}$ *and let* $\mathcal{H}_j$ *be the collection of all* j*-dimensional hyperplanes in* V*. Then for each* $j = 0, 1, 2, \ldots,$ *for any infinite sequence* $\{C_i\}$ *of distinct hyperplanes in* $\mathcal{H}_j$ *such that* $Q(C_i)$ *converges, its limit must be* $Q(F)$ *for some hyperplane* F *of dimension less than* j *such that* $F \subset C_i$ *for infinitely many* i*. In particular,* $Q(C_i)$ *cannot be strictly increasing. The same is true for vector subspaces in place of hyperplanes.*

Proof. Hyperplanes of dimension 0 are singletons $\{x\}$. The empty set $\emptyset$ will be considered as a hyperplane of dimension -1. Let $W_{-1} := \emptyset$. *Claim 1*: For each $j = 0, 1, \ldots,$ there exists a finite or countable sequence $\{V_{ji}\} \subset \mathcal{H}_j$ such that for $W_j := W_{j-1} \cup \bigcup_i V_{ji}$, $Q(V \setminus W_j) = 0$ for all $V \in \mathcal{H}_j$. Let $V_{0i} = \{x_i\}$ for some unique i if and only if $Q(\{x_i\}) > 0$. The set of such x_i is clearly countable. Let $W_0 := \cup_i V_{0i} = \{x \in V : Q(\{x\}) > 0\}$. Clearly, for any $x \in V$, $Q(\{x\} \setminus W_0) = 0$. Recursively, for $j \geq 1$, assuming W_{j-1} has the given properties, suppose for $r = 1, 2$, $H_r \in \mathcal{H}_j$ and $Q(H_r \setminus W_{j-1}) > 0$. If $H_1 \neq H_2$, then $H_1 \cap H_2$ is a hyperplane of dimension at most $j-1$, so $Q(H_1 \cap H_2 \setminus W_{j-1}) = 0$ and the sets $H_r \setminus W_{j-1}$ are disjoint up to sets with $Q = 0$. Thus there are at most countably many different $H_r \in \mathcal{H}_j$ with $Q(H_r \setminus W_{j-1}) > 0$. Let $V_{jr} := H_r$ for such H_r and set $W_j := W_{j-1} \cup \bigcup_r V_{jr}$. It's then clear that for any $H \in \mathcal{H}_j$, $Q(H \setminus W_j) = 0$, so the recursion can continue and Claim 1 is proved.

Claim 2 is that if C is any hyperplane of dimension j or larger, and $s = 0, 1, \ldots, j$, then for each r, either $C \supset V_{sr}$ or $Q(C \cap (V_{sr} \setminus W_{s-1})) = 0$. If C doesn't include V_{sr}, then $C \cap V_{sr}$ is a hyperplane of dimension $\leq s - 1$, and so included in W_{s-1} up to a set with $Q = 0$, so Claim 2 follows.

Now, given distinct $C_i \in \mathcal{H}_j$ with $Q(C_i)$ converging, let B be a hyperplane of largest possible dimension b included in C_i for infinitely many i. Then $b < j$. Taking a subsequence, we can assume that $B \subset C_i$ for all i. *Claim 3* is that then $Q(C_i \setminus B) \to 0$ as $i \to \infty$. For any $s = 0, 1, \ldots, j-1$, and each r, by Claim 2, if $C_i \supset V_{sr}$ for infinitely many i, then $V_{sr} \subset B$, since otherwise C_i includes the smallest hyperplane including V_{sr} and B, which has dimension larger than b, a contradiction. So $\lim_{i \to \infty} Q((C_i \setminus B) \cap (V_{sr} \setminus W_{s-1})) = 0$ for each $s < j$ and r. It follows by induction on s that $Q(C_i \cap W_s \setminus B) \to 0$ as $i \to \infty$ for $s = 0, 1, \ldots, j-1$.

By the proof of Claim 1, the sets $C_i \setminus W_{j-1}$ are disjoint up to sets with $Q = 0$, so Claim 3 follows, and so the statement of the lemma for hyperplanes. The proof for vector subspaces is parallel and easier. The fact that $Q(C_i)$ cannot be strictly increasing then clearly follows, as a subsequence would also be strictly increasing. So the lemma is proved. $\square$

Dümbgen and Tyler [10], Lemma 5.1 show that $\sup\{Q(V) : V \in \mathcal{H}_j\}$ is attained for each Q and j and is weakly upper semicontinuous in Q.

4. Location and scatter t functionals

As Kent and Tyler [15, Section 3] and Kent, Tyler and Vardi [16] showed, (t) location-scatter estimation in $\mathbb{R}^d$ can be reduced to pure scatter estimation in $\mathbb{R}^{d+1}$, beginning with the following.

Proposition 10. (i) *For any $d = 1, 2, \ldots,$ there is a 1-1 correspondence, C^∞ in either direction, between matrices $A \in \mathcal{P}_{d+1}$ and triples (Σ, μ, γ) where $\Sigma \in \mathcal{P}_d$, $\mu \in \mathbb{R}^d$, and $\gamma > 0$, given by*

$$A = A(\Sigma, \mu, \gamma) = \gamma \begin{bmatrix} \Sigma + \mu\mu' & \mu \\ \mu' & 1 \end{bmatrix}. \tag{12}$$

The same holds for $A \in \mathcal{P}_{d+1}$ with $\gamma = A_{d+1,d+1} = 1$ and pairs $(\mu, \Sigma) \in \mathbb{R}^d \times \mathcal{P}^d$.
(ii) *If* (12) *holds, then for any $y \in \mathbb{R}^d$ (a column vector),*

$$(y', 1)A^{-1}(y', 1)' \;=\; \gamma^{-1}\left(1 + (y-\mu)'\Sigma^{-1}(y-\mu)\right). \tag{13}$$

For M-estimation of location and scatter in $\mathbb{R}^d$, we will have a function $\rho : [0, \infty) \mapsto [0, \infty)$ as in the previous section. The parameter space is now the set of pairs (μ, Σ) for $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}_d$, and we have a multivariate ρ function

$$\rho(y, (\mu, \Sigma)) \;:=\; \frac{1}{2}\log\det\Sigma + \rho((y-\mu)'\Sigma^{-1}(y-\mu)).$$

For any $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}_d$ let $A_0 \;:=\; A_0(\mu, \Sigma) \;:=\; A(\Sigma, \mu, 1) \in \mathcal{P}_{d+1}$ by (12) with $\gamma = 1$, noting that $\det A_0 = \det\Sigma$. Now ρ can be adjusted, in light of (9) and (13), by defining

$$h(y, (\mu, \Sigma)) \;:=\; \rho(y, (\mu, \Sigma)) - \rho(y'y). \tag{14}$$

Laws P on $\mathbb{R}^d$ correspond to laws $Q \;:=\; P \circ T_1^{-1}$ on $\mathbb{R}^{d+1}$ concentrated in $\{y : \; y_{d+1} = 1\}$, where $T_1(y) \;:=\; (y', 1)' \in \mathbb{R}^{d+1}$, $y \in \mathbb{R}^d$. We will need a hypothesis on P corresponding to $Q \in \mathcal{U}_{d+1,a(0)}$. Kent and Tyler [15] gave these conditions for empirical measures.

Definition. For any $a_0 > 0$ let $\mathcal{V}_{d,a(0)}$ be the set of all laws P on $\mathbb{R}^d$ such that $P(J) < 1 - (d-q)/a_0$ for every affine hyperplane J of dimension $q < d$.

The next fact is rather easy to prove. Here $a > d + 1$ avoids the contradictory $Q(\{0\}) < 0$.

Proposition 11. *If P is a law on $\mathbb{R}^d$, $a > d + 1$, and $Q \;:=\; P \circ T_1^{-1}$ on $\mathbb{R}^{d+1}$, then $P \in \mathcal{V}_{d,a}$ if and only if $Q \in \mathcal{U}_{d+1,a}$.*

A family of ρ functions for which $\gamma = 1$ automatically, as noted by Kent and Tyler [15, (1.5), (1.6), Section 4], is given by elliptically symmetric multivariate t densities with ν degrees of freedom as follows: for $0 < \nu < \infty$ and $0 \le s < \infty$ let

$$\rho_\nu(s) \;:=\; \rho_{\nu,d}(s) \;:=\; \frac{\nu + d}{2}\log\left(\frac{\nu + s}{\nu}\right). \tag{15}$$

For this ρ, u is $u_\nu(s) \;:=\; u_{\nu,d}(s) \;:=\; (\nu + d)/(\nu + s)$, which is decreasing, and $s \mapsto su_{\nu,d}(s)$ is strictly increasing and bounded, i.e. (9) holds, with supremum and limit at $+\infty$ equal to $a_{0,\nu} \;:=\; a_0(u_\nu(\cdot)) = \nu + d$.

The following fact is in part given by Kent and Tyler [15] and further by Kent, Tyler and Vardi [16], for empirical measures; equation (16) was not found explicitly in either. Here a proof will be given for any $P \in \mathcal{V}_{d,\nu+d}$, assuming Theorem 8 and Propositions 6 and 10.

Theorem 12. *For any $d = 1, 2, \ldots$, $\nu > 1$, law P on $\mathbb{R}^d$, and $Q = P \circ T_1^{-1}$ on $\mathbb{R}^{d+1}$, letting $\nu' := \nu - 1$, assuming $P \in \mathcal{V}_{d,\nu+d}$ in parts (a) through (e),*
(a) *For $A \in \mathcal{P}_{d+1}$, $A \mapsto Qh(A)$ defined by (8) for $\rho = \rho_{\nu',d+1}$ has a unique critical point $A(\nu') := A_{\nu'}(Q)$ which is an absolute minimum;*
(b) *$A(\nu')_{d+1,d+1} = \int u_{\nu',d+1}(y'A(\nu')^{-1}y)dQ(y) = 1$;*
(c) *For any $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}_d$ let $A = A(\Sigma, \mu, 1) \in \mathcal{P}_{d+1}$ in (12). Then for any $y \in \mathbb{R}^d$ and $z := (y', 1)'$, we have*

$$u_{\nu',d+1}(z'A^{-1}z) \equiv u_{\nu,d}((y-\mu)'\Sigma^{-1}(y-\mu)). \tag{16}$$

In particular, this holds for $A = A(\nu')$ and its corresponding $\mu = \mu_\nu \in \mathbb{R}^d$ and $\Sigma = \Sigma_\nu \in \mathcal{P}_d$.
(d)

$$\int u_{\nu,d}((y-\mu_\nu)'\Sigma_\nu^{-1}(y-\mu_\nu))dP(y) = 1. \tag{17}$$

(e) *For h defined by (14) with $\rho = \rho_{\nu,d}$, (μ_ν, Σ_ν) is the M-functional θ_1 for P.*
(f) *If, on the other hand, $P \notin \mathcal{V}_{d,\nu+d}$, then $(\mu, \Sigma) \mapsto Ph(\mu, \Sigma)$ for h as in (e) has no critical points.*

Proof. (a): Theorem 8(b) applies since $u_{\nu',d+1}$ satisfies its hypotheses, with $a_0(u_{\nu',d+1}) = \nu' + d + 1 = \nu + d > d + 1$.
(b): By (10), multiplying by $A(\nu')^{-1}$ and taking the trace gives

$$d + 1 = \int u_{\nu',d+1}\left(z'A(\nu')^{-1}z\right) z'A(\nu')^{-1}z dQ(z).$$

We also have, since $z_{d+1} \equiv 1$, that $A(\nu')_{d+1,d+1} = \int u_{\nu',d+1}(z'A(\nu')^{-1}z)dQ(z)$. For any $s \geq 0$, we have $su_{\nu',d+1}(s) + \nu' u_{\nu',d+1}(s) = \nu + d$. Combining gives

$$d + 1 = \nu + d - \nu' \int u_{\nu',d+1}\left(z'A(\nu')^{-1}z\right) dQ(z),$$

and (b) follows.
(c): We can just apply (13) with $\gamma = 1$, and for $A = A(\nu')$, part (b).
(d): This follows from (b) and (c).
(e): By Proposition 10, for $\gamma = 1$ fixed, the relation (12) is a homeomorphism between $\{A \in \mathcal{P}_{d+1} : A_{d+1,d+1} = 1\}$ and $\{(\mu, \Sigma) : \mu \in \mathbb{R}^d, \Sigma \in \mathcal{P}_d\}$. So this also follows from Theorem 8.
(f): We have $\nu + d > d + 1$, so $Q \notin \mathcal{U}_{d+1,\nu+d}$ by Proposition 11. By Theorem 8(a), Qh defined by (8) for $\rho = \rho_{\nu',d+1}$ has no critical point A. Suppose Ph has a critical point (μ, Σ) for $\rho = \rho_{\nu,d}$. Let $A := A(\Sigma, \mu, 1) \in \mathcal{P}_{d+1}$. By an affine transformation we can assume $\mu = 0$ and $\Sigma = I_d$, the $d \times d$ identity matrix, so $A = I_{d+1}$. Equations for $\Sigma = I_d$ to be a critical point can be written in the form $\partial/\partial(\Sigma^{-1})_{ij} = 0$, $1 \leq i \leq j \leq d$. By (16) it follows easily that equation (10) holds for $B = A$ and $u = u_{\nu',d+1}$ with the possible exception of the $(d+1, d+1)$ entry. Summing the equations for the diagonal (i, i) entries for $i = 1, \ldots, d$, it follows that the $(d+1, d+1)$ equation and so (10) holds. By Proposition 6, we get that A is a critical point of the given Qh, a contradiction. $\square$

Kent, Tyler and Vardi [16, Theorem 3.1] show that if $u(s) \geq 0$, $u(0) < +\infty$, $u(\cdot)$ is continuous and nonincreasing for $s \geq 0$, and $su(s)$ is nondecreasing for $s \geq 0$, up to $a_0 > d$, and if (17) holds with u in place of $u_{\nu,d}$ at each critical point (μ, Σ) of

Qh, then u must be of the form $u(s) = u_{\nu,d}(s) = (\nu + d)/(\nu + s)$ for some $\nu > 0$. Thus, the method of relating pure scatter functionals in $\mathbb{R}^{d+1}$ to location-scatter functionals in $\mathbb{R}^d$ given by Theorem 12 for t functionals defined by functions $u_{\nu,d}$ does not extend directly to any other functions u.

When $d = 1$, $P \in \mathcal{V}_{1,\nu+1}$ requires that $P(\{x\}) < \nu/(1+\nu)$ for each point x. Then Σ reduces to a number σ^2 with $\sigma > 0$. If $\nu > 1$, and $P \notin \mathcal{V}_{1,\nu+1}$, then for some unique x, $P(\{x\}) \geq \nu/(\nu+1)$. One can extend (μ_ν, σ_ν) by setting $\mu_\nu(P) := x$ and $\sigma_\nu(P) := 0$, with (μ_ν, σ_ν) then being weakly continuous at all P [6, Section 6].

The t_ν functionals (μ_ν, Σ_ν) defined in this section can't have a weakly continuous extension to all laws for $d \geq 2$, because such an extension of μ_ν would give a weakly continuous affinely equivariant location functional defined for all laws, which is impossible by Theorem 4(b). Here is an example showing that for $d = 2$ and empirical laws with $n = 6$, invariant under interchanging x and $-x$, and/or y and $-y$, so that an affinely equivariant μ must be 0, there is no continuous extension of the scatter matrix Σ_ν to laws concentrated in lines. For $k = 1, 2, \ldots$ let

$$\begin{aligned} P^{(k)} &:= \frac{1}{6}\left[\delta_{(-1,-1/k)} + \delta_{(-1,1/k)} + \delta_{(1,-1/k)} + \delta_{(1,1/k)}\right] + \frac{1}{3}\delta_{(0,0)}, \\ Q^{(k)} &:= \frac{1}{6}\left[2\delta_{(-1,0)} + \delta_{(0,-1/k)} + \delta_{(0,1/k)} + 2\delta_{(1,0)}\right]. \end{aligned}$$

Then for each $\nu > 1$, all members of both sequences have mass $\leq 1/3 < \nu/(\nu+2)$ at each point and mass $\leq 2/3 < (\nu+1)/(\nu+2)$ on each line, so are in $\mathcal{U}_{2,\nu+2}$ and the functionals μ_ν, Σ_ν are defined for them. By the symmetries $x \leftrightarrow -x$ and $y \leftrightarrow -y$, $\mu_\nu \equiv 0$ and Σ_ν is diagonal on all these laws. Both sequences converge to the limit $P = \frac{1}{3}\left[\delta_{(-1,0)} + \delta_{(0,0)} + \delta_{(1,0)}\right]$, which is concentrated in a line and so is not in $\mathcal{U}_{2,\nu+2}$ for any ν. $\Sigma_\nu(P^{(k)})$ converges to $\left(\begin{smallmatrix} a(\nu) & 0 \\ 0 & 0 \end{smallmatrix}\right)$ but $\Sigma_\nu(Q^{(k)})$ converges to $\left(\begin{smallmatrix} b(\nu) & 0 \\ 0 & 0 \end{smallmatrix}\right)$ where $a(\nu) := 2(1-\nu^{-1})/3 \neq b(\nu) := (2+\nu^{-1})/3$. We also have $\Sigma_\nu(Q^{(1)}) = \left(\begin{smallmatrix} b(\nu) & 0 \\ 0 & c(\nu) \end{smallmatrix}\right)$ with $c(\nu) = \frac{1}{3}(1-\nu^{-1})$, so that, in contrast to Theorem 5(a), Σ_ν is not proportional to the covariance matrix $\left(\begin{smallmatrix} 2/3 & 0 \\ 0 & 1/3 \end{smallmatrix}\right)$ for any $\nu < \infty$, but Σ_ν converges to the covariance as $\nu \to +\infty$, as is not surprising since the t_ν distribution converges to a normal one.

5. Differentiability of t functionals

Let (S, e) be any separable metric space, in our case $\mathbb{R}^d$ with its usual Euclidean metric. Recall the space $BL(S, e)$ of all bounded Lipschitz real functions on S, with its norm $\|f\|_{BL}$. The dual Banach space $BL^*(S, e)$ has the dual norm $\|\phi\|^*_{BL}$, which metrizes the weak topology on probability measures [5, Theorem 11.3.3].

Let V be an open set in a Euclidean space $\mathbb{R}^d$. For $k = 1, 2, \ldots$, let $C^k_b(V)$ be the space of all real-valued functions f on V such that all partial derivatives $D^p f$, for $D^p := \partial^{[p]}/\partial x_1^{p_1} \cdots \partial x_d^{p_d}$ and $0 \leq [p] := p_1 + \cdots + p_d \leq k$, are continuous and bounded on V. On $C^k_b(V)$ we have the norm

$$(18) \qquad \|f\|_{k,V} := \sum_{0 \leq [p] \leq k} \|D^p f\|_{\sup,V}, \quad \text{where} \quad \|g\|_{\sup,V} := \sup_{x \in V} |g(x)|.$$

Then $(C^k_b(V), \|\cdot\|_{k,V})$ is a Banach space. For $k = 1$ and V convex in $\mathbb{R}^d$ it's straightforward that $C^1_b(V)$ is a subspace of $BL(V, e)$, with the same norm for $d = 1$ and an equivalent one if $d > 1$.

Substituting $\rho_{\nu,d}$ from (15) into (6) gives for $y \in \mathbb{R}^d$ and $A \in \mathcal{P}_d$,

$$L_{\nu,d}(y, A) \ := \ \frac{1}{2} \log \det A + \frac{\nu + d}{2} \log \left[1 + \nu^{-1} y' A^{-1} y\right],$$

so that in (7) we get

$$h_\nu(y, A) \ := \ h_{\nu,d}(y, A) \ := \ L_{\nu,d}(y, A) - L_{\nu,d}(y, I).$$

Differentiating with respect to entries C_{ij} where $C = A^{-1}$, and recalling $u_{\nu,d}(s) \equiv (\nu + d)/(\nu + s)$, we get as shown in [6, Section 5]

$$\frac{\partial h_{\nu,d}(y, A)}{\partial C_{ij}} = \frac{\partial L_{\nu,d}(y, A)}{\partial C_{ij}} = -\frac{A_{ij}}{1 + \delta_{ij}} + \frac{(\nu + d) y_i y_j}{(1 + \delta_{ij})(\nu + y' C y)}. \tag{19}$$

For $0 < \delta < 1$ and $d = 1, 2, \ldots$, let

$$\mathcal{W}_\delta \ := \ \mathcal{W}_{\delta,d} \ := \ \{A \in \mathcal{P}_d : \ \max(\|A\|, \|A^{-1}\|) < 1/\delta\}.$$

The following is proved in [6, Section 5].

Lemma 13. *For any $\delta \in (0, 1)$ let $U \ := \ U_\delta \ := \ \mathbb{R}^d \times \mathcal{W}_{\delta,d}$. Let $A \in \mathcal{P}_d$ be parameterized by the entries C_{kl} of $C = A^{-1}$. For any $\nu \geq 1$, the functions $\partial L_{\nu,d}/\partial C_{kl}$ in* (19) *are in $C_b^j(U_\delta)$ for all $j = 1, 2, \ldots$.*

To treat t functionals of location and scatter in any dimension p we will need functionals of pure scatter in dimension $p + 1$, so in the following lemma we only need dimension $d \geq 2$. The next lemma, proved in [6, Section 5], helps to show differentiability of t functionals via implicit function theorems, as it implies that the derivative of the gradient (the Hessian) of Qh is non-singular at a critical point. Let $T(d) \ := \ \{(i, j) : \ 1 \leq i \leq j \leq d\}$.

Lemma 14. *For each $\nu > 0$, $d = 2, 3, \ldots$, and $Q \in \mathcal{U}_{d,\nu+d}$, let $A(\nu) = A_\nu(Q) \in \mathcal{P}_d$ be the unique critical point of $Qh(\cdot)$ defined by (8) for $\rho = \rho_{\nu,d}$ defined by (15). For $C = A^{-1}$, the Hessian matrix $\partial^2 Qh(A)/\partial C_{ij} \partial C_{kl}$ with rows indexed by $(i, j) \in T(d)$ and columns by $(k, l) \in T(d)$ is positive definite at $A = A(\nu)$.*

For any $\nu > 0$ and $A \in \mathcal{P}_d$, let $L_{i,j,\nu}(x, A) \ := \ \partial L_{\nu,d}(x, A)/\partial C_{ij}$ from (19). Let $X \ := \ BL^*(\mathbb{R}^d, e)$ for the usual metric $e(s, t) \ := \ |s - t|$. Again, parameterize $A \in \mathcal{P}_d$ with inverse C by $\{C_{ij}\}_{1 \leq i \leq j \leq d} \in \mathbb{R}^{d(d+1)/2}$. Consider the open set $\Theta \ := \mathcal{P}_d \subset \mathbb{R}^{d(d+1)/2}$ and the function $F \ := \ F_\nu$ from $X \times \Theta$ into $\mathbb{R}^{d(d+1)/2}$ defined by

$$F(\phi, A) \ := \ \{\phi(L_{i,j,\nu}(\cdot, A))\}_{1 \leq i \leq j \leq d}. \tag{20}$$

Then F is well-defined because $L_{i,j,\nu}(\cdot, A)$ are all bounded and Lipschitz functions of x for each $A \in \Theta$; in fact, they are C^1 with bounded derivatives equal except possibly for signs to second partials of $L_{\nu,d}$ with respect to C_{ij}. The next fact, proved in [6, Section 5], uses some basic notions and facts from infinite-dimensional calculus, given in [2] and reviewed in the Appendix of [6].

Theorem 15. *Let $X := BL^*(\mathbb{R}^d, e)$. In parts* (a) *through* (c)*, let $\nu > 0$.*
(a) *The function $F = F_\nu$ is C^∞ (for Fréchet differentiability) from $X \times \Theta$ into $\mathbb{R}^{d(d+1)/2}$.*
(b) *Let $Q \in \mathcal{U}_{d,\nu+d}$, and take the corresponding $\phi_Q \in X$. At $A_\nu(Q)$, the $d(d+1)/2 \times d(d+1)/2$ matrix $\partial F(\phi_Q, A)/\partial C \ := \ \{\partial F(\phi_Q, A)/\partial C_{kl}\}_{1 \leq k \leq l \leq d}$ is invertible.*
(c) *The functional $Q \mapsto A_\nu(Q)$ is C^∞ for the BL^* norm on $\mathcal{U}_{d,\nu+d}$.*
(d) *For each $\nu > 1$, the functional $P \mapsto (\mu_\nu, \Sigma_\nu)(P)$ given by Theorems 8 and 12 is C^∞ on $\mathcal{V}_{d,\nu+d}$ for the BL^* norm.*

To prove asymptotic normality of $\sqrt{n}(T(P_n) - T(P))$ for $T = (\mu_\nu, \Sigma_\nu)$, the dual-bounded-Lipschitz norm $\|\cdot\|_{BL}^*$ is too strong for some heavy-tailed distributions. Giné and Zinn [12] proved that for $d = 1$, $\{f : \|f\|_{BL} \le 1\}$ is a P-Donsker class if and only if $\sum_{j=1}^{\infty} \Pr(j - 1 < |X| \le j)^{1/2} < \infty$ for X with distribution P. To define norms better suited to present purposes, for $\delta > 0$ and $r = 1, 2, \ldots$, let $\mathcal{F}_{\delta,r}$ be the set of all functions of y appearing in (19) and their partial derivatives with respect to C_{ij} through order r, for any $A \in \mathcal{W}_\delta$. Then each $\mathcal{F}_{\delta,r}$ is a uniformly bounded VC major class as in Theorem 3. Let $Y_{\delta,r}$ be the linear span of $\mathcal{F}_{\delta,r}$. Let $X_{\delta,r}$ be the set of all real-valued linear functionals ϕ on $Y_{\delta,r}$ for which $\|\phi\|_{\delta,r} := \sup\{|\phi(f)| : f \in \mathcal{F}_{\delta,r}\} < \infty$. For $A \in \mathcal{W}_{\delta,d}$ and $\phi \in X_{\delta,r}$, define $F(\phi, A)$ again by (20), which makes sense since each $L_{i,j,\nu}(\cdot, A) \in \mathcal{F}_{\delta,r}$ for any $r = 0, 1, 2, \ldots$ by definition.

The next two theorems are also proved in [6, Section 5]. Theorem 17 is a delta-method fact.

Theorem 16. *Let $0 < \delta < 1$. For any positive integers d and r, Theorem 15 holds for $X = X_{\delta,r+3}$ in place of $BL^*(\mathbb{R}^d, e)$, $\mathcal{W}_{\delta,d}$ in place of Θ, and C^r in place of C^∞ wherever it appears (parts (a), (c), and (d)).*

Theorem 17. (a) *For any $d = 2, 3, \ldots$ and $\nu > 0$, let $Q \in \mathcal{U}_{d,\nu+d}$. Then the empirical measures $Q_n \in \mathcal{U}_{d,\nu+d}$ with probability $\to 1$ as $n \to \infty$ and $\sqrt{n}(A_\nu(Q_n) - A_\nu(Q))$ converges in distribution to a normal distribution with mean 0 on $\mathbb{R}^{d(d+1)/2}$ if $A \in \mathcal{P}_d$ is parameterized by $\{A_{ij}\}_{1 \le i \le j \le d}$, or a different normal distribution for the parameterization by $\{A_{ij}^{-1}\}_{1 \le i \le j \le d}$ as above. The limit distributions can also be taken on $\mathbb{R}^{d^2}$, concentrated on symmetric matrices.*
(b) *Let $d = 1, 2, \ldots$ and $1 < \nu < \infty$. For any $P \in \mathcal{V}_{d,\nu+d}$, the empirical measures $P_n \in \mathcal{V}_{d,\nu+d}$ with probability $\to 1$ as $n \to \infty$ and the functionals μ_ν and Σ_ν are such that as $n \to \infty$, $\sqrt{n}\,[(\mu_\nu, \Sigma_\nu)(P_n) - (\mu_\nu, \Sigma_\nu)(P)]$ converges in distribution to some normal distribution with mean 0 on $\mathbb{R}^d \times \mathbb{R}^{d^2}$, whose marginal on $\mathbb{R}^{d^2}$ is concentrated on $d \times d$ symmetric matrices.*

Now, here is a statement on uniformity as P and Q vary, proved in [6, Section 5].

Proposition 18. *For any $\delta > 0$ and $M < \infty$, the rate of convergence to normality in Theorem 17*(a) *is uniform over the set $\mathcal{Q} := \mathcal{Q}(\delta, M)$ of all $Q \in \mathcal{U}_{d,\nu+d}$ such that $A_\nu(Q) \in \mathcal{W}_\delta$ and*

$$Q(\{z : |z| > M\}) \le (1 - \delta)/(\nu + d), \tag{21}$$

or in part (b), *over all $P \in \mathcal{V}_{d,\nu+d}$ such that $\Sigma_\nu(P) \in \mathcal{W}_\delta$ and (21) holds for P in place of Q.*

Acknowledgements

I thank Lutz Dümbgen and David Tyler very much for kindly sending me copies of their papers [9], [22], and [10]. I also thank Zuoqin Wang and Xiongjiu Liao for helpful literature searches.

References

[1] Davies, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.* **21** 1843–1899.

[2] Dieudonné, J. (1960). *Foundations of Modern Analysis.* 2d printing, "enlarged and corrected," 1969. Academic Press, New York.

[3] Dudley, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15** 1306–1326.

[4] Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press.

[5] Dudley, R. M. (2002). *Real Analysis and Probability*, 2d ed. Cambridge University Press.

[6] Dudley, R. M. (2005). Differentiability of t-functionals of location and scatter. Preprint.

[7] Dudley, R. M., Giné, E., and Zinn, J. (1991). Uniform and universal Glivenko-Cantelli classes. *J. Theoretical Probab.* **4** 485–510.

[8] Dudley, R. M. and Haughton, D. (2002). Asymptotic normality with small relative errors of posterior probabilities of half-spaces. *Ann. Statist.* **30** 1311–1344.

[9] Dümbgen, L. (1997). The asymptotic behavior of Tyler's M-estimator of scatter in high dimension. Preprint.

[10] Dümbgen, L., and Tyler, D. E. (2005). On the breakdown properties of some multivariate M-functionals. *Scand. J. Statist.* **32** 247–264.

[11] Frenkin, B. R. (2000). Some topological minimax theorems. *Mat. Zametki* **67** (1) 141–149 (Russian); English transl. in *Math. Notes* **67** (1-2) 112–118. Corrections, *ibid.* **68** (3) (2000), 480 (Russian), 416 (English).

[12] Giné, E., and Zinn, J. (1986). Empirical processes indexed by Lipschitz functions. *Ann. Probab.* **14** 1329–1338.

[13] Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Sympos. Math. Statist. Probability* **1**. Univ. California Press, Berkeley and Los Angeles, pp. 221–233.

[14] Huber, P. J. (1981). *Robust Statistics.* Wiley, New York. Reprinted, 2004.

[15] Kent, J. T. and Tyler, D. E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.* **19** 2102–2119.

[16] Kent, J. T., Tyler, D. E. and Vardi, Y. (1994). A curious likelihood identity for the multivariate T-distribution. *Commun. Statist.—Simula.* **23** 441–453.

[17] Murphy, S. A., and van der Vaart, A. W. (2000). On profile likelihood (with discussion). *J. Amer. Statist. Assoc.* **95** 449–485.

[18] Obenchain, R. L. (1971). Multivariate procedures invariant under linear transformations. *Ann. Math. Statist.* **42** 1569–1578.

[19] Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications*, Vol. B (W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, eds.). Dordrecht, Reidel,* pp. 283-297.

[20] Talagrand, M. (1987). The Glivenko-Cantelli problem. *Ann. Probab.* **15** 837–870.

[21] Talagrand, M. (1996). The Glivenko-Cantelli problem, ten years later. *J. Theoret. Probab.* **9** 371–384.

[22] Tyler, D. E. (1986). Breakdown properties of the M-estimators of multivariate scatter. Technical Report, Rutgers University.

* The author has seen the Rousseeuw (1985) paper cited in secondary sources (MathSciNet and several found by JSTOR) but not in the original.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 220–237

DOI: 10.1214/074921706000000879

Uniform error bounds for smoothing splines

P. P. B. Eggermont[1] **and V. N. LaRiccia**[1]

University of Delaware

Abstract: Almost sure bounds are established on the uniform error of smoothing spline estimators in nonparametric regression with random designs. Some results of Einmahl and Mason (2005) are used to derive uniform error bounds for the approximation of the spline smoother by an "equivalent" reproducing kernel regression estimator, as well as for proving uniform error bounds on the reproducing kernel regression estimator itself, uniformly in the smoothing parameter over a wide range. This admits data-driven choices of the smoothing parameter.

1. Introduction

In this paper, we study uniform error bounds for the smoothing spline estimator of arbitrary order for a nonparametric regression problem. In effect, we approximate the smoothing spline by a kernel-like estimator, and give sharp bounds on the approximation error under very mild conditions on the nonparametric regression problem, as well as on the uniform error on the kernel-like estimator. An application to obtaining confidence bands is pointed out.

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random sample of the bivariate random variable (X, Y) with $X \in [\,0\,,\,1\,]$, almost surely. Assume that

$$f_o(x) = \mathbb{E}[\,Y \mid X = x\,] \tag{1.1}$$

exists, and that for some natural number m,

$$f_o \in W^{m,\infty}(\,0\,,\,1\,), \tag{1.2}$$

where for $a < b$, the Sobolev spaces $W^{m,p}(a, b)$, $1 \le p \le \infty$, are defined as

$$W^{m,p}(a,b) = \left\{ f \in C^{m-1}[\,a\,,\,b\,] \;\middle|\; \begin{array}{l} f^{(m-1)} \text{ abs. continuous} \\ f^{(m)} \in L^p(a,b) \end{array} \right\}, \tag{1.3}$$

see, e.g., [2].

Regarding the design, assume that

(1.4) $X_1, X_2, \ldots, X_n$ are independent and identically distributed, having a probability density function w with respect to Lebesgue measure on $(\,0\,,\,1\,)$,

[1]Food and Resource Economics, University of Delaware, Newark, Delaware 19716-1303, e-mail: eggermon@udel.edu; lariccia@udel.edu

Keywords and phrases: spline smoothing, random designs, equivalent kernels.
AMS 2000 subject classifications: 62G08, 62G20.

and that

$$w_1 \le w(t) \le w_2 \quad \text{for all} \quad t \in [0, 1], \tag{1.5}$$

for positive constants w_1 and w_2.

With the random variable (X, Y), associate the noise D by

$$D = Y - f_o(X), \tag{1.6}$$

and define $D_i = Y_i - f_o(X_i)$, $i = 1, 2, \ldots, n$. Assume that

$$\sup_{x \in [0,1]} \mathbb{E}[\,|D|^\kappa \mid X = x\,] < \infty \quad \text{for some } \kappa > 2. \tag{1.7}$$

(With the assumption (1.2), this is equivalent to $\sup_x \mathbb{E}[\,|Y|^\kappa \mid X = x\,] < \infty$.)

Under the above conditions, uniform error bounds for the Nadaraya-Watson estimator have been established by Deheuvels and Mason [8] for a random choice of the smoothing parameter, and by Einmahl and Mason [12] uniformly in the smoothing parameter over a wide range. We recall that the Nadaraya-Watson estimator is defined as

$$\widehat{f}_n(t) = \tfrac{1}{n} \sum_{i=1}^n Y_i \, K_h(t - X_i) \Big/ \tfrac{1}{n} \sum_{i=1}^n K_h(t - X_i),$$

where, $K_h(t) = h^{-1} K(h^{-1} t)$ for some nice "kernel" K. In this case, $\widehat{f_n(x)}$ is an estimator of $f_o(x) = \mathbb{E}[Y \mid X = x]$. For some earlier results on uniform error bounds for Nadaraya-Watson estimators, see, e.g., [16] and [15].

For the smoothing spline estimator, we must come to terms with the fact that the estimator is defined implicitly as the solution, denoted by $f = f^{nh}$, of a minimization problem,

$$\begin{aligned} &\text{minimize} \quad \mathrm{LS}(f) \stackrel{\text{def}}{=} \tfrac{1}{n} \sum_{i=1}^n |\, f(X_i) - Y_i\,|^2 + h^{2m} \,\|\, f^{(m)} \,\|^2 \\ &\text{subject to} \quad f \in W^{m,2}(0, 1), \end{aligned} \tag{1.8}$$

where $\|\cdot\|$ denotes the $L^2(0,1)$ norm. Thus, the bulk of the paper is devoted to establishing that for all $t \in [0, 1]$,

$$f^{nh}(t) - \mathbb{E}[\, f^{nh}(t) \mid X_1, \ldots, X_n\,] = \tfrac{1}{n} \sum_{i=1}^n D_i \, \mathfrak{R}_{wmh}(X_i, t) + \varepsilon^{nh}(t), \tag{1.9}$$

where $\|\,\varepsilon^{nh}\,\|_\infty$ is negligible compared to the leading term in (1.9), and $\mathfrak{R}_{wmh}$ is the Green's function for a suitable boundary value problem, see (2.13). Here, $\|\cdot\|_\infty$ denotes the $L^\infty(0, 1)$ norm. The approach taken follows Eggermont and LaRiccia [10].

The precise results are as follows. For $\gamma > 0$, define the intervals

$$\mathcal{H}_n(\gamma) = \Big[\, \gamma \,\big(\frac{\log n}{n}\big)^{1-2/\kappa}, \tfrac{1}{2} \Big], \; \mathcal{G}_n(\gamma) = \Big[\, \gamma \,\big(\frac{\log n}{n}\big)^{1-2/\lambda}, \tfrac{1}{2} \Big], \tag{1.10}$$

where λ is unspecified but satisfies $2 < \lambda < \min(\kappa, 4)$.

Theorem 1. *Under the assumptions* (1.4) - (1.7) *on the model* (1.1)*, the error term* ε^{nh} *in* (1.9) *satisfies almost surely,*

$$T_{UE}(\gamma) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \sup_{h\in\mathcal{H}_n(\gamma)} \frac{\|\,\varepsilon^{nh}\,\|_\infty}{h^{-1/2}(nh)^{-1}\,\{\,\log(1/h) \vee \log\log n\,\}} < \infty .$$

The uniform-in-bandwidth character of this theorem (which admits *random* choices of the smoothing parameter) stands out. Regarding the actual error bound, if $h \in \mathcal{G}_n(\gamma)$, then $h \gg (n^{-1}\log n)^{1/2}$ and the error term in (1.9) can be ignored. Note that for $m \geq 2$ and $\kappa > 3$, this covers the optimal h, which behaves like $(n^{-1}\log n)^{1/(2m+1)}$. The theorem makes the smoothing spline much more accessible as an object of study. Here, we consider uniform error bounds on the estimator. For *cubic* smoothing splines in a somewhat different setting, uniform error bounds were derived by Chiang, Rice and Wu [5].

Main Theorem. *Assume the conditions* (1.2) *through* (1.7) *on the model* (1.1)*. Then, the spline estimator of order* m *satisfies almost surely,*

$$Q_{UE}(\gamma) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \sup_{h\in\mathcal{G}_n(\gamma)} \frac{\|\,f^{nh} - f_o\,\|_\infty}{\sqrt{h^{2m} + (nh)^{-1}\,\{\,\log(1/h) \vee \log\log n\,\}}} < \infty .$$

The constant Q_{UE} depends on the unknown regression function f_o through the bias. If we restrict h such that $h \ll (n^{-1}\log n)^{1/(2m+1)}$, then this dependence disappears, e.g., if for $m \geq 2$ and $\kappa > 2 + (1/m)$, we let

$$\mathcal{F}_n(\gamma) = \Big[\,\gamma\,\big(\,n^{-1}\log n\,\big)^{1-2/\kappa}\,,\; n^{-1/(2m+1)}\,\Big], \tag{1.12}$$

then

$$\mathfrak{Q}_{UE}(\gamma) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \sup_{h\in\mathcal{F}_n(\gamma)} \frac{\|\,f^{nh} - f_o\,\|_\infty}{\sqrt{(nh)^{-1}\,\{\,\log(1/h) \vee \log\log n\,\}}} < \infty, \tag{1.13}$$

and $\mathfrak{Q}_{UE}$ does not depend on f_o. This has obvious consequences for the construction of confidence bands. Since it seems reasonable that the value of $\mathfrak{Q}_{UE}$ can be determined via bootstrap techniques, then almost sure confidence bands in the spirit of Deheuvels and Mason [8] and the Bonferroni bounds of Eubank and Speckman [14] may be obtained. The full import of this will be explored elsewhere.

2. The smoothing spline estimator

Let $m \in \mathbb{N}$ and $h > 0$ be fixed. The smoothing spline estimator, denoted by f^{nh}, is defined as the solution of the minimization problem (1.8). The problem (1.8) always has solutions, and for $n \geq m$, the solution is unique, almost surely. For more on spline smoothing, see, e.g., [13] or [24].

A closer look at the spline smoothing problem reveals that $f(X_i)$ is well-defined for any $f \in W^{m,2}(\,0\,,\,1\,)$. In particular, there exists a constant c such that for all $f \in W^{m,2}(\,0\,,\,1\,)$ and all $x \in [\,0\,,\,1\,]$,

$$|\,f(x)\,| \leq c\,\{\,\|\,f\,\|^2 + \|\,f^{(m)}\,\|^2\,\}^{1/2}, \tag{2.1}$$

see, e.g., [2]. Then, a simple scaling argument shows that there exists a constant c_m such that for all $0 < h \le 1$, all $f \in W^{m,2}(0, 1)$, and all $t \in [0, 1]$,

$$| f(t) | \le c_m h^{-1/2} \| f \|_{mh}^2. \tag{2.2}$$

Here,

$$\| f \|_{mh} \stackrel{\text{def}}{=} \{ \| f \|^2 + h^{2m} \| f^{(m)} \|^2 \}^{1/2}. \tag{2.3}$$

Of course, the inequality (2.2) is geared towards the uniform design. For the present, "arbitrary" design, it is more appropriate to consider the inner products

$$\langle f , g \rangle_{wmh} = \langle f , g \rangle_{L^2((0,1),w)} + h^{2m} \langle f^{(m)} , g^{(m)} \rangle_{L^2(0,1)}, \tag{2.4}$$

where $\langle \cdot, \cdot \rangle_{L^2(0,1)}$ is the usual $L^2(0,1)$ inner product and

$$\langle f , g \rangle_{L^2((0,1),w)} = \int_0^1 f(t) g(t) w(t) \, dt. \tag{2.5}$$

The norms are then defined by $\| f \|_{wmh} = \{ \langle f , f \rangle_{wmh} \}^{1/2}$. With the design density being bounded and bounded away from zero, see (1.5), it is obvious that the norms $\| \cdot \|_{mh}$ and $\| \cdot \|_{wmh}$ are equivalent, uniformly in h. In particular, with the constants w_1 and w_2 as in (1.5), for all $f \in W^{m,2}(0, 1)$,

$$w_1 \| f \|_{mh} \le \| f \|_{wmh} \le w_2 \| f \|_{mh}. \tag{2.6}$$

(Note that, actually, $w_1 \le 1 \le w_2$.) Then, the analogue of (2.3) holds: There exists a constant c_m such that for all $0 < h \le 1$, all $f \in W^{m,2}(0, 1)$, and all $t \in [0, 1]$,

$$| f(t) | \le c_m h^{-1/2} \| f \|_{wmh}. \tag{2.7}$$

For later use, we quote the following multiplication result which follows readily with Cauchy-Schwarz: There exists a constant c such that for all f and $g \in W^{1,2}(0, 1)$,

$$\| f g \|_{L^1((0,1),w)} + h \| (f g)' \|_{L^1(0,1)} \le c \| f \|_{w,1,h} \| g \|_{w,1,h}. \tag{2.8}$$

Also, there exist constants $c_{k,k+1}$ such that for all $f \in W^{k+1,2}(0, 1)$,

$$\| f \|_{w,k,h} \le c_{k,k+1} \| f \|_{w,k+1,h}. \tag{2.9}$$

The inequality (2.7) says that the linear functionals $f \mapsto f(t)$ are continuous in the $\| \cdot \|_{wmh}$-topology, so that $W^{m,2}(0, 1)$ with the inner product $\langle \cdot, \cdot \rangle_{wmh}$ is a reproducing kernel Hilbert space, see [3]. Thus, by the Riesz-Fischer theorem on the representation of bounded linear functionals on Hilbert space, for each t, there exists an element $\mathfrak{R}_{wmht} \in W^{m,2}(0, 1)$ such that for all $f \in W^{m,2}(0, 1)$,

$$f(t) = \langle f , \mathfrak{R}_{wmht} \rangle_{wmh}. \tag{2.10}$$

Applying this to $\mathfrak{R}_{wmht}$ itself gives $\mathfrak{R}_{wmht}(s) = \langle \mathfrak{R}_{wmht} , \mathfrak{R}_{wmhs} \rangle_{wmh}$, so that it makes sense to define

$$\mathfrak{R}_{wmh}(t, s) = \mathfrak{R}_{wmht}(s) = \mathfrak{R}_{wmhs}(t) \quad \text{for all } s, t \in [0, 1]. \tag{2.11}$$

Then, again the inequality (2.7) implies that

$$\| \mathfrak{R}_{wmh}(t \,, \cdot) \|_{wmh} \le c_m \, h^{-1/2}, \tag{2.12}$$

with the same constant c_m.

Finally, we observe that reproducing kernels may be interpreted as the Green's functions for appropriate boundary value problems, see, e.g., [9]. In the present case, $\mathfrak{R}_{wmh}(t \,, s)$ is the Green's function for

$$\begin{aligned} &(-h^2)^m \, u^{(2m)} + w \, u = v \quad \text{on} \quad (\, 0 \,, 1 \,), \\ &u^{(k)}(0) = u^{(k)}(1) = 0, \quad k = m, \dots, 2m-1. \end{aligned} \tag{2.13}$$

In case $w(t) = 1$ for all t (the uniform density), we denote $\mathfrak{R}_{wmh}$ by $\mathcal{R}_{mh}$.

We finish this section by showing that the little information we have on the reproducing kernels suffices to prove some useful bounds on random sums of the form

$$\tfrac{1}{n} \sum_{j=1}^{m} D_i \, f(X_i),$$

with $D_1, D_2, \dots, D_n$ and $X_1, X_2, \dots, X_n$ as in Section 1, and $f \in W^{m,2}(\, 0 \,, 1 \,)$ random, i.e., depending on the D_i and X_i. To obtain these bounds, let

$$\mathfrak{S}_{nh}(t \,) \stackrel{\text{def}}{=} \tfrac{1}{n} \sum_{i=1}^{n} D_i \; \mathfrak{R}_{wmh}(X_i, \, t \,), \quad t \in [\, 0 \,, 1 \,]. \tag{2.14}$$

This is a reproducing-kernel regression estimator for pure noise data.

Lemma 1. *For every $f \in W^{m,2}(\, 0 \,, 1 \,)$, random or not,*

$$\Big| \, \tfrac{1}{n} \sum_{j=1}^{m} D_i \, f(X_i) \, \Big| \le \| \, f \, \|_{wmh} \, \| \, \mathfrak{S}_{nh} \, \|_{wmh},$$

and under the assumptions (1.4), (1.5) *and* (1.7), *there exists a constant c_m not depending on h such that $\mathbb{E}\big[\, \| \, \mathfrak{S}_{nh} \, \|^2_{wmh} \, \big] \le c_m \, (nh)^{-1}$.*

Proof. The identity $\frac{1}{n} \sum_{i=1}^{n} D_i \, f(X_i) = \big\langle \, f \,, \, \mathfrak{S}_{nh} \, \big\rangle_{wmh}$ implies the first bound by way of Cauchy-Schwarz. For the expectation, we have

$$\mathbb{E}\big[\, D_i \, \mathfrak{R}_{wmh}(X_i, \, t \,) \, \big] = \mathbb{E}\big[\, \mathbb{E}[\, D_i \mid X_i \,] \, \mathfrak{R}_{wmh}(X_i, \, t \,) \, \big] = 0$$

and so, since $D_i \, \mathfrak{R}_{wmh}(X_i, \, t \,)$, $i = 1, 2, \dots, n$, are independent and identically distributed (iid), it follows that

$$\begin{aligned} \mathbb{E}[\, \| \, \mathfrak{S}_{nh} \, \|^2_{L^2((\,0\,,1\,),w)} \,] &= n^{-2} \sum_{i=1}^{n} \mathbb{E}[\, D_i^2 \, \| \, \mathfrak{R}_{wmh}(\, X_i \,, \cdot \,) \, \|^2_{L^2((\,0\,,1\,),w)} \,] \\ &\le n^{-1} \, M \, \mathbb{E}[\, \| \, \mathfrak{R}_{wmh}(\, X \,, \cdot \,) \, \|^2_{L^2((\,0\,,1\,),w)} \,]. \end{aligned}$$

where

$$M = \sup_{x \in [\, 0 \,, 1 \,]} \mathbb{E}[\, D^2 \mid X = x \,]. \tag{2.15}$$

By (1.7), we have $M < \infty$.

Similarly, since $D_i \, \mathfrak{R}^{(m)}_{wmh}(X_i, t)$, $i = 1, 2, \dots, n$, are iid, then

$$\mathbb{E}[\, \| \, \mathfrak{S}^{(m)}_{nh} \, \|^2_{L^2(0,1)} \,] \le n^{-1} \, M \, \mathbb{E}[\, \| \, \mathfrak{R}^{(m)}_{wmh}(\, X \,, \, \cdot \,) \, \|^2_{L^2(0,1)} \,].$$

It follows that

$$\mathbb{E}[\, \| \, \mathfrak{S}_{nh} \, \|^2_{wmh} \,] \le n^{-1} \, M \, \mathbb{E}[\, \| \, \mathfrak{R}_{wmh}(X, \, \cdot \,) \, \|^2_{wmh} \,].$$

Now, (2.12) takes care of the last norm. □

3. Random sums

In this section, we discuss sharp bounds on the "random sums" $\mathfrak{S}_{nh}$ of (2.14), using results of Einmahl and Mason [12] regarding convolution-kernel estimators (in a more general setting). Thus, let

$$K \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \qquad \int_{\mathbb{R}} K(x) \, dx = 1. \tag{3.1}$$

We also need some restrictions on the "size" of the set of functions on $[\,0\,,\,1\,]$,

$$\mathcal{K} = \big\{ \, K\big(h^{-1}(\, x - \cdot \,)\big) \;\big|\; x \in [\,0\,,\,1\,]\,, \; 0 < h \le 1 \, \big\}. \tag{3.2}$$

First, we need to assume that

$$\mathcal{K} \quad \text{is pointwise measurable.} \tag{3.3}$$

For the definition of *pointwise measurability*, see van der Vaart and Wellner [23].

Let Q be a probability measure on $([\,0\,,\,1\,], \mathcal{B})$, and let $\| \cdot \|_Q$ denote the $L^2(Q)$ metric. For $\varepsilon > 0$, let $\mathcal{N}(\varepsilon, \mathcal{K}, \| \cdot \|_Q)$ denote the smallest number of balls in the $\| \cdot \|_Q$ metric needed to cover $\mathcal{K}$, i.e.,

$$\mathcal{N}\big(\varepsilon, \mathcal{K}, \| \cdot \|_Q\big) = \min \left\{ n \in \mathbb{N} \;\middle|\; \begin{array}{c} \exists \; g_1, g_2, \dots, g_n \in \mathcal{K} \quad \forall k \in \mathcal{K} \\ \displaystyle\min_{1 \le i \le n} \| \, k - g_i \, \|_Q \le \varepsilon \end{array} \right\}. \tag{3.4}$$

Then, let

$$\mathcal{N}(\varepsilon, \mathcal{K}) = \sup \mathcal{N}\big(\varepsilon, \mathcal{K}, \| \cdot \|_Q\,\big), \tag{3.5}$$

where the supremum is over *all* probability measures Q on $([\,0\,,\,1\,], \mathcal{B})$.

The restriction on the size of $\mathcal{K}$ now takes the form that there exist positive constants C and ν such that

$$\mathcal{N}(\varepsilon, \mathcal{K}) \le C \, \mathrm{e}^{-\nu}, \quad 0 < \varepsilon < 1. \tag{3.6}$$

Nolan and Pollard [19], see also [23], show that the condition (3.6) holds if the kernel K satisfies (3.1) and (3.3), and has bounded variation,

$$K \in BV(\mathbb{R}). \tag{3.7}$$

Whenever K has left and right limits everywhere (so in particular, when (3.7) holds), then (3.3) holds also.

The object of study is the following kernel "estimator" with "pure noise" data,

$$\mathbb{S}_{nh}(t) \stackrel{\text{def}}{=} \tfrac{1}{n} \sum_{i=1}^{n} D_i \, K_h(X_i - t), \quad t \in [0, 1]. \tag{3.8}$$

We quote the following slight modification as it applies to (3.8) of Proposition 2 of Einmahl and Mason [12] without proof. The modification involves the omission of the condition of compact support of the kernel K, which is permissible since the design is contained in a compact set, to wit the interval $[0, 1]$, [17]. Recall the definition of $\mathcal{H}_n(\gamma)$ from (1.10).

Proposition 1 (after Einmahl and Mason [12]). *Under the assumptions* (3.1), (3.3), (3.6), (3.7), *and* (1.4), (1.5), *and* (1.7), *for every* $\gamma > 0$,

$$\limsup_{n\to\infty} \sup_{h \in \mathcal{H}_n(\gamma)} \frac{\| \mathbb{S}_{nh} \|_\infty}{\sqrt{(nh)^{-1} \left\{ \log(1/h) \vee \log\log n \right\}}} < \infty \quad \textit{almost surely.}$$

Proof. The proof needs updating in only one spot, viz. the bound (3.20) of the Einmahl and Mason [12] paper needs to be established under the present conditions. However, that just amounts to showing that

$$\sup_{0<h\le 1} \sup_{t \in [0,1]} h \, \mathbb{E}[\, | D \, K_h(t - X) |^2 \,] < \infty.$$

Observe that

$$\mathbb{E}[\, | D \, K_h(t - X) |^2 \,] = \mathbb{E}[\, \mathbb{E}[\, D^2 \mid X \,] \, | K_h(t - X) |^2 \,] \le M \, \mathbb{E}[\, | K_h(t - X) |^2 \,],$$

with M as in (2.15). Now,

$$\begin{aligned} \mathbb{E}[\, | K_h(t - X) |^2 \,] &= \int_0^1 h^{-2} \, K^2\big(h^{-1}(t - x) \big) \, w(x) \, dx \\ &\le w_2 \, h^{-1} \int_{\mathbb{R}} K^2(x) \, dx \le w_2 \, \| K \|_{L^1(\mathbb{R})} \, \| K \|_{L^\infty(\mathbb{R})} \, h^{-1} \le c \, h^{-1}, \end{aligned}$$

for a suitable constant c, not depending on t. □

Now, we have the task of relating the random sums involving the reproducing kernels to sums involving convolution kernels. Obviously, some convolution-kernel-like properties of the reproducing kernel are required.

Definition 1. We say a family A_h, $0 < h < 1$, defined on $[0, 1] \times [0, 1]$, is convolution-like if it satisfies the following conditions: There exists a constant c such that for all $t \in [0, 1]$ and all h, $0 < h < 1$,

$$\| A_h(\cdot, t) \|_{L^1(0,1)} \le c, \;\; \| A_h(t, \cdot) \|_\infty \le c \, h^{-1}, \;\; | A_h(\cdot, t) |_{BV} \le c \, h^{-1}.$$

Here, $| f |_{BV}$ denotes the total variation of the function f over $[0, 1]$.

The families $h^\ell \, \mathfrak{R}^{(\ell)}_{wmh}(t, s)$, $0 < h < 1$, $\ell = 0, 1, \dots, m$, are indeed convolution-like, as shown in [11]. Here,

$$\mathfrak{R}^{(\ell)}_{wmh}(t, s) = \frac{d^\ell}{ds^\ell} \, \mathfrak{R}_{wmh}(t, s) \tag{3.9}$$

denotes the ℓ-th order derivative of $\mathfrak{R}_{wmh}(\,t\,,\,s\,)$ with respect to s (or by symmetry, with respect to t). This result is in the style of results on the "equivalent" kernel for spline smoothing, except that that the kernel is *not* a convolution kernel and that it handles arbitrary design densities subject to the condition (1.5) *and* treats the boundary conditions in (2.13) exactly. The relevant references on equivalent kernels for spline smoothing are [1, 5–7, 11, 18, 20, 21].

Now, there is an interesting way of connecting the reproducing kernel sum $\mathfrak{S}_{nh}$ to a sum $\mathbb{S}_{nh}$ for an appropriate kernel K. Define

$$\begin{aligned} g(x) &= \exp\bigl(-x\bigr)\,\mathbb{1}(\,x \geq 0\,), \\ g_h(x) &= h^{-1}\,g\bigl(\,h^{-1}x\,\bigr), \quad x \in \mathbb{R}. \end{aligned} \tag{3.10}$$

One verifies that $h\,g_h$ is the fundamental solution for the initial value problem

$$\begin{aligned} &h\,u' + u = v \quad \text{on } (\,0\,,\,1\,), \\ &u(0) = a, \end{aligned} \tag{3.11}$$

i.e., for $1 \leq p \leq \infty$ and $v \in L^p(\,0\,,\,1\,)$, the solution u of the initial value problem (3.11) satisfies $u \in L^p(\,0\,,\,1\,)$, and is given by

$$u(x) = h\,g_h(x)\,u(0) + \int_0^1 g_h(x-z)\,v(z)\,dz, \quad x \in [\,0\,,\,1\,], \tag{3.12}$$

see, e.g., [4], Section 2.1, formulas (10) through (14). Note that the last integral is really only over the interval $[\,0\,,\,x\,]$. Since $v = h\,u' + u$, this leads to the integral representation of the function u,

$$u(x) = h\,g_h(x)\,u(0) + \int_0^1 g_h(x-z)\,\bigl\{\,h\,u'(z) + u(z)\,\bigr\}\,dz, \quad x \in [\,0\,,\,1\,]. \tag{3.13}$$

Now, one verifies that

$$\text{the kernel } g \text{ satisfies (3.1), (3.3), (3.7),} \tag{3.14}$$

so that the class Γ generated by g,

$$\Gamma = \bigl\{\,g\bigl(h^{-1}(\,x - \,\cdot\,)\bigr) \;\bigm|\; x \in [\,0\,,\,1\,],\ 0 < h \leq 1\,\bigr\}. \tag{3.15}$$

satisfies (3.6), i.e.,

$$\mathcal{N}(\varepsilon, \Gamma) \leq C\,\mathrm{e}^{-\nu}, \quad 0 < \varepsilon < 1. \tag{3.16}$$

Thus, Proposition 1 would apply to the random sum

$$s^{nh}(z) = \tfrac{1}{n}\sum_{i=1}^{n} D_i\,g_h(X_i - z), \quad z \in [\,0\,,\,1\,], \tag{3.17}$$

but first we connect the sums $\mathfrak{S}_{nh}$ and s^{nh}.

Lemma 2. *Assume that the functions A_h, $0 < h \leq 1$, are convolution-like in the sense of Definition 1. Then, there exists a constant c such that for all h, $0 < h \leq 1$, all $D_1, D_2, \ldots, D_n \in \mathbb{R}$, and all positive $X_1, X_2, \ldots, X_n \in [\,0\,,\,1\,]$,*

$$\bigl\|\,\tfrac{1}{n}\sum_{i=1}^{n} D_i\,A_h(X_i, \,\cdot\,)\,\bigr\|_\infty \leq c\,\bigl\|\,\tfrac{1}{n}\sum_{i=1}^{n} D_i\,g_h(X_i - \,\cdot\,)\,\bigr\|_\infty.$$

Proof. Assume that A_h is differentiable with respect to its first argument. Then,

$$| A_h(\,\cdot\,,\, t\,) |_{BV} = \| A_h'(\,\cdot\,,\, t\,) \|_{L^1(0,1)},$$

where the prime $'$ denotes differentiation with respect to the first argument. Now, apply (3.13) to the function $u = A_h(\,\cdot\,,\, t\,)$ (for fixed t), so for all x,

$$A_h(x,\, t\,) = h\, g_h(x)\, A_h(0,\, t\,) + \int_0^1 g_h(x-z) \,\{\, h\, A_h'(z,\, t\,) + A_h(z,\, t\,) \,\}\, dz. \tag{3.18}$$

Next, take $x = X_i$ and substitute this into

$$S_{nh}(\,t\,) \stackrel{\text{def}}{=} \tfrac{1}{n} \sum_{i=1}^{n} D_i\, A_h(X_i,\, t\,), \qquad t \in [\,0\,,\,1\,].$$

Then, we have

$$S_{nh}(\,t\,) = h\, A_h(0,\, t\,)\, s^{nh}(0) + \int_0^1 \{h\, A_h'(z,\, t\,) + A_h(z,\, t\,)\,\}\, s^{nh}(z)\, dz. \tag{3.19}$$

Now, straightforward bounding gives

$$\|\, S_{nh} \,\|_\infty \le C \,|\, s^{nh}(0) \,| + C_1 \,\|\, s^{nh} \|_\infty,$$

where $C = h \,\|\, A_h(0,\,\cdot\,) \,\|_\infty$ and

$$C_1 = \sup_{t\,\in[\,0\,,\,1\,]} \|\, A_h(\,t\,,\,\cdot\,) \,\|_{L^1(\,0\,,\,1\,)} + h \,|\, A_h(\,t\,,\,\cdot\,) \,|_{BV}\,.$$

So, by the convolution-like properties of A_h, the constants C and C_1 are bounded, uniformly in h. Also, since all of the X_i are positive, then

$$\lim_{z \to 0} s^{nh}(z) = s^{nh}(0),$$

and so, for $C_2 = C + C_1$, we have $\|\, S_{nh} \,\|_\infty \le C_2 \,\|\, s^{nh} \,\|_\infty$.

The extension to the case where A_h is not necessarily differentiable with respect to its first argument follows readily. □

Since the families $h^\ell\, \mathfrak{R}^{(\ell)}_{wmh}(\,t\,,\, s\,)$, $0 < h < 1$, $\ell = 0, 1, \ldots, m$, are convolution-like in the sense of Definition 1, we may apply the lemma to the sum $\mathfrak{S}_{nh}$ of (2.14) and its derivatives. This yields

$$h^\ell \,\|\, \mathfrak{S}^{(\ell)}_{nh} \,\|_\infty \le c \,\|\, s^{nh} \,\|_\infty, \quad \ell = 0, 1, \ldots, m. \tag{3.20}$$

Now, for the model (1.1) through (1.7), the sum s^{nh} of (3.17) may be treated by the above formulated Proposition 1. This proves the following result. (Recall the definition (1.10) of $\mathcal{H}_n(\gamma)$.)

Theorem 2. *Under the assumptions* (1.4), (1.5) *and* (1.7), *for* $\gamma > 0$, *and for* $\ell = 0, 1, \ldots, m$,

$$Q_{\infty,\ell}(\gamma) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \sup_{h \in \mathcal{H}_n(\gamma)} \frac{h^\ell \,\|\, \mathfrak{S}^{(\ell)}_{nh} \,\|_\infty}{\sqrt{(nh)^{-1} \{\log(1/h) \vee \log\log n\}}} < \infty,$$

almost surely.

It turns out that we need a similar result for the $\|\cdot\|_{wmh}$ norm, which requires a result for the L^2 norm. The following is good enough for our purposes. Obviously, with $\mathfrak{S}_{nh}^{(m)}$ denoting the m-th order derivative of $\mathfrak{S}_{nh}$, we have

$$\| \mathfrak{S}_{nh} \|_{L^2((0,1),w)} \le c \, \| \mathfrak{S}_{nh} \|_\infty , \quad \| \mathfrak{S}_{nh}^{(m)} \|_{L^2(0,1)} \le \| \mathfrak{S}_{nh}^{(m)} \|_\infty ,$$

with $c = \sqrt{w_2}$, and then Theorem 2 gives useful bounds for the $\|\cdot\|_{wmh}$ norm.

Corollary 3. *Under the conditions of Theorem 2, we have almost surely,*

$$Q_{wm} \stackrel{\text{def}}{=} \limsup_{n\to\infty} \sup_{h\in\mathcal{H}_n(\gamma)} \frac{\| \mathfrak{S}_{nh} \|_{wmh}}{\sqrt{(nh)^{-1} \{\log(1/h) \vee \log\log n\}}} < \infty .$$

4. The design sums

The reproducing kernel Hilbert space set-up is also useful for connecting *random design sums*

$$\tfrac{1}{n} \sum_{i=1}^{n} |\, f(X_i)\,|^2$$

to their (partial?) expectations

$$\int_0^1 |\, f(x)\,|^2 \, w(x)\, dx ,$$

for *random* functions f. In particular, we prove the following almost sure result. The range of the smoothing parameter is much larger here than it was before, although we only need it for $h \in \mathcal{H}_n(\gamma)$, see (1.10). Here, let

$$\mathcal{D}_n(\gamma) = \big[\, \gamma\, n^{-1} \log n \,,\, \tfrac{1}{2} \,\big]. \tag{4.1}$$

Theorem 4. *Under the assumptions* (1.4) *and* (1.5), *for all* $f \in W^{m,2}(0\,,1)$,

$$\tfrac{1}{n} \sum_{j=1}^{n} |\, f(X_i)\,|^2 + h^{2m} \| f^{(m)} \|^2 \ge r_{nh} \| f \|_{wmh}^2 ,$$

where $\displaystyle \liminf_{n\to\infty} \inf_{h\in\mathcal{D}_n(\gamma)} r_{nh} = 1$ *almost surely.*

To prove this, let W be the (cumulative) distribution function corresponding to the design density w and let W_n be the empirical distribution function of the design $X_1, X_2, \ldots, X_n$, and introduce the "design sums"

$$w^{nh}(\,t\,) = g_h * dW_n(\,t\,) \stackrel{\text{def}}{=} \tfrac{1}{n} \sum_{i=1}^{n} g_h(X_i - t\,), \;\; t \in [\,0\,,1\,], \tag{4.2}$$

which is a *convolution*-kernel density estimator and its expectation,

$$\mathbb{E}[\, w^{nh}(\,t\,)\,] = g_h * dW(\,t\,) = \int_0^1 g_h(\tau - t\,)\, w(\tau)\, d\tau, \quad t \in [\,0\,,1\,]. \tag{4.3}$$

We will use Theorem 1 of Einmahl and Mason [12], quoted here for convenience. (This time, no modifications are necessary.)

Proposition 2 ([12]). *Under the assumptions* (3.1), (3.3), (3.6), *and* (3.7), *and* (1.4) *and* (1.5), *for every* $\gamma > 0$,

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{D}_n(\gamma)} \frac{\|\, w^{nh} - \mathbb{E}[\, w^{nh}\,]\,\|_\infty}{\sqrt{nh\,\bigl(\log(1/h)\vee\log\log n\bigr)}} < \infty \quad \textit{almost surely.}$$

To prove Theorem 4, we start with simple "design sums".

Lemma 3. *Under the assumptions* (1.4) *and* (1.5), *for all* $f \in W^{1,1}(\,0\,,\,1\,)$,

$$\Bigl|\int_0^1 f(\,t\,)\,\{\,dW_n(\,t\,) - dW(\,t\,)\,\}\Bigr| \le \zeta_{nh}\,\bigl\{\,\|\,f\,\|_{L^1((\,0\,,\,1\,),w)} + h\,\|\,f'\,\|_{L^1(\,0\,,\,1\,)}\,\bigr\}$$

where $$\limsup_{n\to\infty} \sup_{h\in\mathcal{D}_n(\gamma)} \frac{\zeta_{nh}}{\sqrt{(nh)^{-1}\,\{\,\log(1/h)\vee\log\log n\,\}}} < \infty$$
almost surely.

Proof. With the reproducing kernel Hilbert space trick,

$$f(\,t\,) = \langle\, f\,,\,\mathfrak{R}_{w,1,h}(\,t\,,\,\cdot\,)\,\rangle_{w,1,h},$$

we obtain by linearity and Fubini's theorem that

$$\int_0^1 f(\,t\,)\,\{\,dW_n(\,t\,) - dW(\,t\,)\,\} = \langle\, f\,,\,\delta^{nh}\,\rangle_{w,1,h},$$

where

$$\delta^{nh}(s) = \int_0^1 \mathfrak{R}_{w,1,h}(\,t\,,\,s\,)\,\{\,dW_n(\,t\,) - dW(\,t\,)\},\quad s\in[\,0\,,\,1\,],$$

is the variance part of the pointwise error of a reproducing-kernel estimator of the design density w. Now, straightforward bounding gives

$$\begin{aligned}\langle\, f\,,\,\delta^{nh}\,\rangle_{L^2((\,0\,,\,1\,),w)} &\le \|\,f\,\|_{L^1((\,0\,,\,1\,),w)}\,\|\,\delta^{nh}\,\|_\infty,\\ \langle\, f'\,,\,(\delta^{nh})'\,\rangle_{L^2(0,1)} &\le \|\,f'\,\|_{L^1(\,0\,,\,1\,)}\,\|\,(\delta^{nh})'\,\|_\infty,\end{aligned}$$

so that

$$\langle\, f\,,\,\delta^{nh}\,\rangle_{w,1,h} \le \bigl\{\,\|\,f\,\|_{L^1((\,0\,,\,1\,),w)} + h\,\|\,f'\,\|_{L^1(\,0\,,\,1\,)}\,\bigr\}\,\bigl\{\,\|\,\delta^{nh}\,\|_\infty + h\,\|\,(\delta^{nh})'\,\|_\infty\,\bigr\},$$

with, explicitly,

$$\begin{aligned}\|\,\delta^{nh}\,\|_\infty &= \Bigl\|\,\tfrac{1}{n}\sum_{i=1}^n \mathfrak{R}_{w,1,h}(X_i,\,\cdot\,) - \mathbb{E}[\,\mathfrak{R}_{w,1,h}(X_1\,,\,\cdot\,)\,]\,\Bigr\|_\infty,\\ h\,\|\,(\delta^{nh})'\,\|_\infty &= \Bigl\|\,\tfrac{1}{n}\sum_{i=1}^n h\,\mathfrak{R}'_{w,1,h}(X_i,\,\cdot\,) - h\,\mathbb{E}[\,\mathfrak{R}'_{w,1,h}(X_1\,,\,\cdot\,)\,]\,\Bigr\|_\infty.\end{aligned}$$

Both of these may be interpreted as the variance part of the uniform error of (reproducing) kernel estimators. As already noted, the families $\mathfrak{R}_{wmh}(\,t\,,\,s\,)$ and $h\,\mathfrak{R}'_{w,1,h}(\,t\,,\,s\,)$ are convolution-like in the sense of Definition 1. Then, by an appeal to Lemma 2 with $D_i = 1$ for all i,

$$\begin{aligned}\|\,\delta^{nh}\,\|_\infty &\le C\,\|\,g_h * \{\,dW_n - dW\,\}\,\|_\infty,\\ \|\,(\delta^{nh})'\,\|_\infty &\le C_1\,\|\,g_h * \{\,dW_n - dW\,\}\,\|_\infty,\end{aligned}$$

for suitable constants C and C_1. Now, an appeal to Theorem 1 of Einmahl and Mason [12], see Proposition 2 above, clinches the deal. □

We now get the following lemma, which immediately implies Theorem 4.

Lemma 4. *Under the assumptions* (1.4) *and* (1.5), *for all* $f,\, g \in W^{m,2}(0\,,\,1)$,

$$\Bigl|\int_0^1 f(t)\,g(t)\,\{\,dW_n(t) - dW(t)\,\}\Bigr| \le \eta_{nh}\,\|\,f\,\|_{wmh}\,\|\,g\,\|_{wmh},$$

where
$$\limsup_{n\to\infty}\ \sup_{h\in\mathcal{D}_n(\gamma)} \frac{\eta_{nh}}{\sqrt{(nh)^{-1}\,\{\,\log(1/h)\vee\log\log n\,\}}} < \infty$$
almost surely.

Proof. From Lemma 3, we get the bound

$$\zeta_{nh}\,\{\,\|\,f\,g\,\|_{L^1((\,0\,,\,1\,),w)} + h\,\|\,(f\,g)'\,\|_{L^1(\,0\,,\,1\,)}\,\},$$

with the requisite behavior of ζ^{nh}. Now, from (2.8),

$$\|\,f\,g\,\|_{L^1((\,0\,,\,1\,),w)} +\ h\,\|\,(f\,g)'\,\|_{L^1(\,0\,,\,1\,)} \le\ c\,\|\,f\,\|_{w,1,h}\,\|\,g\,\|_{w,1,h}$$

for an appropriate constant c. Finally, (2.9) gives $\|\,f\,\|_{w,1,h} \le c\,\|\,f\,\|_{w,m,h}$, again for an appropriate constant c, and likewise for g. Thus, η_{nh} satisfies $\eta_{nh} \le c\,\zeta_{nh}$, and the lemma follows. □

5. L^2 error bounds

We are now ready to prove almost sure bounds on $\|\,f^{nh} - f_o\,\|^2_{wmh}$ for the spline smoother f^{nh}. The starting point is the quadratic Taylor expansion of the objective function LS(f) of (1.8) around its minimizer. Let

$$\varepsilon \equiv f^{nh} - f_o. \tag{5.1}$$

Since the Gateaux variation of LS at its minimizer vanishes, this gives

$$\tfrac{1}{n}\sum_{i=1}^{n} |\,\varepsilon(X_i)\,|^2 + h^{2m}\,\|\,\varepsilon^{(m)}\,\|^2 = \mathrm{LS}(\,f_o\,) - \mathrm{LS}(\,f^{nh}\,). \tag{5.2}$$

Now, again, simple quadratic Taylor expansion around f_o gives

$$\begin{aligned}\mathrm{LS}(\,f_o\,) - \mathrm{LS}(\,f^{nh}\,) = -\,\tfrac{1}{n}\sum_{i=1}^{n} |\,\varepsilon(X_i)\,|^2 + \tfrac{2}{n}\sum_{i=1}^{n} D_i\,\varepsilon(X_i)& \\ -\,h^{2m}\,\|\,\varepsilon^{(m)}\,\|^2 + 2\,h^{2m}\,\langle\,f_o^{(m)}\,,\,\varepsilon^{(m)}\,\rangle&,\end{aligned} \tag{5.3}$$

and so, after substitution into (5.2),

$$\tfrac{1}{n}\sum_{i=1}^{n} |\,\varepsilon(X_i)\,|^2 + h^{2m}\,\|\,\varepsilon^{(m)}\,\|^2 = \tfrac{1}{n}\sum_{i=1}^{n} D_i\,\varepsilon(X_i) + h^{2m}\,\langle\,f_o^{(m)}\,,\,\varepsilon^{(m)}\,\rangle. \tag{5.4}$$

This is similar to the development in [22].

Now, with Lemma 1, Theorem 4, and Cauchy-Schwarz, one obtains

$$r^{nh} \, \| \, \varepsilon \, \|_{wmh}^2 \le \| \, \varepsilon \, \|_{wmh} \, \{ \, \| \, \mathfrak{S}_{nh} \, \|_{wmh} + h^m \, \| \, f_o^{(m)} \, \| \, \}, \tag{5.5}$$

where we took the liberty of using $h^m \| \, \varepsilon^{(m)} \, \| \le \| \, \varepsilon \, \|_{wmh}$. It follows that

$$r^{nh} \, \| \, \varepsilon \, \|_{wmh} \le \| \, \mathfrak{S}_{nh} \, \|_{wmh} + h^m \, \| \, f_o^{(m)} \, \|, \tag{5.6}$$

and the following result emerges.

Theorem 5. *For the model* (1.1)*, under the assumptions* (1.4)*,* (1.6)*,* (1.2)*, and* (1.5)*, with $\mathcal{H}_n(\gamma)$ defined in* (1.10)*, for $\gamma > 0$, almost surely,*

$$\limsup_{n \to \infty} \sup_{h \in \mathcal{H}_n(\gamma)} \frac{\| \, f^{nh} - f_o \, \|_{wmh}}{\sqrt{h^{2m} + (nh)^{-1} \{ \log(1/h) \vee \log\log n \, \}}} < \infty,$$

and for $h \asymp (n^{-1} \log n)^{1/(2m+1)}$ (deterministic or random),

$$\| \, f^{nh} - f_o \, \|_{wmh} = \mathcal{O}\big(\big(n^{-1} \log n \big)^{m/(2m+1)} \big) \quad \textit{almost surely.}$$

Proof. This follows from (5.6) and Corollary 3. □

The error bound (5.6) appears to be quite sharp. In the next section, we show that $f^{nh}(\, t\,) - \mathbb{E}[\, f^{nh}(\, t\,) \mid X_1, \dots, X_n \,] \approx \mathfrak{S}_{nh}$ in a precise, useful sense.

6. C-splines

In this section, we determine a useful, accurate expression for the variance part $f^{nh}(\, t\,) - \mathbb{E}[\, f^{nh}(\, t\,) \mid X_1, \dots, X_n \,]$ of the pointwise error $f^{nh}(\, t\,) - f_o(\, t\,)$, with an eye towards almost sure uniform error bounds. Since the estimator f^{nh} is linear in the data, one sees that

$$\varphi^{nh} = f^{nh} - \mathbb{E}[\, f^{nh} \mid X_1, \dots, X_n \,] \tag{6.1}$$

is the solution to the "pure noise" problem

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{n} \sum_{i=1}^n | \, f(X_i) - D_i \, |^2 + h^{2m} \, \| \, f^{(m)} \, \|^2 \\ &\text{subject to} \quad f \in W^{m,2}(\, 0 \,, 1 \,). \end{aligned} \tag{6.2}$$

In fact, we show that $\varphi^{nh}(\, t\,) \approx \psi^{nh}(\, t\,)$, where $f = \psi^{nh}$ solves the C(ontinuous)-spline problem

$$\begin{aligned} &\text{minimize} \quad \| \, f \, \|_{L^2((\, 0 \,, 1 \,), w)}^2 - \tfrac{2}{n} \sum_{i=1}^n D_i \, f(X_i) + h^{2m} \, \| \, f^{(m)} \, \|^2 \\ &\text{subject to} \quad f \in W^{m,2}(\, 0 \,, 1 \,). \end{aligned} \tag{6.3}$$

By the interpretation of the reproducing kernel $\mathfrak{R}_{wmh}$ as the Green's function for the boundary value problem (2.13), one observe that ψ^{nh} is given by

$$\psi^{nh}(\, t\,) = \tfrac{1}{n} \sum_{i=1}^n D_i \, \mathfrak{R}_{wmh}(X_i, \, t\,), \tag{6.4}$$

so that the following almost sure error bounds apply.

Theorem 6. *Under the assumptions of Theorem* 5, *almost surely, uniformly in* $h \in H_n(\gamma)$, *see* (1.10),

$$\| \varphi^{nh} - \psi^{nh} \|_{wmh} = \mathcal{O}\big(h^{-1/2} (nh)^{-1} \{ \log(1/h) \vee \log\log n \} \big),$$
$$\| \varphi^{nh} - \psi^{nh} \|_{\infty} = \mathcal{O}\big(h^{-1} (nh)^{-1} \{ \log(1/h) \vee \log\log n \} \big).$$

Proof. Let $\varepsilon = \varphi^{nh} - \psi^{nh}$. Similar to the inequality (5.4), one obtains quadratic inequalities for the discrete and the continuous spline problems. Adding these gives

$$\| \varepsilon \|^2 + 2 h^{2m} \| \varepsilon^{(m)} \|^2 + \tfrac{1}{n} \sum_{i=1}^{n} | \varepsilon(X_i) |^2 = \text{rhs}, \tag{6.5}$$

where

$$\text{rhs} = \int_0^1 g(t) \{ dW_n(t) - dW(t) \},$$

with $g = | \varphi^{nh} |^2 - | \psi^{nh} |^2 = (\varphi^{nh} + \psi^{nh}) \varepsilon$, and ε as above.

Now let $\gamma > 0$ be fixed. Then, the following statements hold uniformly in $h \in \mathcal{H}_n(\gamma)$. Using Lemma 3, one obtains, almost surely

$$\begin{aligned} \text{rhs} &= \mathcal{O}\big(\sqrt{(nh)^{-1} \{ \log(1/h) \vee \log\log n \}} \big) \big\{ \| g \|_{L^1((0,1),w)} + h \| g \|_{L^1(0,1)} \big\} \\ &= \mathcal{O}\big(\sqrt{(nh)^{-1} \{ \log(1/h) \vee \log\log n \}} \big) \| \varepsilon \|_{wmh} \| \varphi^{nh} + \psi^{nh} \|_{wmh}, \end{aligned}$$

where we used the multiplication result (2.8), and (2.9). Substituting this into (6.5), one obtains almost surely,

$$\| \varepsilon \|_{wmh} = \mathcal{O}\big(\sqrt{(nh)^{-1} \{ \log(1/h) \vee \log\log n \}} \big) \| \varphi^{nh} + \psi^{nh} \|_{wmh}.$$

Now,

$$\| \varphi^{nh} + \psi^{nh} \|_{wmh} \le \| \varphi^{nh} \|_{wmh} + \| \psi^{nh} \|_{wmh},$$

and consequently, by Theorem 5 applied with $f_o = 0$, and (6.5),

$$\| \varphi^{nh} + \psi^{nh} \|_{wmh} = \mathcal{O}\big(\sqrt{(nh)^{-1} \{ \log(1/h) \vee \log\log n \}} \big) \quad \text{almost surely.}$$

Thus,

$$\| \varepsilon \|_{wmh} = \mathcal{O}\big((nh)^{-1} \{ \log(1/h) \vee \log\log n \} \big),$$

and so, at the loss of a factor $h^{-1/2}$,

$$\| \varepsilon \|_{\infty} = \mathcal{O}\big(h^{-1/2} (nh)^{-1} \{ \log(1/h) \vee \log\log n \} \big).$$

The theorem has been proved. □

The above completes the proof of Theorem 1.

7. C-splines: the Bias

In considering the bias of the estimator f^{nh}, note that $f_h = \mathbb{E}[f^{nh} \mid X_1, \dots, X_n]$ is the solution of (1.8) with $D_i = 0$, $i = 1, 2, \dots, n$, i.e., the solution of the discrete noiseless problem

$$\begin{aligned} &\text{minimize} \quad DN(f) \stackrel{\text{def}}{=} \tfrac{1}{n} \sum_{i=1}^{n} | f(X_i) - f_o(X_i) |^2 + h^{2m} \| f^{(m)} \|^2 \\ &\text{subject to} \quad f \in W^{m,2}(0, 1). \end{aligned} \tag{7.1}$$

Note that the randomness in f_h is due to the randomness of the design. We must compare f_h to f_o, but it is easier to first compare f_h to φ_h, the solution of the continuous noiseless problem

$$
\text{(7.2)} \qquad \begin{aligned} &\text{minimize} \quad CN(f) \stackrel{\text{def}}{=} \| f - f_o \|^2_{L^2((0,1),w)} + h^{2m} \| f^{(m)} \|^2 \\ &\text{subject to} \quad f \in W^{m,2}(0,1). \end{aligned}
$$

In this section we prove the following theorem on the conditional bias. Note the "restricted" set $\mathcal{G}_n(\gamma)$ of allowable values of h.

Theorem 7. *Under the assumptions of Theorem* 5, *almost surely,*

$$
\limsup_{n\to\infty} \sup_{h\in\mathcal{G}_n(\gamma)} \frac{\| f_h - f_o \|_\infty}{\sqrt{h^{2m} + (nh)^{-1}\{\log(1/h) \vee \log\log n\}}} < \infty.
$$

The proof goes again by way of the reproducing kernel approximation, and follows without further ado from the following two lemmas.

Lemma 5. *Under the assumptions of Theorem* 5, *almost surely,*

$$
\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_n(\gamma)} \frac{\| \varphi_h - f_h \|_\infty}{h^{-1/2}\{ h^{2m} + (nh)^{-1} \{\log(1/h) \vee \log\log n\}\}} < \infty.
$$

Proof. Let $\varepsilon = \varphi_h - f_h$. Similar to the derivation of (5.6), one obtains

$$
\text{(7.3)} \qquad 2\,\| \varepsilon \|^2_{wmh} = \text{rhs}
$$

where "rhs" $= DN(\varphi_h) - DN(f_h) + CN(f_h) - CN(\varphi_h)$. This simplifies to

$$
\text{rhs} = \int_0^1 g(t)\,\{ dW_n(t) - dW(t) \},
$$

with W and W_n as in Lemma 3, and

$$
\begin{aligned} g(t) &= |\,\varphi_h(t) - f_o(t)\,|^2 - |\,f_h(t) - f_o(t)\,|^2 \\ &= \big(\varphi_h(t) + f_h(t) - 2\,f_o(t)\big)\,\varepsilon(t). \end{aligned}
$$

By Lemma 4, we get that

$$
\text{(7.4)} \qquad \text{rhs} \le \eta_{nh}\,\| \varphi_h + f_h - 2\,f_o \|_{wmh}\,\| \varepsilon \|_{wmh},
$$

with

$$
\text{(7.5)} \qquad \eta_{nh} = \mathcal{O}\big(\sqrt{(nh)^{-1}\{\log(1/h) \vee \log\log n\}}\,\big)
$$

almost surely, uniformly in $h \in \mathcal{H}_n(\gamma)$.

Substituting (7.4) into (7.3), we obtain

$$
\text{(7.6)} \qquad \| \varepsilon \|_{wmh} \le \tfrac{1}{2}\,\eta_{nh}\,\| \varphi_h + f_h - 2\,f_o \|_{wmh}.
$$

Now, with regards to bounding $\| \varphi_h + f_h - 2\,f_o \|_{wmh}$, the situation is as in Section 5, except that the $D_i = 0$, $i = 1, 2, \ldots, n$. Then, almost surely,

$$
\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_n(\gamma)} h^{-m}\,\| f_h - f_o \|_{wmh} < \infty.
$$

(Note that there is still randomness in f_h due to the design.) In the same way, one obtains that deterministically,

$$\| \varphi_h - f_o \|_{wmh} = \mathcal{O}(h^m).$$

It then follows from (7.6) that

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_n(\gamma)} \frac{\| \varepsilon \|_{wmh}}{h^m \sqrt{(nh)^{-1}\{ \log(1/h) \vee \log\log n\}}} < \infty.$$

Of course,

$$2\, h^m \sqrt{(nh)^{-1}\{ \log(1/h) \vee \log\log n\}} \le h^{2m} + (nh)^{-1}\{ \log(1/h) \vee \log\log n\},$$

so that

$$\frac{2 \| \varepsilon \|_{wmh}}{h^{2m} + (nh)^{-1}\{ \log(1/h) \vee \log\log n\}} \le \frac{\| \varepsilon \|_{wmh}}{h^m \sqrt{(nh)^{-1}\{ \log(1/h) \vee \log\log n\}}}.$$

At the loss of a factor $h^{1/2}$, this gives us the required bound on $\| \varepsilon \|_\infty$. □

Lemma 6. *Under the assumption* (1.2)*, there exists a constant* c*, such that*

$$\| \varphi_h - f_o \|_\infty \le c\, h^m \| f_o^{(m)} \|_\infty,$$

provided f_o *satisfies* (1.5).

Proof. One verifies that φ_h is the solution to the differential equation

$$(-h^2)^m f^{(2m)} + w\, f = w\, f_o \quad \text{on } (0, 1), \tag{7.7}$$

supplemented with the natural boundary conditions. Now, we assume that the regression function f_o satisfies $f_o \in W^{m,\infty}(0, 1)$, so certainly, $f_o \in W^{m,2}(0, 1)$. Then, cf. (2.13), the solution of (7.7) is given by

$$\varphi_h(t) = \int_0^1 \mathfrak{R}_{wmh}(t, s)\, w(s)\, f_o(s)\, ds = \big\langle \mathfrak{R}_{wmh}(t, \cdot), f_o \big\rangle_{L^2((0,1),w)},$$

and so

$$\varphi_h(t) = \big\langle \mathfrak{R}_{wmh}(t, \cdot), f_o \big\rangle_{wmh} - h^{2m} \big\langle \mathfrak{R}^{(m)}_{wmh}(t, \cdot), f_o^{(m)} \big\rangle_{L^2(0,1)},$$

with $\mathfrak{R}^{(m)}_{wmh}(t, s)$ the m-th derivative of $\mathfrak{R}_{wmh}(t, s)$ with respect to s, as in (3.9). Since, $\big\langle \mathfrak{R}_{wmh}(t, \cdot), f_o \big\rangle_{wmh} = f_o(t)$, and

$$\begin{aligned} \Big| \big\langle \mathfrak{R}^{(m)}_{wmh}(t, \cdot), f_o^{(m)} \big\rangle_{L^2(0,1)} \Big| &\le \| \mathfrak{R}^{(m)}_{wmh}(t, \cdot) \|_{L^1((0,1),w)} \| f_o^{(m)} \|_\infty \\ &\le c\, h^{-m} \| f_o^{(m)} \|_\infty, \end{aligned}$$

the last inequality by the convolution-likeness of $h^m\, \mathfrak{R}^{(m)}_{wmh}$, the lemma follows. □

8. Uniform error bounds

As an application of the reproducing-kernel approximation to the spline smoother, we obtain uniform error bounds on the spline smoother, uniformly in the bandwidth over a wide (useful) range.

Proof of the Main Theorem. First, let us consider the result of Theorem 1. Recall that the range of the bandwidths is $\mathcal{G}_n(\gamma) = [\,\gamma\,(n^{-1}\log n)^{1-2/\lambda}\,,\,1/2\,]$ and that $2 < \lambda < \min(\kappa, 4)$ with $\kappa > 3$. So, $h \in \mathcal{G}_n(\gamma)$ implies that $h \gg (n^{-1}\log n)^{1/2}$. Now, from Theorem 6, with $f_h = \mathbb{E}[\, f^{nh} \mid X_1, \cdots, X_n\,]$, and $\mathfrak{S}_{nh}$ given by (2.14),

$$f^{nh}(\,t\,) - f_h(\,t\,) = \mathfrak{S}_{nh}(\,t\,) + \varepsilon^{nh}(\,t\,),$$

with, almost surely, uniformly in $h \in \mathcal{H}_n(\gamma)$,

$$\|\,\varepsilon^{nh}\,\|_\infty = \mathcal{O}\Big(\,h^{-1/2}(nh)^{-1}\,\{\,\log(1/h) \vee \log\log n\,\}\,\Big).$$

For $h \gg (n^{-1}\log n)^{1/2}$, we may conclude that

$$\|\,\varepsilon^{nh}\,\|_\infty = o\Big(\,\sqrt{\,(nh)^{-1}\,\{\,\log(1/h) \vee \log\log n\,\}\,}\,\Big),$$

which is negligible compared to the upperbound of Theorem 2,

$$\|\,\mathfrak{S}_{nh}\,\|_\infty = \mathcal{O}\big(\,\sqrt{\,(nh)^{-1}\,\{\,\log(1/h) \vee \log\log n\,\}\,}\,\big)$$

almost surely, uniformly in $h \in \mathcal{H}_n(\gamma)$. Finally,

$$\begin{aligned}\|\,f^{nh} - f_o\,\|_\infty &\le \|\,f^{nh} - f_h\,\|_\infty + \|\,f_h - f_o\,\|_\infty \\ &\le \|\,\mathfrak{S}_{nh}\,\|_\infty + \|\,\varepsilon^{nh}\,\|_\infty + \|\,f_h - f_o\,\|_\infty,\end{aligned}$$

and Theorem 7 takes care of the last term. □

Acknowledgment. We thank David Mason for patiently explaining the results of Einmahl and Mason [12] to us, and for straightening out the required modification of their Proposition 2.

References

[1] Abramovich, F. and Grinshtein, V. (1999). Derivation of equivalent kernels for general spline smoothing: a systematic approach. *Bernoulli* **5** 359–379.
[2] Adams, R. A. and Fournier, J. J. F. (2003). *Sobolev Spaces*, 2nd edition. Academic Press, Amsterdam.
[3] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.
[4] Boyce, W. E. and DiPrima, R. C. (1977). *Elementary Differential Equations and Boundary Value Problems*, 3rd edition. John Wiley and Sons, New York.
[5] Chiang, C., Rice, J. and Wu, C. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc.* **96** 605–619.
[6] Cox, D. D. (1984a). Asymptotics of M-type smoothing splines. *Ann. Statist.* **11** 530–551.

[7] Cox, D. D. (1984b). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

[8] Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.* **7** 225–277.

[9] Dolph, C. L. and Woodbury, M. A. (1952). On the relation between Green's functions and covariances of certain stochastic processes and its application to unbiased linear prediction. *Trans. Amer. Math. Soc.* **72** 519–550.

[10] Eggermont, P. P. B. and LaRiccia, V. N. (2006a). *Maximum Penalized Likelihood Estimation. Volume II: Regression.* Springer-Verlag, New York, in preparation.

[11] Eggermont, P. P. B. and LaRiccia, V. N. (2006b). Equivalent kernels for smoothing splines. *J. Integral Equations Appl.* **18** 197–225.

[12] Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403.

[13] Eubank, R. L. (1999). *Spline Smoothing and Nonparametric Regression.* Marcel Dekker, New York.

[14] Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301.

[15] Härdle, W., Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16** 1428–1449.

[16] Konakov, V. D. and Piterbarg, V. I. (1984). On the convergence rate of maximal deviation distribution for kernel regression estimates. *J. Multivariate Anal.* **15** 279–294.

[17] Mason, D. M. (2006). Private communication.

[18] Messer, K. and Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation, *Ann. Statist.* **21** 179–196.

[19] Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15** 780–799.

[20] Nychka, D. (1995). Splines as local smoothers. *Ann. Statist.* **23** 1175–1197.

[21] Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12** 898–916.

[22] van de Geer, S. A. (2000). *Applications of Empirical Process Theory.* Cambridge University Press.

[23] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

[24] Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 238–259

DOI: 10.1214/074921706000000888

Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results

Evarist Giné[1,*] and Vladimir Koltchinskii[1,†]

University of Connecticut and Georgia Institute of Technology

Abstract: Let M be a compact Riemannian submanifold of $\mathbf{R}^m$ of dimension d and let $X_1, \ldots, X_n$ be a sample of i.i.d. points in M with uniform distribution. We study the random operators

$$\Delta_{h_n,n} f(p) := \frac{1}{nh_n^{d+2}} \sum_{i=1}^{n} K(\frac{p - X_i}{h_n})(f(X_i) - f(p)), \ p \in M$$

where $K(u) := \frac{1}{(4\pi)^{d/2}} e^{-\|u\|^2/4}$ is the Gaussian kernel and $h_n \to 0$ as $n \to \infty$. Such operators can be viewed as graph laplacians (for a weighted graph with vertices at data points) and they have been used in the machine learning literature to approximate the Laplace-Beltrami operator of M, $\Delta_M f$ (divided by the Riemannian volume of the manifold). We prove several results on a.s. and distributional convergence of the deviations $\Delta_{h_n,n} f(p) - \frac{1}{|\mu|}\Delta_M f(p)$ for smooth functions f both pointwise and uniformly in f and p (here $|\mu| = \mu(M)$ and μ is the Riemannian volume measure). In particular, we show that for any class $\mathcal{F}$ of three times differentiable functions on M with uniformly bounded derivatives

$$\sup_{p\in M} \sup_{f\in\mathcal{F}} \left| \Delta_{h_n,p} f(p) - \frac{1}{|\mu|}\Delta_M f(p) \right| = O\Big(\sqrt{\frac{\log(1/h_n)}{nh_n^{d+2}}} \Big) \ \text{ a.s.}$$

as soon as

$$nh_n^{d+2}/\log h_n^{-1} \to \infty \ \text{ and } \ nh_n^{d+4}/\log h_n^{-1} \to 0,$$

and also prove asymptotic normality of $\Delta_{h_n,p} f(p) - \frac{1}{|\mu|}\Delta_M f(p)$ (functional CLT) for a fixed $p \in M$ and uniformly in f.

1. Introduction

Recently, there have been several developments in statistical analysis of data supported on a submanifold in a high dimensional space based on the idea of approximation of the Laplace-Beltrami operator of the manifold (and some more general operators that contain information not only about the geometry of the manifold, but also about the unknown density of data points) by empirical graph Laplacian operators. If V is a finite set of vertices and $W := (w_{ij})_{i,j\in V}$ is a symmetric nonnegative definite matrix of weights with $w_{ij} \geq 0$ ("adjacency matrix"), then the

[1]University of Connecticut, Department of Mathematics, U-3009, Storrs, CT 06269, USA, School of Mathematics, Georgia Inst. of Technology, Atlanta, GA 30332, USA, e-mail: gine@math.uconn.edu; vlad@math.gatech.edu

*Research partially supported by NSA Grant H98230-1-0075.

†Research partially supported by NSF Grant DMS-03-04861.

AMS 2000 subject classifications: primary 60F15, 60F05; secondary 53A55, 53B99.

Keywords and phrases: Laplace-Beltrami operator, graph Laplacian, large sample approximations.

graph Laplacian of the weighted graph (V, W) is defined as the matrix (operator) $L = D - W$, where D is the diagonal matrix with the degrees of vertices

$$\deg(i) := \sum_{j \in V} w_{ij},\ i \in V$$

on the diagonal. Such (unnormalized) graph Laplacians along with their normalized counterparts $\tilde{L} := I - D^{-1/2} L D^{-1/2}$ have been studied extensively in spectral graph theory. If now $X_1, \ldots, X_n$ are i.i.d. points uniformly distributed in a compact Riemannian submanifod M of $\mathbf{R}^m$ of dimension $d < m$, it has been suggested in the literature to view $\{X_1, \ldots, X_n\}$ as the set V of vertices of the graph and to define the weights as $w_{ij} \asymp e^{-\|X_i - X_j\|^2/4h^2}$ with a small parameter $h > 0$, to approximate the Laplace-Beltrami operator Δ_M of M, $\Delta_M(f) = \mathrm{div}(\mathrm{grad}(f))$. More precisely, the estimate is defined as

$$\Delta_{h_n,n} f(p) := \frac{1}{n h_n^{d+2}} \sum_{i=1}^{n} K\Big(\frac{p - X_i}{h_n}\Big)(f(X_i) - f(p)),\ p \in M$$

where $K(u) := \frac{1}{(4\pi)^{d/2}} e^{-\|u\|^2/4}$ is the Gaussian kernel and $h_n \to 0$ as $n \to \infty$ (if the functions f are restricted to V, this can be viewed, up to a sign, as a graph Laplacian operator). We will call such operators *empirical graph Laplacians* and their limit as $n \to \infty$ on smooth functions f is $\frac{1}{|\mu|}\Delta_M f(p)$, where $|\mu|$ is the Riemannian volume of M. There are numerous statistical applications of such an approximation of the manifold Laplacian by its empirical version. In particular, one can look at projections of the data on eigenspaces of the empirical Laplacian $\Delta_{h_n,n}$ (the technique sometimes called diffusion maps) in order to try to recover geometrically relevant features of the data (as in the method of spectral clustering) or use the kernels associated with this operator to approximate the heat kernel of the manifold and to use it to design kernel machines suitable, for instance, for classification of the manifold data.

Convergence properties of empirical graph Laplacians have been first studied by Belkin and Niyogi [1] and Hein, Audibert and von Luxburg [8]. Our goal in this paper is to provide a more subtle probabilistic analysis of such operators. In particular, for proper classes of smooth functions $\mathcal{F}$ and for a fixed $p \in M$, we establish a functional CLT for $\sqrt{n h_n^{d+2}}(\Delta_{h_n,p} f(p) - \frac{1}{|\mu|}\Delta_M f(p)),\ f \in \mathcal{F}$, and also show that

$$\sup_{p \in M} \sup_{f \in \mathcal{F}} \Big|\Delta_{h_n,p} f(p) - \frac{1}{|\mu|}\Delta_M f(p)\Big| = O\bigg(\sqrt{\frac{\log(1/h_n)}{n h_n^{d+2}}}\bigg) \text{ a.s.}$$

(under suitable assumptions on h_n). The asymptotic properties of empirical laplacians are closely related to the well developed theory of kernel density and kernel regression estimators, which can be viewed as examples of so called local empirical processes, as in [6]. Our proofs are essentially based on an extension of this type of results to the case of data on the manifolds (for kernel density estimation on manifolds, see, e.g., [11] and references therein). For simplicity, we are considering in the current paper only uniform distributions on manifolds and Gaussian kernels K, but more general types of operators that occur in the case when the distribution of the data is not uniform and more general kernels (as in the paper of Hein, Audibert and von Luxburg [8]) can be dealt with quite similarly using the methods of the paper.

2. Some geometric background

We refer to [4] for the basic definitions and notations from Riemannian geometry. Given a manifold M and $p \in M$, $T_p(M)$ will denote the tangent space to M at p, and TM the tangent bundle. Let M be a complete connected (embedded) Riemannian submanifold of $\mathbf{R}^m$, of dimension $d < m$, meaning that M is a complete connected Riemannian manifold and that the inclusion map $\phi : M \mapsto \mathbf{R}^m$ is an isometric embedding, that is, (i) ϕ is differentiable and injective, (ii) $d\phi_p : T_p(M) \mapsto T_{\phi(p)}(\mathbf{R}^m)$ is an isometry onto its image, $T_{\phi(p)}(\phi(M))$, and (iii) ϕ is a homeomorphism onto $\phi(M)$ with the topology inherited from $\mathbf{R}^m$. When no confusion may arise, we identify M with $\phi(M)$. M being complete, by the Hopf and Rinow theorem (e.g., [4], p. 146) the closed bounded sets of M are compact.

Given $p \in M$ and $v \in T_p(M)$, let $\gamma(t, p, v)$, $t > 0$, be the geodesic starting at p with velocity v, $\gamma(0, p, v) = p$ and $\gamma'(0, p, v) = v$. The exponential map $\mathcal{E}_p : T_p(M) \mapsto M$ (the usual notation is $\exp_p$) is defined by $\mathcal{E}_p(v) = \gamma(1, p, v)$. This map is defined on all of $T_p(M)$ by the Hopf and Rinow theorem.

A normal neighborhood V of $p \in M$ is one for which a) every point q in V can be joined to p by a *unique* geodesic $\gamma(t, p, v)$, $0 \le t \le 1$, with $\gamma(0) = p$, $\gamma(1) = q$ and b) the exponential map centered at p, $\mathcal{E}_p$, is a diffeomorphism between a neighborhood of $0 \in T_p(M)$ and V. If $B \subset V$ is a normal ball of center p, that is, the image by the exponential map of a ball around zero in $T_p(M)$, then the unique geodesic joining p to $q \in B$ is a minimizing geodesic, which means that if d_M denotes the distance in M and $|\cdot|$ denotes the norm of $T_p(M)$ defined by the Riemannian structure of M, then $d_M(p, \mathcal{E}_p(v)) = |v|$. Given an orthonormal basis $e_1, \dots, e_d$ of $T_p(M)$, the normal coordinates centered at p (or the p-normal coordinates) of $q \in V$ are the components $q_i^p = \langle \mathcal{E}_p^{-1}(q), e_i \rangle$ of $\mathcal{E}_p^{-1}(q)$ in this basis. (The super-index p will be omitted when no confusion may arise, but we will need it when considering normal coordinates based at different points.) Every point in M has a normal neighborhood. See [4], Propositions 2.7, and 3.6, pp. 64 and 70 for these facts. Actually, more is true (e.g., [4], Theorem 3.7 and Remark 3.8):

Proposition 2.1. *For every $p \in M$ there exist a neighborhood W of p and a number $\delta > 0$ such that:* (a) *for every q in W, $\mathcal{E}_q$ is defined on the δ ball around $0 \in T_q(M)$, $B_\delta(0) \subset T_q(M)$, and $\mathcal{E}_q(B_\delta(0))$ is a normal ball for q,* (b) $W \subset \mathcal{E}_q(B_\delta(0))$ *(W is a normal neighborhood of all of its points), and* (c) *the function*

$$F(q, v) := (q, \mathcal{E}_q(v))$$

is a diffeomorphism from $W_\delta := W \times B_\delta(0) = \{(q, v) \in TM : q \in W, |v| < \delta\}$ onto its image in $M \times M$ and $|dF|$ is bounded away from zero on W_δ.

Such a neighborhood W of p is called *totally or uniformly normal*. In particular, $\mathcal{E}_q(v)$ is jointly differentiable in $(q, v) \in W_\delta$ if W is a uniformly normal neighborhood. Moreover, for every $q \in W$ and $v \in T_q(M)$ such that $|v| < \delta$, $d_M(q, \mathcal{E}_q(v)) = |v|$.

Remark 2.1. By shrinking W and taking $\delta/2$ instead of δ if necessary, we can assume in Proposition 2.1 that the closure of W and the closure of W_δ (which are compact because M is complete) are contained in W' and $W'_{\delta'}$ satisfying the properties described in the previous proposition. Moreover, we can also assume that for all q in W, $\mathcal{E}_q(B_\delta(0))$ is contained in a strongly convex normal ball around p (these points are at distances less than 2δ from p, so this assumption can be met by further shrinking W and taking a smaller δ if necessary, since every point in M has a strongly convex geodesic ball, e.g. Proposition 4.2 in do Carmo, loc. cit.; strongly convex set: for any two points in the set, the minimizing geodesic joining them lies in the set). We will assume without loss of generality and without further mention that our uniformly normal neighborhoods W satisfy these two conditions.

Let W be a uniformly normal neighborhood of p as in the remark, let W' be a uniformly normal neighborhood of p containing the closure of W, and let $0 < \delta < \delta'$ be as in the proposition and the remark. Let us choose an orthonormal basis $e_1, \dots, e_d$ of $T_p(M)$ and define an orthonormal frame $e_1^q, \dots, e_d^q$, $q \in W'$, by parallel transport of $e_1, \dots, e_d$ from p to q along the unique minimizing geodesic joining p and q. So, $e_1^q, \dots, e_d^q$ is an orthonormal basis of $T_q(M)$ for each $q \in W'$. This frame depends differentiably on q as parallel transport is differentiable (and preserves length and angle). So, we have on W' a system of normal coordinates centered at q for every $q \in W'$, namely, if $x \in \mathcal{E}_q(B_{\delta'}(0))$ is $x = \mathcal{E}_q(\sum_{i=1}^d v_i e_i^q)$, then the coordinates of x, x_i^q are $x_i^q = v_i$, the components of $\mathcal{E}_q^{-1}(x)$. Let now f be a differentiable function $f : M \mapsto \mathbf{R}$, and define $\tilde{f} : W'_{\delta'} \mapsto \mathbf{R}$ by

$$\tilde{f}(q, v) := f(\pi_2(F(q, v)) = f(\mathcal{E}_q(v)),$$

where π_2 is the projection of $M \times M$ onto its second component. This map is differentiable by the previous proposition. In particular, if we take as coordinates of $(q, v) \in W'_{\delta'}$ the normal coordinates centered at p of q, $(q_1, \dots, q_d) = (q_1^p, \dots, q_d^p)$ and for $v \in T_q(M)$ the coordinates $v_1, \dots v_d$ in the basis $e_1^q, \dots, e_d^q$, which coincide with the normal coordinates centered at q of $\mathcal{E}_q(v)$, then the real function of $2d$ variables (which we keep calling $\tilde{f}$; the same convention applies to other similar cases below)

$$\tilde{f}(q_1, \dots, q_d, v_1, \dots, v_d) = \tilde{f}(q, v)$$

is differentiable on the preimage of $W'_{\delta'}$ by this system of coordinates. Moreover, by compactness, each of its partial derivatives (of any order) is uniformly bounded on the preimage of W_δ. If we denote by x_i^q the normal coordinates centered at q, we obviously have that for each $r \in \mathbf{N}$ and $(i_1, \dots, i_r) \in \{1, \dots, d\}^r$,

$$\frac{\partial^r f}{\partial x_{i_1}^q \partial x_{i_2}^q \dots \partial x_{i_r}^q}(x) = \frac{\partial^r \tilde{f}}{\partial v_{i_1} \partial v_{i_2} \dots \partial v_{i_r}}(q_1, \dots, q_d, x_1^q, \dots, x_d^q).$$

We then conclude that

$$\text{(2.1)}\quad \begin{array}{c} \textit{each of the partial derivatives (any order) of } f \textit{ with respect to the} \\ q-\textit{normal coordinates } x_i^q \textit{ is uniformly bounded in } q \in W \textit{ and } x \in \mathcal{E}_q(B_\delta(0)). \end{array}$$

In particular, the error term in any limited Taylor development of f in q-normal coordinates can be bounded uniformly in q for all $|v| < \delta$, that is, if $P_k^q(x_1^q, \dots, x_d^q)$ is the Taylor polynomial of degree k in these coordinates, we have, for $q \in W$ and $|\mathcal{E}_q^{-1}(x)| < \delta$,

$$|f(x) - P_k^q(x_1^q, \dots, x_d^q)| \le C_k (d_M(q, x))^{k+1}, \tag{2.2}$$

where C_k is a constant that depends only on k. Moreover, the coefficients of the polynomials P_k^q are differentiable functions of q, in particular bounded on W. The q-uniformity of these Taylor developments for $|v| < \delta$ and the q-differentiability of their coefficients will be very useful below. We will apply these properties to the canonical (in $\mathbf{R}^m$) coordinates of the embedding ϕ and also to the functions $\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\rangle(x)$, where $x_i = x_i^p$ are the p-normal coordinates.

In what follows, we often deal with classes $\mathcal{F}$ of functions on M whose partial derivatives up to a given order k are uniformly bounded in M or in a neighborhood U of a point $p \in M$. In such cases, we say that $\mathcal{F}$ is *unformly bounded up to the k-th order* in M (or in U). Clearly, this property does not depend on the choice of normal (or even arbitrary) local coordinates. In the case when we choose an orthonormal frame $e_1^q, \ldots, e_d^q$ and define normal coordinates and the corresponding partial derivatives as described above, we can also deal with continuity of partial derivatives. We say that $\mathcal{F}$ is a *uniformly bounded and equicontinuous up to the k-th order* class of functions iff there exists a finite covering of M with uniformly normal neighborhoods such that, in each neighborhood, the sets of partial derivatives of any order $\le k$ are uniformly bounded and equicontinuous. This definition does not depend on the choice of orthonormal frames in the neighborhoods. Such classes are useful because the remainders in Taylor developments are uniform both in $q \in M$ and in $f \in \mathcal{F}$.

Consider now, for $q \in W'$ and $x \in \mathcal{E}_q(B_{\delta'}(0))$, the tangent vector fields $\frac{\partial}{\partial x_i^q}(x)$, $i = 1, \ldots, d$, and simply write $\frac{\partial}{\partial x_i}(x)$ for $\frac{\partial}{\partial x_i^p}(x)$. Taking the previous coordinates q_i, v_j in $W'_{\delta'}$, denote by $\chi_i(\mathcal{E}_q(v)) = \chi_i(q_1, \ldots, q_d, v_1, \ldots, v_d)$ the p-normal coordinates of $\mathcal{E}_q(v)$, which are differentiable. By the chain rule,

$$\frac{\partial}{\partial x_i^q}(x) = \sum_{j=1}^{d} \frac{\partial \chi_j}{\partial v_i}(q, \mathcal{E}_q^{-1}x)\frac{\partial}{\partial x_j}(x).$$

Hence, if $g_{ij}^q(x)$ are the components of the metric tensor at x in q-normal coordinates, we have

$$g_{ij}^q(x) = \langle \frac{\partial}{\partial x_i^q}, \frac{\partial}{\partial x_j^q}\rangle(x) = \sum_{1\le r,s\le d} \frac{\partial \chi_r}{\partial v_i}(q, \mathcal{E}_q^{-1}x)\frac{\partial \chi_s}{\partial v_j}(q, \mathcal{E}_q^{-1}x)\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\rangle(x).$$

By (2.2), we conclude that if $P_k^q(x_1^q, \ldots, x_d^q)$ is the Taylor polynomial of degree k in the expansion of $g_{ij}^q(x)$ in q-normal coordinates, then there are constants C_k that depend only on k such that, for all $q \in W$ and $x \in \mathcal{E}_q(B_\delta(0))$,

$$|g_{ij}^q(x) - P_k^q(x_1^q, \ldots, x_d^q)| \le C_k (d_M(q, x))^{k+1}. \tag{2.3}$$

This will also be useful below. These remarks allow us to strengthen several results based on Taylor expansions by making them uniform in q, as follows.

Proposition 2.2. *Given $p \in M$, let W and W_δ be as in Remark 2.1, and consider for each $q \in W$, the q-normal system of coordinates defined above. Then,*
(a) *for every $q \in W$ the components $g_{ij}^q(x_1^q, \ldots, x_d^q)$ of the metric tensor in q-normal coordinates admit the following expansion, uniform in q and in $x \in \mathcal{E}_q(B_\delta(0))$ ($B_\delta(0) \in T_q(M)$):*

$$g_{ij}^q(x_1^q, \ldots, x_d^q) = \delta_{ij} - \frac{1}{3}R_{irsj}^q(0)x_r^q x_s^q + O(d_M^3(q, x)), \tag{2.4}$$

(Einstein notation) where $R_{irsj}^q(0)$ are the components of the curvature tensor at q in q-normal coordinates, and, as a consequence, the following expansion of the volume element is also uniform in q and x:

$$\sqrt{\det(g_{ij}^q)}(x_1^q, \ldots, x_d^q) = 1 - \frac{1}{6}\mathrm{Ric}_{rs}^q(0)x_r^q x_s^q + O(d_M^3(q, x)), \tag{2.5}$$

where $\mathrm{Ric}^q_{rs}(0)$ *are the components of the Ricci tensor at* q *in* q*-normal coordinates.*
(b) *There exists* $C<\infty$ *such that for all* $q\in W$ *and* $x\in\mathcal{E}_q(B_\delta(0))$,

$$0\le d_M^2(q,x)-\|\phi(q)-\phi(x)\|^2\le Cd_M^4(q,x). \tag{2.6}$$

(c) *For each* $1\le\alpha\le m$, *the* α*-th component in canonical coordinates of* $\mathbf{R}^m$ *of* $\phi(\mathcal{E}_q(v))$, $\phi_\alpha(\mathcal{E}_q(v))$, *admits the following expansion in* q*-normal coordinates* v_i *of* $\mathcal{E}_q(v)$, *uniform in* $q\in W$ *and* $|v|<\delta$,

$$\phi_\alpha(\mathcal{E}_q(v))-\phi_\alpha(q)=\frac{\partial\tilde\phi_\alpha}{\partial v_i}(q,0)v_i+\frac{1}{2}\frac{\partial^2\tilde\phi_\alpha}{\partial v_i\partial v_j}(q,0)v_iv_j+O(|v|^3), \tag{2.7}$$

$\alpha=1,\dots,m$, *where* $\tilde\phi(q,v)=\phi(\mathcal{E}_q(v))$.

Note that $\sum_{i=1}^d\frac{\partial\tilde\phi_\alpha}{\partial v_i}(q,0)v_i$ are the $\mathbf{R}^m$-canonical coordinates centered at $\phi(q)$ of the vector $d\phi_q(v)\in T_{\phi(q)}(\phi(M))\subset\mathbf{R}^m$ since $g^q_{ij}(q)=\delta_{ij}$. Hence, if we identify the tangent space to $\phi(M)$ at $\phi(q)$ with an affine subspace of $\mathbf{R}^m$, part c) says that the difference between $\phi(\mathcal{E}_q(v))\in\mathbf{R}^m$ and the tangent vector to the geodesic $\phi(\gamma(t,q,v))$ at $\phi(q)$ $(t=0)$, $\phi(q)+d\phi_q(v)$, is a vector of the form

$$\Big(\frac{1}{2}\sum_{1\le i,j\le d}\frac{\partial^2\tilde\phi_\alpha}{\partial v_i\partial v_j}(q,0)v_iv_j:\alpha=1,\dots,m\Big)+O(|v|^3)$$

where $O(|v|^3)$ is uniform in $|v|<\delta$ and q. Ignoring the embedding, this gives an expansion of the exponential map as

$$\mathcal{E}_q(v)=q+v+Q_q(v,v)+O(|v|^3) \tag{2.7$'$}$$

uniform in $q\in W$ and $|v|<\delta$, where Q_q is a $\mathbf{R}^m$-valued bilinear map on $T_q(M)$ (actually, on $T_{\phi(q)}(\phi(M))$) that depends differentiably on q, hence uniformly bounded in $q\in W$.

Proof of Proposition 2.2. (a) follows from the expansions of g^q_{ij} and $\sqrt{\det(g^q_{ij})}$ in q-normal coordinates (e.g. in [12], p. 41), the expansion of its determinant (e.g., [12], p. 45), and the uniformity provided by (2.3).

(c) follows by direct application of (2.2) to $f=\phi_\alpha$, $\alpha=1,\dots,m$.

(b) Following Smolyanov, Weizsäcker and Wittich [13], for $q\in W$ and $x=\mathcal{E}_q(v)$, $|v|<\delta$, and applying (2.2) for $f=\phi_\alpha$, $\alpha=1,\dots,m$, we have

$$\begin{aligned}0\le\frac{d_M^2(q,x)-\|\phi(q)-\phi(x)\|^2}{d_M^4(q,x)}&=\frac{|v|^2-\sum_{\alpha=1}^m\big(\phi_\alpha(\mathcal{E}_q(v))-\phi_\alpha(q)\big)^2}{|v|^4}\\ &=\frac{|v|^2-\sum_{\alpha=1}^m\big(\frac{\partial\tilde\phi_\alpha}{\partial v_i}(q,0)v_i+\frac{1}{2}\frac{\partial^2\tilde\phi_\alpha}{\partial v_i\partial v_j}(q,0)v_iv_j+\frac{1}{6}\frac{\partial^3\tilde\phi_\alpha}{\partial v_i\partial v_j\partial v_k}(q,0)v_iv_jv_k\big)^2}{|v|^4}\\ &\quad+O(|v|),\end{aligned}$$

where the term $O(|v|)$ is dominated by $C_4|v|$ for a constant C_4 that does not depend on q or v. But now, continuing the proof in this reference, which consists in developing and simplifying the ratio above, we obtain that, uniformly in $q\in W$, $x=\mathcal{E}_q(v)$, $|v|<\delta$,

$$0\le\frac{d_M^2(q,x)-\|\phi(q)-\phi(x)\|^2}{d_M^4(q,x)}=\frac{1}{12}\sum_\alpha\frac{\big(\frac{\partial^2\tilde\phi_\alpha}{\partial v_i\partial v_j}(q,0)v_iv_j\big)^2}{|v|^4}+O(|v|),$$

and note also that, by compactness, the main term is bounded by a fixed finite constant in this domain. □

Although we have been using [4] as our main reference on Riemannian geometry, another nice user-friendly reference for the exponential map in particular and Riemannian manifolds in general is [9]. We thank Jesse Ratzkin for reading this section and making comments (of course, any mistakes are ours).

3. Approximation of the Laplacian by averaging kernel operators

Let M be a compact connected Riemannian submanifold of $\mathbf{R}^m$, $m > d$ (if M is compact, it is automatically embedded, that is, conditions (i) and (ii) on the immersion ϕ imply that ϕ is a homeomorphism onto its image). [See a remark at the end of this section for a relaxation of this condition.] Let μ be its Riemannian volume measure and $|\mu| = \mu(M)$. Let $K : \mathbf{R}^m \mapsto \mathbf{R}$ be the Gaussian kernel of $\mathbf{R}^m$,

$$K(x) = \frac{1}{(4\pi)^{d/2}} e^{-\|x\|^2/4}, \tag{3.1}$$

where $\|x\|$ is the norm of x in $\mathbf{R}^m$. Let X be a random variable taking values in M with law the normalized volume element, $\mu/|\mu|$, and let $f : M \mapsto \mathbf{R}$ be a differentiable function. The object of this section is to show that the Laplace-Beltrami operator or Laplacian of M,

$$\Delta_M f(p) = \text{div grad}(f)(p)$$

(in coordinates, $\Delta_M(f) = \frac{1}{\sqrt{\det(g_{ij})}} \frac{\partial}{\partial x_i}(g^{ij}\sqrt{\det(g_{ij})}\frac{\partial f}{\partial x_j})$, where $(g^{ij}) = (g_{ij})^{-1}$) can be approximated, uniformly in f (with some partial derivatives bounded), and in $p \in M$, by the *averaging kernel operator*

$$\Delta_{h_n} f(p) := \frac{1}{h_n^{d+2}} E\left[K\left(\frac{\phi(p) - \phi(X)}{h_n}\right)(f(X) - f(p))\right] \tag{3.2}$$

with rates depending on $h_n \to 0$. Note that, by the expansion (2.4) of the metric tensor in normal coordinates centered at p, we have, in these coordinates,

$$\Delta_M f(p) = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2}(p). \tag{3.3}$$

(where $p = (0, \dots, 0)$ in these coordinates).

With some abuse of notation, given $p \in M$, we denote the derivatives with respect to the components of v of $\tilde{f}(p, v) = f \circ \mathcal{E}_p(v)$ at (p, v), $v = \mathcal{E}_p^{-1}(x)$, by $f'(x)$, $f''(x)$, etc. (so, for instance, if $x = \mathcal{E}_p(v)$, $f'(x) = \left(\frac{\partial \tilde{f}}{\partial v_1}(p, v), \dots, \frac{\partial \tilde{f}}{\partial v_d}(p, v)\right)$) (in fact, $f^{(k)}(x)$ depends on p and therefore it should have been denoted $f_p^{(k)}(x)$, but in the context we are using this notation p is typically fixed, so, we will drop p, hopefully, without causing a confusion).

Theorem 3.1. *We have, for any p, any normal neighborhood U_p of p and a class $\mathcal{F}$ uniformly bounded up to the third order in U_p, that*

$$\sup_{f \in \mathcal{F}} \left|\Delta_{h_n} f(p) - \frac{1}{|\mu|}\Delta_M f(p)\right| = O(h_n). \tag{3.4}$$

as $h_n \to 0$. Moreover, for any class of functions uniformly bounded up to the third order in M,

$$\sup_{f\in\mathcal{F}}\sup_{p\in M}\left|\Delta_{h_n}f(p)-\frac{1}{|\mu|}\Delta_M f(p)\right| = O(h_n). \tag{3.5}$$

as $h_n \to 0$.

Proof. M being regular, the embedding ϕ is a homeomorphism of M onto $\phi(M)$, and M being compact, the uniformities defined respectively on M by the intrinsic metric $d_M(p,q)$ and by the metric from $\mathbf{R}^m$, $d_{\mathbf{R}^m}(p,q) := \|\phi(p)-\phi(q)\|$ coincide (e.g., Bourbaki (1940), Theorem II.4.1, p. 107), that is, given $\varepsilon > 0$ there exists $\delta > 0$ such that if $d_M(p,q) < \delta$ for $p, q \in M$, then $d_{\mathbf{R}^m}(p,q) < \varepsilon$, and conversely. Hence, in Proposition 2.2, we can replace $B_\delta(0) \subset T_q(M)$ by $B'_{\delta'}(0) := \mathcal{E}_q^{-1}\{x \in M : \|\phi(q)-\phi(x)\| < \delta'\}$ for some δ' depending on δ but not on p or q. From here on, we identify M with $\phi(M)$ (that is, we leave ϕ implicit). Let $p \in M$. Given $h_n \searrow 0$, let

$$\mathcal{B}_n := \{x \in M : \|p-x\| < Lh_n(\log h_n^{-1})^{1/2}\} \tag{3.6}$$

for a constant L to be chosen later. As soon as $Lh_n(\log h_n^{-1})^{1/2} < \delta'$, the neighborhood of $0 \in T_p(M)$,

$$\tilde{\mathcal{B}}_n := \mathcal{E}_p^{-1}\mathcal{B}_n$$

is well defined, and, by (2.6), since $|v| = d_M(p, \mathcal{E}_p(v))$, we have on $\tilde{B}_n$ that

$$|v|^2 \geq \|p-\mathcal{E}_p(v)\|^2 \geq |v|^2(1-C|v|^2)$$

with C independent of $p \in M$. Hence, for all $n \geq N_0$, for some $N_0 < \infty$ independent of p, we have

$$\begin{aligned}\{v \in T_pM : |v| < Lh_n(\log h_n^{-1})^{1/2}\} &\subseteq \tilde{\mathcal{B}}_n \\ &\subseteq \{v \in T_pM : |v| < 2Lh_n(\log h_n^{-1})^{1/2}\},\end{aligned} \tag{3.7}$$

where the coefficient 2 can be replaced by $\lambda_n \to 1$. Assume $n \geq N_0$.

By the definitions of K and $\mathcal{B}_n$,

$$\begin{aligned}&E\left|K\left(\frac{p-X}{h_n}\right)(f(X)-f(p))I(X \in M\setminus\mathcal{B}_n)\right| \\ &\quad\leq \frac{2\|f\|_\infty}{(4\pi)^{d/2}}\int_{M\setminus\mathcal{B}_n} e^{-\|p-x\|^2/4h_n^2}\frac{d\mu(x)}{|\mu|} \\ &\quad\leq \frac{2\|f\|_\infty}{(4\pi)^{d/2}}h_n^{L^2/4}.\end{aligned} \tag{3.8}$$

Taking into account that the measure μ has density $\sqrt{\det(g_{ij})}$ in p-normal coordinates (hence on $\mathcal{B}_n$), we have

$$\begin{aligned}&E\left[K\left(\frac{p-X}{h_n}\right)(f(X)-f(p))I(X \in \mathcal{B}_n)\right] \\ &\quad= \frac{1}{(4\pi)^{d/2}|\mu|}\int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/4h_n^2}(f(\mathcal{E}_p(v))-f(\mathcal{E}_p(0)))\sqrt{\det(g_{ij})}(v)dv.\end{aligned} \tag{3.9}$$

With the notation introduced just before the statement of the theorem, the Taylor expansion of f in p-normal coordinates can be written as

$$f(\mathcal{E}_p(v)) - f(\mathcal{E}_p(0)) = \langle f'(p), v\rangle + \frac{1}{2}f''(p)(v,v) + \frac{1}{3!}f'''(\xi_v)(v,v,v).$$

where $\xi_v = \mathcal{E}_p(\theta_v v)$ for some $\theta_v \in [0,1]$. Next we will estimate the three terms that result from combining this Taylor development with equation (3.9). Recall that, by Proposition 2.2, there are C_1 and C independent of p such that

$$\sqrt{\det(g_{ij})}(v) \le 1 + C_1|v|^2, \quad \frac{1}{2}|v|^2 \le |v|^2 - C|v|^4 \le \|p - \mathcal{E}_p(v)\|^2 \le |v|^2 \tag{3.10}$$

for $v \in \tilde{\mathcal{B}}_n$, and recall also (3.7) on the size of $\tilde{\mathcal{B}}_n$. Using these facts and the development of the exponential about $-|v|^2/4h_n^2$ immediately gives

$$\int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/4h_n^2}\langle f'(p), v\rangle \sqrt{\det(g_{ij})}(v)dv = \int_{\tilde{\mathcal{B}}_n} e^{-|v|^2/4h_n^2}\langle f'(p), v\rangle dv + R_n$$

where

$$\begin{aligned}
|R_n| &\le \int_{\tilde{\mathcal{B}}_n} \left(e^{-(|v|^2-C|v|^4)/4h_n^2} - e^{-|v|^2/4h_n^2}\right)|f'(p)||v|dv \\
&\quad + C_1 \int_{\tilde{\mathcal{B}}_n} e^{-|v|^2/8h_n^2}|f'(p)||v|^3 dv \\
&\le \int_{\tilde{\mathcal{B}}_n} e^{-|v|^2/8h_n^2}|f'(p)|(C|v|^5/(4h_n^2) + C_1|v|^3)dv \\
&\le h_n^{3+d} \int_{\mathbf{R}^d} e^{-|v|^2/8}|f'(p)|(C|v|^5/4 + C_1|v|^3)dv \\
&= D|f'(p)|h_n^{3+d},
\end{aligned}$$

and D only depends on C, C_1 and d. Moreover, since $\mathcal{B}_n^c \subseteq \{|v| \ge Lh_n(\log h_n^{-1})^{1/2}\}$ and

$$\int_{\mathbf{R}^d} e^{-|v|^2/4h_n^2}\langle f'(p), v\rangle dv = 0,$$

we also have

$$\begin{aligned}
\left|\int_{\tilde{\mathcal{B}}_n} e^{-|v|^2/4h_n^2}\langle f'(p), v\rangle dv\right| &= \left|\int_{\tilde{\mathcal{B}}_n^c} e^{-|v|^2/4h_n^2}\langle f'(p), v\rangle dv\right| \\
&\le |f'(p)| \int_{|v|\ge Lh_n(\log h_n^{-1})^{1/2}} e^{-|v|^2/4h_n^2}|v|dv \\
&= |f'(p)|h_n^{1+d} \int_{|u|\ge L(\log h_n^{-1})^{1/2}} e^{-|u|^2/4}|u|du \\
&= C_d|f'(p)|h_n^{1+d} \int_{r\ge L(\log h_n^{-1})^{1/2}} e^{-r^2/4}r^d dr \\
&\le C_d'|f'(p)|L^{d-1}h_n^{1+d+L^2/4}(\log h_n^{-1})^{(d-1)/2},
\end{aligned}$$

where C_d and C_d' are constants depending only on d. Collecting terms and assuming

$$L^2/4 > 2,$$

we obtain

$$\left|\int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/4h_n^2}\langle f'(p), v\rangle \sqrt{\det(g_{ij})}(v)dv\right| \le D_2|f'(p)|h_n^{3+d}, \tag{3.11}$$

for all $n \ge N_0$, and where D_2 does not depend on p.

The remainder term is of a similar order if $|f'''|$ is uniformly bounded in a neighborhood of p: if c is such a bound,

$$\begin{aligned}(3.12)\qquad &\left|\int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/4h_n^2} f'''(\xi_v)(v,v,v)\sqrt{\det(g_{ij})}(v)dv\right| \\ &\qquad \le c\int_{\tilde{\mathcal{B}}_n} e^{-|v|^2/8h_n^2}|v|^3|1+C_1h_n^2\log h_n^{-1}|dv \le D_3ch_n^{3+d},\end{aligned}$$

where D_3 does not depend on f or p (as long as $n \ge N_0$).

Finally, we consider the second term, which is the one that gives the key relationship to the Laplacian. Proceeding as we did for the first term, we see that

$$\begin{aligned}(3.13)\qquad &\int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/4h_n^2} f''(p)(v,v)\sqrt{\det(g_{ij})}(v)dv \\ &\qquad = h_n^{d+2}\int_{\mathbf{R}^d} e^{-|v|^2/4} f''(p)(v,v)dv + R_n,\end{aligned}$$

where now

$$(3.14)\quad |R_n| \le D_4|f''(p)|h_n^{4+d} + D_5|f''(p)|h_n^{2+d+L^2/4}(\log h_n^{-1})^{d/2} \le D_6|f''(p)|h_n^{4+d}$$

if

$$L^2/4 > 2$$

and $n \ge N_0$, where the constants D do not depend on f or p. Now, by definition

$$f''(p) = (f\circ\mathcal{E}_p)''(0) = \left(\frac{\partial^2(f\circ\mathcal{E}_p)}{\partial v_i\partial v_j}(0)\right)_{i,j=1}^d = \left(\frac{\partial^2 f}{\partial x_i\partial x_j}(p)\right)_{i,j=1}^d,$$

so that, on account of (3.3),

$$\begin{aligned}(3.15)\qquad \int_{\mathbf{R}^d} e^{-|v|^2/4} f''(p)(v,v)dv &= \left(\int_{\mathbf{R}^d} e^{-|v|^2/4}v_1^2dv\right)\sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}(p) \\ &= 2(4\pi)^{d/2}\sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}(p) = 2(4\pi)^{d/2}\Delta_M f(p).\end{aligned}$$

Combining the bounds (3.11), (3.12), (3.13)-(3.14) and the identity (3.15) with (3.9), we obtain the first part of the theorem. Note that we need to choose L such that $L^2/4 > 2$ and then N_0 such that $Lh_{N_0}(\log h_{N_0}^{-1})^{1/2} < \delta$, and that with these choices the bounds obtained on the terms that tend to zero in the proof depend only on the sup of the derivatives of f and on the sup of certain differentiable functions of (q,v) on W_δ, where W is a uniformly normal neighborhood of p and δ the corresponding number from Proposition 2.2 and Remark 2.1. These bounds are the same if we replace p by any $q \in W$ by Proposition 2.2. M being compact, it can be covered by a finite number of uniformly normal neighborhoods W_i, $i \le k$, with numbers δ_i as prescribed in Proposition 2.1 and Remark 2.1. Taking δ to be the minimum of δ_i, $i = 1,\dots,k$, and the constants in the bounds in the first part of the proof as the maximum of the constants in these bounds for each of the k neighborhoods, the above estimates work uniformly on $q \in M$, giving the second part of the theorem. $\square$

Remark 3.1. (1) Obviously, the first part of the theorem, namely the limit (3.4), does not require the manifold M to be compact. (2) If instead of assuming existence and boundedness of the third order partial derivatives in a neighborhood of p we assume that the second order derivatives are continuous in a neighborhood of p, then we can proceed as in the above proof except for the remainder term (3.12), that now can be replaced by

$$\left| \int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/4h_n^2} (f''(\xi_v) - f''(p))(v,v)\sqrt{\det(g_{ij})(v)}dv \right| = o(h_n^{2+d}). \tag{3.16}$$

Hence, in this case we still have

$$\Delta_{h_n} f(p) \to \frac{1}{|\mu|}\Delta_M f(p) \quad \text{as} \quad h_n \to 0. \tag{3.17}$$

A similar observation can be made regarding (3.5).

Remark 3.2. Suppose N is a compact Riemannian d-dimensional submanifold of $\mathbf{R}^m$ *with boundary* (for the definition, see [12], p. 70-71). The Riemannian volume measure μ is still finite. Then, Theorem 3.1 is still true if X is a N-valued random variable with law $\mu/|\mu|$ with $|\mu| = \mu(N)$, and M a compact subset of N interior to N. The proof is essentially the same.

The first part of Theorem 3.1, without uniformity in f, is proved in a more general setting in [8].

Theorem 3.1 provides the basis for the estimation of the Laplacian of M by independent sampling from the space according to the normalized volume element, which is what we do for the rest of this article.

4. Pointwise approximation of the Laplacian by graph Laplacians

Let M be a compact Riemannian submanifold of $\mathbf{R}^d$ (or, in more generality, let M be as in Remark 3.2), and let X, X_i, $i \in \mathbf{N}$, be independent identically distributed random variables with law $\mu/|\mu|$. The 'empirical counterpart' of the averaging kernel operator from Section 3 corresponding to such a sequence is the so called *graph Laplacian*

$$\Delta_{h_n,n} f(p) := \frac{1}{nh_n^{d+2}} \sum_{i=1}^n K\left(\frac{p - X_i}{h_n}\right)(f(X_i) - f(p)), \tag{4.1}$$

with K given by (3.1) (other kernels are possible).

We begin with the pointwise central limit theorem for a single function f, as a lemma for the CLT uniform in f.

Proposition 4.1. *Assume f has partial derivatives up to the third order continuous in a neighborhood of p. Let $h_n \to 0$ be such that $nh_n^d \to \infty$ and $nh_n^{d+4} \to 0$. Then,*

$$\sqrt{nh_n^{d+2}}\Big[\Delta_{h_n,n} f(p) - \frac{1}{|\mu|}\Delta_M f(p)\Big] \to sg \quad \text{in distribution,} \tag{4.2}$$

where g is a standard normal random variable and

$$s^2 = \frac{1}{(4\pi)^d|\mu|}\int_{\mathbf{R}^d} e^{-|v|^2/2}\Big(\sum_{j=1}^d \frac{\partial f}{\partial x_j^p}(p)v_j\Big)^2 dv = \frac{1}{2^d(2\pi)^{d/2}|\mu|}\sum_{j=1}^d \Big|\frac{\partial f}{\partial x_j^p}(p)\Big|^2. \tag{4.3}$$

Proof. Since by Theorem 3.1,

$$\sqrt{nh_n^{d+2}}\Big(\Delta_M f(p) - \Delta_{h_n} f(p)\Big) = O\Big(\sqrt{nh_n^{d+4}}\Big) \to 0,$$

it suffices to prove that the sequence

$$Z_n = \sqrt{nh_n^{d+2}}\left[\frac{1}{nh_n^{d+2}}\sum_{i=1}^n \left(K\left(\frac{p-X_i}{h_n}\right)(f(X_i)-f(p)) - EK\left(\frac{p-X}{h_n}\right)(f(X_i)-f(p))\right)\right] \tag{4.4}$$

is asymptotically centered normal with variance s^2. To prove this we first observe that we can restrict to $X_i \in \mathcal{B}_n$ because, as in (3.8),

$$\begin{aligned}&\frac{1}{h_n^{d+2}}E\left[K^2\left(\frac{p-X_i}{h_n}\right)(f(X_i)-f(p))^2 I(X\in M\setminus\mathcal{B}_n)\right]\\&\quad\le \frac{4\|f\|_\infty^2}{(4\pi)^d h_n^{d+2}}\int_M e^{-L^2(\log h_n^{-1})/2}d\mu/|\mu| = \frac{4\|f\|_\infty^2}{(4\pi)^d}h_n^{L^2/2-(d+2)} \to 0\end{aligned} \tag{4.5}$$

if we take $L^2/2 > d+2$. Now, on the restriction to $\mathcal{B}_n$ we replace $f(X_i)-f(p)$ by its Taylor expansion up to the second order plus remainder, as done in the proof of Theorem 3.1. The second term and the remainder parts, namely

$$\begin{aligned}Z_{n,2} := \sqrt{nh_n^{d+2}}\Bigg[\frac{1}{nh_n^{d+2}}\sum_{i=1}^n\Bigg(&K\left(\frac{p-X_i}{h_n}\right)f''(p)(\mathcal{E}_p^{-1}(X_i),\mathcal{E}_p^{-1}(X_i))I(X_i\in\mathcal{B}_n)\\&- EK\left(\frac{p-X}{h_n}\right)f''(p)(\mathcal{E}_p^{-1}(X),\mathcal{E}_p^{-1}(X))I(X\in\mathcal{B}_n)\Bigg)\Bigg]\end{aligned} \tag{4.6}$$

and

$$\begin{aligned}Z_{n,3} := \sqrt{nh_n^{d+2}}\Bigg[\frac{1}{nh_n^{d+2}}\sum_{i=1}^n\Bigg(&K\left(\frac{p-X_i}{h_n}\right)f'''(\xi_i)\\&\times(\mathcal{E}_p^{-1}(X_i),\mathcal{E}_p^{-1}(X_i),\mathcal{E}_p^{-1}(X_i))I(X_i\in\mathcal{B}_n)\\&- EK\left(\frac{p-X}{h_n}\right)f'''(\xi)(\mathcal{E}_p^{-1}(X),\mathcal{E}_p^{-1}(X),\mathcal{E}_p^{-1}(X))I(X\in\mathcal{B}_n)\Bigg)\Bigg]\end{aligned} \tag{4.7}$$

tend to zero in probability: the estimates (3.10) give

$$\begin{aligned}EZ_{n,2}^2 &\le \frac{1}{(4\pi)^d|\mu|h_n^{d+2}}\int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/2h_n^2}(f''(p)(v,v))^2\sqrt{\det(g_{ij})}(v)dv\\&\le \frac{2}{(4\pi)^d|\mu|h_n^{d+2}}\int_{\tilde{\mathcal{B}}_n} e^{-|v|^2/3h_n^2}|f''(p)|^2|v|^4dv\\&\le \frac{2h_n^{d+4}}{(4\pi)^d|\mu|h_n^{d+2}}\int_{\mathbf{R}^d} e^{-|v|^2/3}|f''(p)|^2|v|^4dv = O(h_n^2)\to 0,\end{aligned} \tag{4.8}$$

and, with $c = \sup_{x \in U} |f'''(x)|$,

$$\begin{aligned} E|Z_{n,3}| &\le \frac{2cn}{(4\pi)^{d/2}|\mu|\sqrt{nh_n^{d+2}}} \int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/4h_n^2} |v|^3 \sqrt{\det(g_{ij})(v)}dv \\ &\le \frac{3cn}{(4\pi)^{d/2}|\mu|\sqrt{nh_n^{d+2}}} \int_{\tilde{\mathcal{B}}_n} e^{-|v|^2/5h_n^2}|v|^3 dv \\ &\le \frac{3h_n^{3+d}cn}{(4\pi)^{d/2}|\mu|\sqrt{nh_n^{d+2}}} \int_{R^d} e^{-|v|^2/5}|v|^3 dv \\ &= O\Big(\sqrt{nh_n^{d+4}}\Big) \to 0. \end{aligned} \tag{4.9}$$

Finally, we show that the linear term part,

$$\begin{aligned} Z_{n,1} := \sqrt{nh_n^{d+2}} \Bigg[\frac{1}{nh_n^{d+2}} \sum_{i=1}^n \Bigg(K\Big(\frac{p-X_i}{h_n}\Big) \langle f'(p), \mathcal{E}_p^{-1}(X_i)\rangle I(X_i \in \mathcal{B}_n) \\ - EK\Big(\frac{p-X}{h_n}\Big)\langle f'(p), \mathcal{E}_p^{-1}(X)\rangle I(X \in \mathcal{B}_n)\Bigg)\Bigg] \end{aligned} \tag{4.10}$$

is asymptotically $N(0, s^2)$. Since, by (3.11),

$$\begin{aligned} &\sqrt{\frac{n}{h_n^{d+2}}} EK\Big(\frac{p-X}{h_n}\Big)\langle f'(p), \mathcal{E}_p^{-1}(X)\rangle I(X \in \mathcal{B}_n) \\ &\qquad = \sqrt{\frac{n}{h_n^{d+2}}} O(h_n^{d+3}) = O\Big(\sqrt{nh_n^{4+d}}\Big) \to 0, \end{aligned}$$

and since, by computations similar to the ones leading to (3.11),

$$\begin{aligned} &\frac{1}{h_n^{d+2}}\, E\left[K\Big(\frac{p-X}{h_n}\Big)\langle f'(p), \mathcal{E}_p^{-1}(X)\rangle I(X \in \mathcal{B}_n)\Big)\right]^2 \\ &\quad = \frac{1}{(4\pi)^d|\mu|h_n^{d+2}} \int_{\tilde{\mathcal{B}}_n} e^{-\|p-\mathcal{E}_p(v)\|^2/2h_n^2} \langle f'(p), v\rangle^2 \sqrt{\det(g_{ij})(v)}dv \\ &\quad = \frac{1}{(4\pi)^d|\mu|} \int_{\mathbf{R}^d} e^{-|v|^2/2}\langle f'(p), v\rangle^2 dv \\ &\qquad + \frac{1}{h_n^{d+2}}\bigg(O\Big(h_n^{d+4}\Big) + O\Big(h_n^{2+d+L^2/4}(\log h_n^{-1})^{(d+1)/2}\Big)\bigg), \end{aligned}$$

we have that, taking $L^2/4 > 2 + d$,

$$\lim_{n\to\infty} EZ_{n,1}^2 = s^2. \tag{4.11}$$

Therefore, by Lyapunov's theorem (e.g., [2], p. 44), in order to show that

$$\mathcal{L}(Z_{n,1}) \to N(0, s^2), \tag{4.12}$$

it suffices to prove that

$$\begin{aligned} \frac{n}{\Big(\sqrt{nh_n^{d+2}}\Big)^4} E\Big[K\Big(\frac{p-X}{h_n}\Big)\langle f'(p), \mathcal{E}_p^{-1}(X)\rangle I(X \in \mathcal{B}_n) \\ - EK\Big(\frac{p-X}{h_n}\Big)\langle f'(p), \mathcal{E}_p^{-1}(X)\rangle I(X \in \mathcal{B}_n)\Big]^4 \to 0 \end{aligned} \tag{4.13}$$

By the hypothesis on h_n and (3.11) we can ignore the expected value within the square bracket, and for the rest, proceeding as usual, we have

$$\begin{aligned}
&\frac{1}{nh_n^{2(d+2)}}E\Big[K\Big(\frac{p-X}{h_n}\Big)\langle f'(p), \mathcal{E}_p^{-1}(X)\rangle I(X\in\mathcal{B}_n)\Big]^4\\
&\quad=\frac{1}{(4\pi)^2|\mu|nh_n^{2(d+2)}}\int_{\tilde{\mathcal{B}}_n} e^{-(|v|^2+O(|v|^4)/h_n^2}\langle f'(p), v\rangle^4(1+O(|v|^2))dv\\
&\quad\le\frac{2h_n^{d+4}}{(4\pi)^2|\mu|nh_n^{2(d+2)}}\int_{\mathbf{R}^d} e^{-|v|^2/2}|f'(p)|^4dv = O(1/(nh_n^d))\to 0,
\end{aligned}$$

proving (4.13), and therefore, the limit (4.12). Now the theorem follows from Proposition 2.1, (4.5), (4.8), (4.9) and (4.12). □

This result extends without effort to the CLT uniform in f, which is the main result in this section.

Theorem 4.2. *Let U be a normal neighborhood of p and let $\mathcal{F}$ be a class of functions uniformly bounded up to the third order in U. Assume $nh_n^{d+4}\to 0$, and $nh_n^d\to\infty$. Then, as $n\to\infty$, the processes*

$$\Big\{\sqrt{nh_n^{d+2}}\Big[\Delta_{h_n,n}f(p)-\frac{1}{|\mu|}\Delta_M f(p)\Big]:\ f\in\mathcal{F}\Big\} \tag{4.14}$$

converge in law in $\ell^\infty(\mathcal{F})$ to the Gaussian process

$$\Big\{G(f):=\frac{1}{2^{d/2}(2\pi)^{d/4}|\mu|^{1/2}}\sum_{j=1}^d Z_j\frac{\partial f}{\partial x_j^p}(p):\ f\in\mathcal{F}\Big\}, \tag{4.15}$$

where $Z=(Z_1,\dots,Z_d)$ is the standard normal vector in $T_p(M)$ $(=\mathbf{R}^d)$.

Proof. The proof of Proposition 4.1 applied to $f=\sum_{j=1}^r\alpha_j f_j$, with $f_j\in\mathcal{F}$, shows that the finite dimensional distributions of the processes (4.14) converge to those of the process (4.15) (by the definition (4.3) of $s=s(f)$). Also, by Theorem 3.1, we can center the processes (4.14). Hence, the Theorem will follow if we show that the processes $Z_n=Z_n(f)$ in (4.4) are asymptotically equicontinuous with respect to a totally bounded pseudometric on $\mathcal{F}$ (e.g., [5]).

First, by the computation (3.8) we can restrict the range of X_i to $X_i\in\mathcal{B}_n$ by taking $L^2/4>d+3$, because

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n^{d+2}}}E\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n\Big(K\Big(\frac{p-X_i}{h_n}\Big)(f(X_i)-f(p))I(X_i\in M\setminus\mathcal{B}_n)\\
&\qquad\qquad - EK\Big(\frac{p-X}{h_n}\Big)(f(X_i)-f(p))I(X\in M\setminus\mathcal{B}_n)\Big)\Big|\\
&\quad\le\frac{4cnh_n^{L^2/4}}{(4\pi)^{d/2}\sqrt{nh_n^{d+2}}}\to 0,
\end{aligned}$$

since $nh_n^{d+4}\to 0$.

Now that we can restrict to $X_i \in \mathcal{B}_n$, we only need to consider $Z_{n,i}(f)$, $i = 1, 2, 3$, as defined by equations (4.10), (4.6) and (4.7) respectively. Asymptotic equicontinuity of $Z_{n,1}(f)$ follows because

$$\begin{aligned}
E \sup_{\|f'(p)\| \le \delta} |Z_{n,1}(f)|^2 &\le \delta^2 E \Bigg\| \frac{1}{\sqrt{nh_n^{d+2}}} \sum_{i=1}^n \Bigg(K\left(\frac{p - X_i}{h_n}\right) \mathcal{E}_p^{-1}(X_i) I(X_i \in \mathcal{B}_n) \\
&\qquad - EK\left(\frac{p - X}{h_n}\right) \mathcal{E}_p^{-1}(X) I(X \in \mathcal{B}_n) \Bigg) \Bigg\|^2 \\
&\le \frac{\delta^2}{h_n^{d+2}} E \left\| K\left(\frac{p - X}{h_n}\right) \mathcal{E}_p^{-1}(X) I(X \in \mathcal{B}_n) \right\|^2 \\
&= \frac{\delta^2}{(4\pi)^d |\mu| h_n^{d+2}} \int_{\tilde{\mathcal{B}}_n} e^{-\|p - \mathcal{E}_p(v)\|^2/2h_n^2} |v|^2 \sqrt{\det(g_{ij})}(v) dv \\
&\le \frac{\delta^2}{(4\pi)^d |\mu|} \int_{\mathbf{R}^d} e^{-|v|^2/3} |v|^2 dv
\end{aligned}$$

which tends to zero when we take sup over n and then limit as $\delta \to 0$.

Next, by the computation in (4.9),

$$E \sup_{f \in \mathcal{F}} |Z_{n,3}(f)| = O\left(\sqrt{nh_n^4}\right) \to 0.$$

Finally we consider $Z_{n,2}(f)$. Let ε_i be i.i.d. Rademacher variables independent of $\{X_i\}$. Then, by symmetrization,

$$\begin{aligned}
&E \sup_{f \in \mathcal{F}} Z_{n,2}^2(f) \\
&\quad \le \frac{4}{nh_n^{d+2}} E E_\varepsilon \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n \varepsilon_i K\left(\frac{p - X_i}{h_n}\right) f''(p)(\mathcal{E}_p^{-1}(X_i), \mathcal{E}_p^{-1}(X_i)) I(X_i \in \mathcal{B}_n) \right|^2.
\end{aligned}$$

Next, we recall that, for an operator A in $\mathbf{R}^d$, (or in $T_p(M)$), we have the following identity for its quadratic form

$$A(u, v) := \langle Au, v \rangle = \langle A, u \otimes v \rangle_{HS},$$

where in orthonormal coordinates, $u \otimes v$ is the $d \times d$-matrix with entries $u_i v_j$, and the Hilbert-Schmidt inner product of two matrices is just the inner product in $\mathbf{R}^{2d}$. Also note in particular that $\|u \otimes v\|_{HS} = |u||v|$. Therefore,

$$\begin{aligned}
&E_\varepsilon \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n \varepsilon_i K\left(\frac{p - X_i}{h_n}\right) f''(p)(\mathcal{E}_p^{-1}(X_i), \mathcal{E}_p^{-1}(X_i)) I(X_i \in \mathcal{B}_n) \right|^2 \\
&\quad = E_\varepsilon \sup_{f \in \mathcal{F}} \left\langle f''(p), \sum_{i=1}^n \varepsilon_i K\left(\frac{p - X_i}{h_n}\right) (\mathcal{E}_p^{-1}(X_i) \otimes \mathcal{E}_p^{-1}(X_i)) I(X_i \in \mathcal{B}_n) \right\rangle_{HS}^2 \\
&\quad \le E_\varepsilon \sup_{f \in \mathcal{F}} |f''(p)|_{HS}^2 E_\varepsilon \\
&\qquad \times \left\| \sum_{i=1}^n \varepsilon_i K\left(\frac{p - X_i}{h_n}\right) (\mathcal{E}_p^{-1}(X_i) \otimes \mathcal{E}_p^{-1}(X_i)) I(X_i \in \mathcal{B}_n) \right\|_{HS}^2 \\
&\quad \le b^2 \sum_{i=1}^n K^2\left(\frac{p - X_i}{h_n}\right) \|(\mathcal{E}_p^{-1}(X_i) \otimes \mathcal{E}_p^{-1}(X_i))\|_{HS}^2 I(X_i \in \mathcal{B}_n) =: c^2 \Lambda_n^2.
\end{aligned}$$

Now, by (4.8),

$$E\Lambda_n^2 = nEK^2\left(\frac{p-X}{h_n}\right)\|\mathcal{E}_p^{-1}(X)\|^4 I(X\in\mathcal{B}_n) \le \frac{2nh_n^{d+4}}{(4\pi)^d|\mu|}\int_{\mathbf{R}^d} e^{-|v|^2/3}|v|^4 dv,$$

which gives

$$E\sup_{f\in\mathcal{F}} Z_{n,2}^2(f) = O(h_n^2)\to 0. \qquad \square$$

A simpler proof along similar lines gives the following law of large numbers:

Theorem 4.3. *Let U be a normal neighborhood of p and let $\mathcal{F}$ be uniformly bounded and equicontinuous up to the second order in U. Assume $h_n\to 0$ and $nh_n^{d+2}\to\infty$. Then*

$$\sup_{f\in\mathcal{F}}\left|\Delta_{h_n,n}f(p) - \frac{1}{|\mu|}\Delta_M f(p)\right| \to 0 \quad \text{in pr.} \tag{4.16}$$

A Law of the Iterated Logarithm is also possible, but we refrain from presenting one since in the next section we will give a law of the logarithm for the sup over $f\in\mathcal{F}$ and $p\in M$, and the same methods, with a simpler proof, give the LIL at a single point.

5. Uniform approximation of the Laplacian by graph Laplacians

This section is devoted to results about approximation of the Laplacian by graph Laplacians not only uniformly on the functions f, but also on the points $p\in M$, M a compact submanifold or M as in Remark 3.2. The distributional convergence requires extra work (recall the Bickel-Rosenblatt theorem on the asymptotic distribution of the sup of the difference between a density and its kernel estimator) and will not be considered here.

Although the results in this section are also valid in the situation of Remark 3.2, we will only state them for M a compact submanifold (without boundary). Also, we will identify M with $\phi(M)$, that is, the imbedding ϕ will not be displayed.

Theorem 5.1. *Let M be a compact Riemannian submanifold of dimenison $d<m$ of $\mathbf{R}^m$, let X, X_i be i.i.d. with law $\mu/|\mu|$ and let K be as defined in* (3.1). *Let $\mathcal{F}$ be a class of functions uniformly bounded and equicontinuous up to the second order in M. If $h_n\to 0$ and $nh_n^{d+2}/\log h_n^{-1}\to\infty$, then*

$$\sup_{f\in\mathcal{F}}\sup_{q\in M}\left|\Delta_{h_n,n}f(q) - \frac{1}{|\mu|}\Delta_M f(q)\right| \to 0 \ \text{ a.s.} \tag{5.1}$$

as $n\to\infty$. Moreover, if $\mathcal{F}$ is a class of functions uniformly bounded up to the third order in M, and, in addition to the previous conditions on h_n, $nh_n^{d+4}/\log h_n^{-1}\to 0$, then

$$\sup_{f\in\mathcal{F}}\sup_{q\in M}\left|\Delta_{h_n,n}f(q) - \frac{1}{|\mu|}\Delta_M f(q)\right| = O\left(\sqrt{\frac{\log(1/h_n)}{nh_n^{d+2}}}\right) \ \text{ as } \ n\to\infty \ \text{ a.s.} \tag{5.1$'$}$$

Proof. By Remark 3.1 on Theorem 3.1 (more precisely, by its uniform version), in order to prove (5.1) it suffices to show that

$$\sup_{f\in\mathcal{F}}\sup_{q\in M}\left|\Delta_{h_n,n}f(q) - \Delta_{h_n}f(q)\right| \to 0 \ \textit{ a.s.} \tag{5.2}$$

Let $\mathcal{B}_{n,q} = \{x \in M : \|x - q\| < Lh_n(\log h_n^{-1})^{1/2}\}$, where, we recall, $\|\cdot\|$ is the norm in $\mathbf{R}^m$. Then, as in (3.8) and (4.5), if $L^2/4 > d+2$,

$$
\begin{aligned}
&\sup_f \sup_q \Big| \frac{1}{nh_n^{d+2}} \sum_{i=1}^n \Big(K((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}^c)(f(X_i) - f(q)) \\
&\qquad\qquad - E\Big(K((q-X)/h_n) I(X \in \mathcal{B}_{n,q}^c)(f(X) - f(q))\Big)\Big)\Big| \qquad (5.3)\\
&\leq \frac{4\|f\|_\infty h_n^{L^2/4}}{h_n^{d+2}} \to 0.
\end{aligned}
$$

To establish (5.1), we show that

$$
\begin{aligned}
E_n := \frac{1}{nh_n^{d+2}} E \sup_f \sup_q \Big| \sum_{i=1}^n \Big(& K((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q})(f(X_i) - f(q)) \\
& - EK((q-X)/h_n) I(X \in \mathcal{B}_{n,q})(f(X) - f(q))\Big)\Big| \to 0, \qquad (5.4)
\end{aligned}
$$

and use Talagrand's [14] concentration inequality to transform this into a statement on a.s. convergence.

Each function $f \in \mathcal{F}$ can be extended to a twice continuously differentiable function $\overline{f}$ on a compact domain N of $\mathbf{R}^m$ with M in its interior such that the classes $\{\overline{f} : f \in \mathcal{F}\}$ $\{\overline{f}' : f \in \mathcal{F}\}$ $\{\overline{f}'' : f \in \mathcal{F}\}$ are uniformly bounded and $\{\overline{f}'' : f \in \mathcal{F}\}$ is equicontinuous on N (use a finite partition of unity to patch together convenient extensions of f in each of the sets in a finite cover of M by e.g., geodesic balls: see e.g. Lee [9], pp. 15-16). Then,

$$f(X_i) - f(q) = \overline{f}'(q + \theta(X_i - q))(X_i - q)$$

for some point $0 \leq \theta = \theta_{q,X_i} \leq 1$.

Note that, M being compact, $\mathcal{B}_{n,q}$ is contained in one of a finite number of uniformly normal neighborhoods for all $n \geq N_0$, with $N_0 < \infty$ independent of q, so, we can use q-normal coordinates and notice that for these coordinates, on $\mathcal{B}_{n,q}$, we have the inequalities (3.10) holding uniformly in q (by Proposition 2.2). Since the derivative $\overline{f}'$ is uniformly bounded, for $n \geq N_1$ (independent of q), we have

$$
\begin{aligned}
&EK^2((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q})(f(X_i) - f(q))^2 \\
&\qquad \leq CEK^2((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) \|X_i - q\|^2,
\end{aligned}
$$

which, in view of (3.10), can be further bounded by

$$
\begin{aligned}
&2CEK^2((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) |\mathcal{E}_q^{-1}(X_i)|^2 \\
&\qquad \leq 2C \int_{\tilde{\mathcal{B}}_{n,q}} e^{-(|v|^2 - C|v|^4)/2h_n^2} |v|^2 (1 + C_1|v|^2) dv \\
&\qquad \leq 2Ch_n^{d+2} \int_{\mathbf{R}^d} e^{-|v|^2/3} |v|^2 (1 + C_1|v|^2) dv \leq C_2 h_n^{d+2},
\end{aligned}
$$

so we have

$$EK^2((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q})(f(X_i) - f(q))^2 \leq C_2 h_n^{d+2} \qquad (5.5)$$

with a constant C_2 that does not depend on q.

To prove (5.4), we replace $f(X_i) - f(q)$ by its Taylor expansion of the second order:

$$f(X_i) - f(q) = \overline{f}'(q)(X_i - q) + \frac{1}{2}\overline{f}''(q)(X_i - q, X_i - q) + r_n(f; q; X_i),$$

where

$$\sup_{q \in M} \sup_{f \in \mathcal{F}} r_n(f; q; X) \leq \delta_n \|X - q\|^2$$

with $\delta_n \to 0$ as $n \to \infty$ (because of equicontinuity of $\{\overline{f}'' : f \in \mathcal{F}\}$ and the fact that $\|X - q\| < Lh_n(\log h_n^{-1})^{1/2} \to 0$).

The first order term leads to bounding the expectation

$$\frac{1}{nh_n^{d+2}} E \sup_f \sup_q \Big| \Big\langle \overline{f}'(q), \sum_{i=1}^n \Big(K((q - X_i)/h_n) I(X_i \in \mathcal{B}_{n,q})(X_i - q) \\ - EK((q - X)/h_n) I(X \in \mathcal{B}_{n,q})(X - q) \Big) \Big\rangle \Big|,$$

which is smaller than

$$\frac{b}{nh_n^{d+2}} E \sup_f \sup_q \Big\| \sum_{i=1}^n \Big(K((q - X_i)/h_n) I(X_i \in \mathcal{B}_{n,q})(X_i - q) \\ - EK((q - X)/h_n) I(X \in \mathcal{B}_{n,q})(X - q) \Big) \Big\|,$$

where b is a uniform upper bound on $\overline{f}'$. Denote the coordinates of $x \in \mathbf{R}^m$ in the canonical basis of $\mathbf{R}^m$ by x_α, $\alpha = 1, \ldots, m$ and consider the class of functions $M \mapsto \mathbf{R}$,

$$\mathcal{G} = \{ f_{q,h,\lambda}(x) := e^{-\|q-x\|^2/4h^2} I(\|x - q\| < \lambda)(x_\alpha - q_\alpha) : q \in M, h > 0, \lambda > 0 \}.$$

By arguments of Nolan and Pollard (1987), the class of functions of x, $\{e^{-\|q-x\|^2/4h^2}$: $q \in M, h > 0\}$ is VC subgraph; and it is well known that the open balls in $\mathbf{R}^m$ are VC and that the class of functions $\{x_\alpha - q_\alpha : q \in M\}$ is also VC subgraph (see, e.g., [5]). The three classes are bounded (resp. by 1, 1 and $2 \sup\{\|x\| : x \in M\}$) and therefore, by simple bounds on covering numbers, the product of the three classes is VC-type with respect to the constant envelope $C = 2 + 2 \sup\{\|x\| : x \in M\}$. In particular, if $N(\mathcal{G}, \varepsilon)$ are the covering numbers for $\mathcal{G}$ in L_2 of any probability measure, then

$$N(\mathcal{G}, \varepsilon) \leq \Big(\frac{A}{\varepsilon} \Big)^v$$

for some $A, v < \infty$ and all ε less than or equal to the diameter of $\mathcal{G}$. Hence, by inequality (2.2) in [7], there exists a constant R such that

$$E \sup_q \Big| \sum_{i=1}^n \Big(K((q - X_i)/h_n) I(\|X_i - q\| < Lh_n(\log h_n^{-1})^{1/2})(X_{i,\alpha} - q_\alpha) \\ - E\Big(K((q - X)/h_n) I(\|X - q\| < Lh_n(\log h_n^{-1})^{1/2})(X_\alpha - q_\alpha) \Big| \\ \leq R\Big(\sqrt{n}\sigma \sqrt{\log \frac{A}{\sigma}} \ \vee \ \log \frac{A}{\sigma} \Big), \tag{5.6}$$

where $\sigma^2 \geq \sup_{f\in\mathcal{G}} Ef^2(X)$. Now, to compute σ we use again our observations before the proof of (5.5). For $n \geq N_1$ (independent of q), we have

$$\sup_{f\in\mathcal{G}} Ef^2(X) \leq \int_{\tilde{\mathcal{B}}_{n,q}} e^{-(|v|^2-C|v|^4)/2h_n^2}|v|^2(1+C_1|v|^2)dv$$
$$\leq h_n^{d+2}\int_{\mathbf{R}^d} e^{-|v|^2/3}|v|^2 dv \leq C_2 h_n^{d+2},$$

for some $C_2 < \infty$ independent of q. So, we can take $\sigma^2 = C_2 h_n^{d+2}$. Hence, by the hypothesis on h_n, the right hand side of (5.6) is bounded by

$$R'\Big(nh_n^{d+2}\log\frac{A}{h_n}\Big)^{1/2}$$

for some $R' < \infty$.

To handle the second order term, note that

$$\frac{1}{nh_n^{d+2}}E\sup_f\sup_q\Big|\sum_{i=1}^n\Big(K((q-X_i)/h_n)I(X_i\in\mathcal{B}_{n,q})\overline{f}''(q)(X_i-q,X_i-q)$$
$$-EK((q-X)/h_n)I(X\in\mathcal{B}_{n,q})\overline{f}''(q)(X-q,X-q)\Big)\Big|$$
$$=\frac{1}{nh_n^{d+2}}E\sup_f\sup_q\Big|\Big\langle\overline{f}''(q),$$
$$\sum_{i=1}^n\Big(K((q-X_i)/h_n)I(X_i\in\mathcal{B}_{n,q})(X_i-q)\otimes(X_i-q)$$
$$-EK((q-X)/h_n)I(X\in\mathcal{B}_{n,q})(X-q)\otimes(X-q)\Big)\Big\rangle_{HS}\Big|,$$

which is dominated by

$$\frac{b}{nh_n^{d+2}}E\sup_f\sup_q\Big\|\sum_{i=1}^n\Big(K((q-X_i)/h_n)I(X_i\in\mathcal{B}_{n,q})(X_i-q)\otimes(X_i-q)$$
$$-EK((q-X)/h_n)I(X\in\mathcal{B}_{n,q})(X-q)\otimes(X-q)\Big)\Big\|_{HS}$$

(with b being a uniform upper bound on $\overline{f}''$). Here $\otimes$ denotes the tensor product of vectors of $\mathbf{R}^m$ and $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm for linear transformations of $\mathbf{R}^m$. This leads to bounding the expectation

$$E\sup_q\Big|\sum_{i=1}^n\Big(K((q-X_i)/h_n)I(\|X_i-q\|<Lh_n(\log h_n^{-1})^{1/2})(X_{i,\alpha}-q_\alpha)(X_{i,\beta}-q_\beta)$$
$$-E\Big(K((q-X)/h_n)I(\|X-q\|<Lh_n(\log h_n^{-1})^{1/2})(X_\alpha-q_\alpha)(X_\beta-q_\beta)\Big| \tag{5.6'}$$

for all $1 \leq \alpha,\beta \leq m$, which is done using the inequality for empirical processes on VC-subgraph classes exactly the same way as in the case (5.6). This time the bound becomes

$$R'\Big(nh_n^{d+4}\log\frac{A}{h_n}\Big)^{1/2}.$$

For the remainder, we have the bound

$$
\begin{aligned}
&\frac{1}{nh_n^{d+2}} E \sup_f \sup_q \Big| \sum_{i=1}^n K((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) r_n(f,q,X_i) \\
&\qquad\qquad - EK((q-X)/h_n) I(X \in \mathcal{B}_{n,q}) r_n(f,q,X)\Big)\Big| \\
&\quad \le \frac{\delta_n}{nh_n^{d+2}} E \sup_q \sum_{i=1}^n \Big(K((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) \|X_i - q\|^2 \\
&\qquad + n \frac{\delta_n}{nh_n^{d+2}} \sup_q E\Big(K((q-X)/h_n) I(X \in \mathcal{B}_{n,q}) \|X - q\|^2 \Big) \\
&\quad \le \frac{\delta_n}{nh_n^{d+2}} E \sup_q \sum_{i=1}^n \Big(K((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) \|X_i - q\|^2 \\
&\qquad\qquad - EK((q-X)/h_n) I(X \in \mathcal{B}_{n,q}) \|X - q\|^2 \Big) \\
&\qquad + \frac{2\delta_n}{h_n^{d+2}} \sup_q EK((q-X)/h_n) I(X \in \mathcal{B}_{n,q}) \|X - q\|^2.
\end{aligned}
$$

The first expectation is bounded again by using the inequality for VC-subgraph classes and the bound in this case is

$$
\frac{\delta_n}{nh_n^{d+2}} \Big(nh_n^{d+4} \log \frac{A}{h_n} \Big)^{1/2} = \delta_n \sqrt{\frac{\log(A/h_n)}{nh_n^d}} \to 0.
$$

The second expectation is bounded by replacing $\|X-q\|^2$ by $|\mathcal{E}_q^{-1}(X)|^2$ and changing variables in the integral (as it has been done before several times). This yields a bound of the order $C\delta_n$, which also tends to 0.

Combining the above bounds establishes (5.4). One of the versions of Talagrand's inequality (e.g., [10]) together with (5.5) gives with some constant $K > 0$ and with probability at least $1 - e^{-t}$

$$
\begin{aligned}
&\frac{1}{n} \sup_f \sup_q \Big| \sum_{i=1}^n \Big(K((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) (f(X_i) - f(q)) \\
&\qquad\qquad - EK((q-X)/h_n) I(X \in \mathcal{B}_{n,q}) (f(X) - f(q)) \Big) \Big| \\
&\quad \le K \Big(h_n^{d+2} E_n + \sqrt{h_n^{d+2} \frac{t}{n}} + \frac{t}{n} \Big).
\end{aligned}
$$

Taking $t := t_n := A \log n$ with large enough A, so that $\sum_n e^{-t_n} < \infty$ and using Borel-Cantelli Lemma shows that a.s. for large enough n

$$
\begin{aligned}
&\frac{1}{nh_n^{d+2}} \sup_f \sup_q \Big| \sum_{i=1}^n \Big(K((q-X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) (f(X_i) - f(q)) \\
&\qquad\qquad - EK((q-X)/h_n) I(X \in \mathcal{B}_{n,q}) (f(X) - f(q)) \Big) \Big| \\
&\quad \le K \Big(E_n + \sqrt{\frac{A \log n}{nh_n^{d+2}}} + \frac{A \log n}{nh_n^{d+2}} \Big)
\end{aligned}
$$

and since, in view of (5.4) and under the condition $nh_n^{d+2}/\log h_n^{-1} \to \infty$, the right hand side tends to 0. This and (5.3) yield (5.1). (Note that $nh_n^{d+2}/\log h_n^{-1} \to \infty$ implies $nh_n^{d+2}/\log n \to \infty$.)

The proof of (5.1') requires the following version of Taylor's expansion of $\overline{f}$:

$$
\begin{aligned}
f(X_i) - f(q) = \overline{f}'(q)(X_i - q) + \frac{1}{2}\overline{f}''(q)(X_i - q, X_i - q) \\
+ \frac{1}{6}\overline{f}'''(q + \theta_i(X_i - q))(X_i - q, X_i - q, X_i - q).
\end{aligned}
$$

The first two terms have been handled before, and the expectations of the sup-norms of the corresponding empirical processes were shown to be $O(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}})$. The third term leads to bounding

$$
\begin{aligned}
\frac{1}{nh_n^{d+2}} E \sup_f \sup_q \Big| \sum_{i=1}^n \Big(& K((q - X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) \\
& - EK((q - X)/h_n) I(X \in \mathcal{B}_{n,q}) \Big) \\
& \times \frac{1}{6}\overline{f}'''(q + \theta_i(X_i - q))(X_i - q, X_i - q, X_i - q) \Big|,
\end{aligned}
$$

which, for f''' uniformly bounded by b, is smaller than

$$
\begin{aligned}
& \frac{b}{6nh_n^{d+2}} E \sup_q \sum_{i=1}^n \Big(K((q - X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) \|X_i - q\|^3 \\
& \qquad + \frac{bn}{6nh_n^{d+2}} \sup_q EK((q - X)/h_n) I(X \in \mathcal{B}_{n,q}) \|X - q\|^3 \Big) \\
& \leq \frac{b}{6nh_n^{d+2}} E \sup_q \Big| \sum_{i=1}^n \Big(K((q - X_i)/h_n) I(X_i \in \mathcal{B}_{n,q}) \|X_i - q\|^3 \\
& \qquad - EK((q - X)/h_n) I(X \in \mathcal{B}_{n,q}) \|X - q\|^3 \Big) \Big| \\
& \quad + \frac{b}{3h_n^{d+2}} \sup_q EK((q - X)/h_n) I(X \in \mathcal{B}_{n,q}) \|X - q\|^3,
\end{aligned}
$$

which can be handled exactly as before and shown to be of the order

$$
h_n \sqrt{\frac{\log h_n^{-1}}{nh_n^2}} + h_n = o\Big(\sqrt{\frac{\log h_n^{-1}}{nh_n^{d+2}}} \Big),
$$

by the conditions on h_n. Using Talagrand's inequality the same way as before, completes the proof of $(5.1')$. □

We conclude with the following theorem, whose proof is a little longer and more involved, but it is based on a methodology that is well known and well described in the literature (see [6] and [7]). Its extension to the case of manifolds requires some work, but is rather straightforward.

Theorem 5.2. *Let M be a compact Riemannian submanifold of dimenison $d < m$ of $\mathbf{R}^m$, let X, X_i be i.i.d. with law $\mu/|\mu|$ and let K be as defined in* (3.1). *Assume that $h_n \to 0$, $nh_n^{d+2}/\log h_n^{-1} \to \infty$, and $nh_n^{d+4}/\log h_n^{-1} \to 0$. Let $\mathcal{F}$ be a class of* *functions uniformly bounded up to the third order in M. Then,*

$$\lim_{n\to\infty}\sqrt{\frac{nh_n^{d+2}}{2\log h_n^{-d}}}\sup_{f\in\mathcal{F}}\sup_{q\in M}\left|\Delta_{h_n,n}f(q)-\frac{1}{|\mu|}\Delta_M f(q)\right|$$
$$=\frac{\sup_{f\in\mathcal{F},q\in M}\left(\sum_{j=1}^{d}\left(\frac{\partial f}{\partial x_j^q}(q)\right)^2\right)^{1/2}}{2^{d/2}(2\pi)^{d/4}|\mu|^{1/2}}\quad \text{a.s.},$$

where x_j^q denote normal coordinates centered at q.

References

[1] Belkin, M. and Niyogi, P. (2005). Towards a theoretical foundation for Laplacian-based manifold methods. In: *Learning Theory: Proceedings of 18th Annual Conference on Learning Theory, COLT 2005*, P. Auer and R. Meir (Eds), *Lecture Notes in Computer Science*, vol. 3559. Springer-Verlag, pp. 486–500.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Bourbaki, N. (1940). *Topologie Générale.* Hermann, Paris.

[4] do Carmo, M. (1992). *Riemannian Geometry.* Birkhäuser, Boston.

[5] Dudley, R. M. (2000). *Uniform Central Limit Theorems.* Cambridge Press, Cambridge.

[6] Einmahl, U. and Mason, D. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** 1–37.

[7] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré* **38** 907–921.

[8] Hein, M., Audibert, J. Y. and Luxburg, U. V. (2005). From graphs to manifolds -weak and strong pointwise consistency of graph Laplacians. In: *Learning Theory: Proceedings of 18th Annual Conference on Learning Theory, COLT 2005*, P. Auer and R. Meir (Eds), *Lecture Notes in Computer Science*, vol. 3559. Springer-Verlag, pp. 470–485.

[9] Lee, J. M. (1997). *Riemannian Manifolds.* Springer, New York.

[10] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863–884.

[11] Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Statistics and Probability Letters* **73** 297–304.

[12] Sakai, T. (1996). *Riemannian Geometry.* AMS, Providence.

[13] Smolyanov, O. G., Weizsäcker, H.V. and Wittich, O. (2004). Chernoff's theorem and discrete time approximations of Brownian Motion on manifolds. Preprint, available at: http://lanl.arxiv.org/abs/math.PR/0409155.

[14] Talagrand, M. (1996). New concentration inequalities in product spaces. *Inventiones Math.* **126** 505–563.

IMS Lecture Notes–Monograph Series
High Dimensional Probability
Vol. 51 (2006) 260–275
In the Public Domain
DOI: 10.1214/074921706000000897

A new concentration result for regularized risk minimizers

Ingo Steinwart,[1] Don Hush[1] and Clint Scovel[1]

Los Alamos National Laboratory

Abstract: We establish a new concentration result for regularized risk minimizers which is similar to an oracle inequality. Applying this inequality to regularized least squares minimizers like least squares support vector machines, we show that these algorithms learn with (almost) the optimal rate in some specific situations. In addition, for regression our results suggest that using the loss function $L_\alpha(y,t) = |y - t^{\alpha}$ with α near 1 may often be preferable to the usual choice of $\alpha = 2$.

1. Introduction

The theoretical understanding of support vector machines (SVMs) and related kernel-based methods has been substantially improved in recent years. Based on Talagrand's concentration inequality and local Rademacher averages it has recently been shown that SVMs for classification can learn with rates up to $\frac{1}{n}$ under somewhat realistic assumptions on the data-generating distribution (see [12] and the related work [3]). However, the currently available technique, namely the so-called "shrinking technique" in [12], for establishing such rates requires choosing the *entire* regularization sequence *a-priori*. Unfortunately, the optimal regularization sequences usually depend on some features of the data-generating distribution typically unknown in practice, and consequently the results derived by the shrinking technique have some serious drawbacks.

In this work we replace the shrinking technique by a localization argument similar to the localization argument used in conjunction with local Rademacher averages. The key observation for this new localization argument is that regularized risk minimizers control the size of the norm in the regularization term by their (excess) risk in a non-trivial manner (see Lemma 4.1 for details). As a consequence of this observation, we can not only localize with respect to small variances but also with respect to small maximum norms.

Using the above (double) localization we obtain oracle-type inequalities for a large class of regularized risk minimizers including support vector machines, and regularization networks. For the former we can easily reproduce rates established in [12, 13], while for the latter we show some minmax rates in specific situations and provide results indicating that using the loss function $L_\alpha(y,t) = |y-t|^\alpha$ with α near 1 to estimate the regression function may be more robust to both outliers and the choice of regularization parameter than the usual choice $\alpha = 2$.

2. An oracle inequality for regularized risk minimizers

Throughout this work we assume that X is compact metric space, $Y \subset [-1,1]$ is compact, P is a Borel probability measure on $X \times Y$, and H is a RKHS of continuous

[1]Modeling, Algorithms and Informatics Group, CCS-3, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA, e-mail: ingo@lanl.gov; dhush@lanl.gov; jcs@lanl.gov

functions over X with closed unit ball B_H. It is well-known that H can then be continuously embedded into the space of continuous functions $C(X)$ equipped with the usual maximum-norm $\|.\|_\infty$. In order to avoid constants we always assume that this embedding has norm 1, i.e. $\|.\|_\infty \leq \|.\|_H$.

Furthermore, $L : Y \times \mathbb{R} \to [0, \infty)$ always denotes a continuous function which is convex in the second variable. In the following we are particularly interested in functions L that satisfy the growth assumptions introduced in [6]:

$$\text{(1)}\quad \sup_{y\in Y} L(y,t) \leq 1 + |t|^\alpha \qquad \text{and} \qquad \sup_{y\in Y} \big| L_{|Y\times[-t,t]}(y,.)\big|_1 \leq c_L\, t^{\alpha-1}$$

for some constants $\alpha \in [1,2]$, $c_L > 0$, and all $t \in \mathbb{R}$, where $|h|_1$ denotes the Lipschitz constant of a function h. The functions L will serve as loss functions and consequently let us recall the associated L-risk

$$\mathcal{R}_{L,P}(f) = \mathbb{E}_{(x,y)\sim P} L(y, f(x)),$$

where $f : X \to \mathbb{R}$ is a measurable function. Note that (1) immediately gives $\mathcal{R}_{L,P}(0) \leq 1$. Furthermore, the minimal L-risk is denoted by $\mathcal{R}^*_{L,P}$, i.e.

$$\mathcal{R}^*_{L,P} = \inf\{\mathcal{R}_{L,P}(f) \,|\, f : X \to \mathbb{R} \text{ measurable}\},$$

and a function attaining this infimum is denoted by $f^*_{L,P}$.

The learning schemes we are interested in are based on an optimization problem of the form

$$f_{P,\lambda} := \arg\min_{f\in H} \Big(\lambda\|f\|^2_H + \mathcal{R}_{L,P}(f)\Big)\,,$$

where $\lambda > 0$. Note that if we identify a training set $T = ((x_1, y_1), \ldots, (x_n, y_n)) \in (X \times Y)^n$ with its empirical measure, then $f_{T,\lambda}$ denotes the empirical estimators of the above learning scheme. Obviously, support vector machines (see e.g. [5]) and regularization networks (see e.g. [8]) are both learning algorithms which fall into the above category.

One way to describe the approximation error of these learning schemes is the *approximation error function*

$$a(\lambda) := \lambda\|f_{P,\lambda}\|^2 + \mathcal{R}_{L,P}(f_{P,\lambda}) - \mathcal{R}^*_{L,P}\,, \qquad \lambda > 0,$$

which we discussed in some detail in [13]. Furthermore in order to deal with the complexity of the used RKHSs let us recall that for a subset $A \subset E$ of a Banach space E the *covering numbers* are defined by

$$\mathcal{N}(A,\varepsilon,E) := \min\Big\{n \geq 1 : \exists x_1, \ldots, x_n \in E \text{ with } A \subset \bigcup_{i=1}^n (x_i + \varepsilon B_E)\Big\}, \qquad \varepsilon > 0,$$

where B_E denotes the closed unit ball of E. Given a finite sequence $T = (z_1, \ldots, z_n) \in Z^n$ we are particularly interested in the Banach space $L_2(T)$ which consists of all equivalence classes of functions $f : Z \to \mathbb{R}$ and which is equipped with the norm

$$\text{(2)}\qquad \|f\|_{L_2(T)} := \Big(\frac{1}{n}\sum_{i=1}^n \big|f(z_i)\big|^2\Big)^{\frac{1}{2}}.$$

In other words, $L_2(T)$ is a L_2-space with respect to the empirical measure of $(z_1, \ldots, z_n)$. Furthermore, if T is of the form $T = ((x_1, y_1), \ldots, (x_n, y_n))$, and

$T_X := (x_1, \dots, x_n)$, then the space $L_2(T_X)$ has the obvious meaning. In addition to the convention $0^0 := 1$ we utilize the following

$$a^\infty := \begin{cases} 0 & \text{if } 0 \le a < 1, \\ 1 & \text{if } a = 1, \\ \infty & \text{if } a > 1\,. \end{cases} \tag{3}$$

Now we can state the main result of this paper:

Theorem 2.1. *Let H be a RKHS of a continuous kernel over X with $\|.\|_\infty \le \|.\|_H$. Assume that there are constants $a \ge 1$ and $0 < p < 2$ such that for all $\delta > 0$ we have*

$$\sup_{T \in Z^n} \log \mathcal{N}\big(B_H, \delta, L_2(T_X)\big) \;\le\; a\delta^{-p}\,. \tag{4}$$

Let $L : Y \times \mathbb{R} \to [0,\infty)$ be a continuous function which is convex in its second variable and satisfies (1). *Furthermore, let P be a distribution on $X \times Y$ such that $f^*_{L,P}$ exists. Moreover, suppose that for all $0 < \lambda \le 1$ and all $f \in \lambda^{-\frac{1}{2}} B_H$ we have*

$$\mathbb{E}_P\big(L \circ f - L \circ f^*_{L,P}\big)^2 \le c\left(\|f\|_\infty + 1\right)^v \left(\mathbb{E}_P L \circ f - L \circ f^*_{L,P}\right)^\vartheta \tag{5}$$

for some constants $c \ge 1$, $\vartheta \in (0,1]$, and $v \in [0,2]$. Then there exists a constant $K \ge 1$ such that for all $0 < \lambda \le 1$, $\varepsilon > 0$, $x \ge 1$ satisfying

$$\varepsilon \ge \max\Bigg\{ a(\lambda) + \lambda, \left(\frac{Ka}{\lambda^{\frac{2\alpha p + v(2-p)}{4}} n}\right)^{\frac{4}{8 - 2\alpha p - (v+2\vartheta)(2-p)}}, \left(\frac{Ka}{\lambda^{\frac{\alpha(2+p)}{4}} n}\right)^{\frac{4}{(2+p)(2-\alpha)}},$$
$$\left(\frac{Kx}{\lambda^{\frac{v}{2}} n}\right)^{\frac{2}{4-v-2\vartheta}}, \left(\frac{Kx}{\lambda^{\frac{\alpha}{2}} n}\right)^{\frac{2}{2-\alpha}} \Bigg\},$$

we have

$$\Pr{}^*\Big(T \in Z^n : \mathcal{R}_{L,P}(f_{T,\lambda}) - \mathcal{R}^*_{L,P} < a(\lambda) + \varepsilon\Big) \;\ge\; 1 - e^{-x}$$

where $\Pr^*$ denotes the outer probability.

Theorem 2.1 is proved in Section 4. Now we proceed to illustrate its utility with some applications.

Example 2.2 (Least square regression with Sobolev spaces). Let us consider the least squares loss function which is defined by $L(y,t) = (y-t)^2$. Furthermore, let us assume that H contains the regression function $x \mapsto \mathbb{E}(y|x)$ and satisfies the complexity exponent condition (4). In addition let (λ_n) be a strictly positive null-sequence with $\lambda_n^{1+p/2} n \to \infty$. Then in Section 5 we show that our learning rate is of the form λ_n. In particular, if H is a Sobolev space of order m on some suitable $X \subset \mathbb{R}^d$, $m > d/2$, then we have $p = d/m$, and consequently, for $\lambda_n := n^{-\frac{2m}{2m+d}} \log n$ our rate becomes $n^{-\frac{2m}{2m+d}} \log n$. This equals the optimal rate $n^{-\frac{2m}{2m+d}}$ up to a logarithmic factor (see e.g. [7] and the references therein).

Example 2.3 (Comparison of different loss functions used for regression). Consider again regression with the squared loss function $L_2(y,t) = (y-t)^2$ defining performance but use the loss function $L_\alpha(y,t) = |y-t|^\alpha$ with $1 \le \alpha \le 2$ to determine the estimate $f_{T,\lambda}$. Suppose that H contains the regression function

$x \mapsto \mathbb{E}(y|x)$, and satisfies the complexity exponent condition (4). In Section 5 we begin by using the oracle inequality of Theorem 2.1 to bound the excess L_α-risk $\mathcal{R}_{L_\alpha,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_\alpha,P}$. When $\alpha = 2$ we produce the results of Example 2.2. When $1 < \alpha < 2$ we set $\lambda = n^{-\kappa}$ with $\kappa > 0$ and observe that when $\kappa < \frac{2}{2+p}$ we obtain the rate $n^{-\kappa}$ independently of the value of α and when $\kappa > \frac{2}{2+p}$ we obtain the rate $n^{-\frac{2}{2+p}+(\kappa-\frac{2}{2+p})\frac{\alpha}{2-\alpha}}$. We conclude that the κ-optimal learning rate for the L_α risk is $n^{-\frac{2}{2+p}}$ and is achieved when $\kappa = \frac{2}{2+p}$. Now suppose that the conditional distributions $P(y|x)$ are symmetric. These results are then combined with a calibration inequality

$$\mathcal{R}_{L_2,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_2,P} \le \Psi(\mathcal{R}_{L_\alpha,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_\alpha,P})$$

derived from [11] to obtain bounds on $\mathcal{R}_{L_2,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_2,P}$ in terms of $1 < \alpha < 2$. We observe that when $\kappa \le \frac{2}{2+p}$ we obtain the rate $n^{-\kappa}$ independently of the value of α and when $\kappa > \frac{2}{2+p}$ we obtain the rate $n^{-\frac{2}{2+p}+(\kappa-\frac{2}{2+p})\frac{2}{2-\alpha}}$. We conclude that the κ-optimal learning rate for the L_2 risk also is $n^{-\frac{2}{2+p}}$ and is achieved when $\kappa = \frac{2}{2+p}$. It is important to observe that the rate for fixed κ gets worse as α increases towards 2 and in particular that we have no rates when $2-(\kappa-\frac{2}{2+p})(2+p) \le \alpha \le 2$. When $\alpha = 1$ [11, Example 3.25] shows how, even though the loss function is not strictly convex, we can obtain a calibration inequality in terms of assumptions concerning the concentration about the mean. Consequently with extra assumptions regarding concentration about the mean we can apply these methods, but do not carry out such calculations here since they they are out of the scope of this paper. Moreover, since $\alpha = 1$ is considered more robust to outliers than $\alpha = 2$, these results suggest that setting α near 1 has some substantial advantages to the usual choice $\alpha = 2$. However, to make such a claim more precise will require considering whether and in which sense the assumptions of symmetry and boundedness have been violated. Finally, let us now consider when H is a Sobolev space as in Example 2.2. Then it is clear that we obtain the same optimal rates for all values of $1 < \alpha \le 2$, although for α near 1 we should concern ourselves with the arising constants.

Example 2.4 (Hinge loss classification). Let $Y := \{-1, 1\}$, L be defined by $L(y,t) := \max\{0, 1-yt\}$, $y \in Y$, $t \in \mathbb{R}$, and P be a distribution with Tsybakov noise exponent $q \in [0,\infty]$ in the sense of [12, 13] (see also [2]). When $q > 0$, it follows from [12, Lemma 6.6] that the assumption (5) is satisfied with $\alpha = 1$, $v = \frac{q+2}{q+1}$, $\vartheta = \frac{q}{q+1}$ and $c = \|(2\eta-1)^{-1}\|_{q,\infty} + 2$. Moreover it is simple to show the same is true when $q = 0$ but with $c = 5$. Hence the condition on ε becomes

$$\varepsilon \ge \max\Big\{a(\lambda)+\lambda, \frac{K}{\lambda}\Big(\frac{a}{n}\Big)^{\frac{4(q+1)}{2q+pq+4}}, \frac{K}{\lambda}\Big(\frac{a}{n}\Big)^{\frac{4}{2+p}}, \frac{K}{\lambda}\Big(\frac{x}{n}\Big)^{\frac{2(q+1)}{q+2}}, \frac{K}{\lambda}\Big(\frac{x}{n}\Big)^{2}\Big\}.$$

Some easy estimates then show that this reduces to

$$\varepsilon \ge a(\lambda) + \lambda + Kx^2\lambda^{-1}\big(\frac{a}{n}\big)^{\frac{4(q+1)}{2q+pq+4}},$$

where $K \ge 1$ is a suitable constant and a and n are assumed to satisfy $n \ge a \ge 1$. From this we immediately obtain the rates established in [12, Thm. 2.8] and [13, Thm. 1].

3. A concentration result for ERM schemes

The proof of our main result Theorem 2.1 is based on a refinement of standard local Rademacher average techniques. Since this refinement may be of its own interest we separate its presentation from the proof of 2.1.

Let us begin by introducing some notations. To this end let $\mathcal{F}$ be a class of bounded measurable functions from Z to $\mathbb{R}$. In order to avoid measurability considerations we always assume that $\mathcal{F}$ *is separable with respect to* $\|.\|_\infty$. Given a probability measure P on Z we define the modulus of continuity of $\mathcal{F}$ by

$$\omega_{P,n}(\mathcal{F},\varepsilon) \ := \ \mathbb{E}_{T\sim P^n}\Big(\sup_{\substack{f\in\mathcal{F},\\ \mathbb{E}_P f\le\varepsilon}} |\mathbb{E}_P f - \mathbb{E}_T f|\Big),$$

where we emphasize that the supremum is, as a function from Z^n to $\mathbb{R}$, measurable by the separability assumption on $\mathcal{F}$. In addition note that the supremum is taken over all $f\in\mathcal{F}$ with $\mathbb{E}_P f\le\varepsilon$, whereas usually the supremum is taken over all $f\in\mathcal{F}$ with $\mathbb{E}_P f^2\le\varepsilon$.

We also need some notations related to ERM-type algorithms: we call $C:\mathcal{F}\times Z\to[0,\infty)$ a *cost function* if $C\circ f:=C(f,.)$ is measurable for all $f\in\mathcal{F}$. Given a probability measure P on Z we denote by $f_{P,\mathcal{F}}\in\mathcal{F}$ a minimizer of

$$f\mapsto \mathcal{R}_{C,P}(f):=\mathbb{E}_{z\sim P}C(f,z).$$

Moreover, if P is an empirical measure with respect to $T\in Z^n$ we write $f_{T,\mathcal{F}}$ and $\mathcal{R}_{C,T}(.)$ as usual. For simplicity, we assume throughout this section that $f_{P,\mathcal{F}}$ and $f_{T,\mathcal{F}}$ do exist. Furthermore, although there may be multiple solutions we use a single symbol for them whenever no confusion regarding the non-uniqueness of this symbol can be expected. An algorithm that produces solutions $f_{T,\mathcal{F}}$ is called an *empirical C-risk minimizer*. Moreover, if $\mathcal{F}$ is convex, we say that C is convex if $C(.,z)$ is convex for all $z\in Z$. Finally, C is called *line-continuous* if for all $z\in Z$ and all $f,\hat f\in\mathcal{F}$ the function $t\mapsto C(tf+(1-t)\hat f,z)$ is continuous on $[0,1]$. If $\mathcal{F}$ is a vector space then every convex C is line-continuous. Now we can formulate the main result of this section:

Theorem 3.1. *Let $\mathcal{F}$ be a convex set of bounded measurable functions from Z to $\mathbb{R}$, $C:\mathcal{F}\times Z\to[0,\infty)$ be a convex, line-continuous cost function, and P be a probability measure on Z. Assume that*

$$\mathcal{G} \ := \ \big\{C\circ f - C\circ f_{P,\mathcal{F}} \ : \ f\in\mathcal{F}\big\}$$

is separable with respect to $\|.\|_\infty$. Furthermore assume that there exist constants $b,B\ge 0$, $\beta\in[0,1]$, and $w,W\ge 0$, $\nu\in[0,2]$, $\vartheta\in[0,2)$, such that

$$\|g\|_\infty \ \le \ b\left(\mathbb{E}_P g\right)^\beta + B \tag{6}$$

and

$$\mathbb{E}_P g^2 \ \le \ \Big(b\big(\mathbb{E}_P g\big)^\beta + B\Big)^\nu\Big(w\big(\mathbb{E}_P g\big)^\vartheta + W\Big) \tag{7}$$

for all $g\in\mathcal{G}$. Then for $n\ge 1$, $x\ge 1$ and $\varepsilon>0$ satisfying

$$\varepsilon\ge 3\omega_{P,n}(\mathcal{G},\varepsilon)+\sqrt{\frac{2x(b\varepsilon^\beta+B)^\nu(w\varepsilon^\vartheta+W)}{n}}+\frac{2x\big(b\varepsilon^\beta+B\big)}{n}$$

we have

$$\Pr{}^*\Big(T\in Z^n:\mathcal{R}_{C,P}(f_{T,\mathcal{F}})<\mathcal{R}_{C,P}(f_{P,\mathcal{F}})+\varepsilon\Big) \ \ge \ 1-e^{-x}\,.$$

In order to prove Theorem 3.1 let us first recall Talagrand's concentration inequality (see [14]). The following version of this inequality is derived from Bousquet's result in [4] using a little trick presented in [1, Lem. 2.5]:

Theorem 3.2. *Let P be a probability measure on Z and $\mathcal{H}$ be a set of bounded measurable functions from Z to $\mathbb{R}$ which is separable with respect to $\|.\|_\infty$ and satisfies $\mathbb{E}_P h = 0$ for all $h \in \mathcal{H}$. Furthermore, let $M > 0$ and $\tau \geq 0$ be constants with $\|h\|_\infty \leq M$ and $\mathbb{E}_P h^2 \leq \tau$ for all $h \in \mathcal{H}$. Then for all $x \geq 1$ and all $n \geq 1$ we have*

$$P^n\Big(T \in Z^n : \sup_{h\in\mathcal{H}} \mathbb{E}_T h > 3\mathbb{E}_{T'\sim P^n} \sup_{h\in\mathcal{H}} \mathbb{E}_{T'} h + \sqrt{\frac{2x\tau}{n}} + \frac{Mx}{n}\Big) \;\leq\; e^{-x}\,.$$

This concentration inequality is used to prove the following lemma which is a generalized version of Lemma 13 in [2] and Lemma 5.4 in [12]:

Lemma 3.3. *Let P be a probability measure on Z and $\mathcal{G}$ be a set of bounded measurable functions from Z to $\mathbb{R}$ which is separable with respect to $\|.\|_\infty$. Let us assume that $\mathcal{G}$ satisfies (6) and (7), and that there is a constant $a \in [0,1)$ such that for all $T \in Z^n$, $\varepsilon > 0$ for which there is a $g \in \mathcal{G}$ with*

$$\mathbb{E}_T g \leq a\varepsilon \qquad \textit{and} \qquad \mathbb{E}_P g \geq \varepsilon$$

there is also an element $g^* \in \mathcal{G}$ with

$$\mathbb{E}_T g^* \leq a\varepsilon \qquad \textit{and} \qquad \mathbb{E}_P g^* = \varepsilon\,.$$

Then for all $n \geq 1$, $x \geq 1$, and all $\varepsilon > 0$ satisfying

$$(1-a)\varepsilon \geq 3\omega_{P,n}(\mathcal{G},\varepsilon) + \sqrt{\frac{2x(b\varepsilon^\beta + B)^\nu(w\varepsilon^\vartheta + W)}{n}} + \frac{2x\big(b\varepsilon^\beta + B\big)}{n}$$

we have

$$\Pr{}^*\Big(T \in Z^n : \textit{ for all } g \in \mathcal{G} \textit{ with } \mathbb{E}_T g \leq a\varepsilon \textit{ we have } \mathbb{E}_P g < \varepsilon\Big) \;\geq\; 1 - e^{-x}\,.$$

Proof. We define $\mathcal{H} := \{\mathbb{E}_P g - g : g \in \mathcal{G}, \mathbb{E}_P g = \varepsilon\}$. Obviously, for all $h \in \mathcal{H}$ we have $\mathbb{E}_P h = 0$ and

$$\begin{aligned}
&\|h\|_\infty \leq 2b\varepsilon^\beta + 2B =: M\,,\\
&\mathbb{E}_P h^2 \leq \mathbb{E}_P g^2 \leq (b\varepsilon^\beta + B)^\nu(w\varepsilon^\vartheta + W) =: \tau\,.
\end{aligned}$$

Moreover, it is also easy to verify that $\mathcal{H}$ is separable with respect to $\|.\|_\infty$. As in the proof of Lemma 5.4 in [12] our assumption on $\mathcal{G}$ now yields

$$\begin{aligned}
&\Pr{}^*\big(T \in Z^n : \exists g \in \mathcal{G} \text{ with } \mathbb{E}_T g \leq a\varepsilon \text{ and } \mathbb{E}_P g \geq \varepsilon\big)\\
&\quad \leq \Pr{}^*\big(T \in Z^n : \exists g \in \mathcal{G} \text{ with } \mathbb{E}_T g \leq a\varepsilon \text{ and } \mathbb{E}_P g = \varepsilon\big)\\
&\quad = \Pr{}^*\big(T \in Z^n : \exists g \in \mathcal{G} \text{ with } \mathbb{E}_P g - \mathbb{E}_T g \geq (1-a)\varepsilon \text{ and } \mathbb{E}_P g = \varepsilon\big)\\
&\quad \leq P^n\Big(T \in Z^n : \sup_{\substack{g\in\mathcal{G}\\ \mathbb{E}_P g=\varepsilon}} (\mathbb{E}_P g - \mathbb{E}_T g) \geq (1-a)\varepsilon\Big)\\
&\quad = P^n\Big(T \in Z^n : \sup_{h\in\mathcal{H}} \mathbb{E}_T h \geq (1-a)\varepsilon\Big)\,.
\end{aligned}$$

In order to bound the last probability we will apply Theorem 3.2. To this end observe

$$3\mathbb{E}_{T'\sim P^n} \sup_{h\in\mathcal{H}} \mathbb{E}_{T'}h + \sqrt{\frac{2x\tau}{n}} + \frac{Mx}{n} \;\leq\; (1-a)\varepsilon\,,$$

and consequently applying Theorem 3.2 yields

$$\Pr{}^*\big(T \in Z^n : \exists g \in \mathcal{G} \text{ with } \mathbb{E}_T g \leq a\varepsilon \text{ and } \mathbb{E}_P g \geq \varepsilon\big) \;\leq\; e^{-x}. \qquad \square$$

With the help of the above lemma we can now prove Theorem 3.1:

Proof of Theorem 3.1. For $a := 0$ we will apply Lemma 3.3 to the class $\mathcal{G}$. To this end it obviously suffices to show the richness condition on $\mathcal{G}$ of Lemma 3.3: let $f \in \mathcal{F}$ satisfy

$$\mathbb{E}_T(C\circ f - C\circ f_{P,\mathcal{F}}) \leq 0 \qquad \text{and} \qquad \mathbb{E}_P(C\circ f - C\circ f_{P,\mathcal{F}}) \geq \varepsilon\,.$$

For $t \in [0,1]$ we define $f_t := tf + (1-t)f_{P,\mathcal{F}}$. Since $\mathcal{F}$ is convex we have $f_t \in \mathcal{F}$ for all $t \in [0,1]$. By the line-continuity of C and Lebesgue's theorem we find that the map $h : t \mapsto \mathbb{E}_P(C\circ f_t - C\circ f_{P,\mathcal{F}})$ is continuous for $t\in[0,1]$. Since $h(0) = 0$ and $h(1) \geq \varepsilon$ there is a $t \in (0,1]$ with

$$\mathbb{E}_P(C\circ f_t - C\circ f_{P,\mathcal{F}}) \;=\; h(t) \;=\; \varepsilon$$

by the intermediate value theorem. Moreover, for this t the convexity of C gives

$$\mathbb{E}_T(C\circ f_t - C\circ f_{P,\mathcal{F}}) \;\leq\; \mathbb{E}_T\Big(tC\circ f + (1-t)C\circ f_{P,\mathcal{F}} - C\circ f_{P,\mathcal{F}}\Big) \;\leq\; 0.$$

Now, let $\varepsilon > 0$ satisfy the assumption of the theorem. Then ε also satisfies the assumptions of Lemma 3.3, and hence we find that with probability at least $1-e^{-x}$ every $f\in\mathcal{F}$ with $\mathbb{E}_T(C\circ f - C\circ f_{P,\mathcal{F}}) \leq 0$ satisfies $\mathbb{E}_P(C\circ f - C\circ f_{P,\mathcal{F}}) < \varepsilon$. Since we always have

$$\mathbb{E}_T\big(C\circ f_{T,\mathcal{F}} - C\circ f_{P,\mathcal{F}}\big) \;\leq\; 0$$

we obtain the assertion. $\square$

4. Proof of the main result

In order to prove our oracle-type inequality we will apply Theorem 3.1. To this end we define the regularized cost function C_λ by

$$C_\lambda(x,y,f) := \lambda\|f\|_H^2 + L(y, f(x))\,, \qquad x\in X,\ y\in Y,\ f\in H,$$

and the induced cost class

$$\mathcal{G}(\lambda) := \big\{C_\lambda\circ f - C_\lambda\circ f_{P,\lambda} : f\in\lambda^{-1/2}B_H\big\}\,, \qquad \lambda > 0.$$

Obviously, the C_λ-risk minimizer produces the functions $f_{P,\lambda}$ and $f_{T,\lambda}$. Note that $\mathcal{R}_{L,P}(0) \leq 1$ implies $f_{P,\lambda}\in\lambda^{-1/2}B_H$ for all distributions P on $X\times Y$, and hence the latter in particular holds for the empirical solutions $f_{T,\lambda}$. However, it was already observed in [12] that, depending on the approximation error function, sharper bounds for $\|f_{T,\lambda}\|$ are possible with high probability. In order to establish such sharper bounds we employed a "shrinking technique" in [12] which is rather complicated. The key idea of this paper is to replace the shrinking technique by a localization argument based on (6). Consequently, let us first show that regularized risk minimizers always satisfy the supremum bound (6):

Lemma 4.1. *Let $0 < \lambda \leq 1$, and suppose that $g \in \mathcal{G}(\lambda)$. Then for any $f \in \lambda^{-\frac{1}{2}} B_H$ such that $g = C_\lambda \circ f - C_\lambda \circ f_{P,\lambda}$ we have*

$$\|g\|_\infty \leq 3\Big(\frac{\mathbb{E}_P g}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + 2 \quad \text{and}$$
$$\|f\|_H \leq \Big(\frac{a(\lambda) + \mathbb{E}_P g}{\lambda}\Big)^{1/2}.$$

Proof. Let us write $\varepsilon := \mathbb{E}_P g$. Then we have

$$\begin{aligned}\lambda\|f\|_H^2 &\leq \lambda\|f\|_H^2 + \mathcal{R}_{L,P}(f) - \mathcal{R}^*_{L,P} \\ &= \lambda\|f_{P,\lambda}\|^2 + \mathcal{R}_{L,P}(f_{P,\lambda}) - \mathcal{R}^*_{L,P} + \varepsilon \\ &= a(\lambda) + \varepsilon\,,\end{aligned}$$

which establishes the second assertion. Consequently, $\|L \circ f\|_\infty \leq 1 + \|f\|_\infty^\alpha$ yields

$$\|C_\lambda \circ f\|_\infty \leq \lambda\|f\|_H^2 + \|L \circ f\|_\infty \leq a(\lambda) + \varepsilon + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{\varepsilon}{\lambda}\Big)^{\frac{\alpha}{2}} + 1\,.$$

Analogously, we obtain $\|C_\lambda \circ f_{P,\lambda}\|_\infty \leq a(\lambda) + \big(\frac{a(\lambda)}{\lambda}\big)^{\frac{\alpha}{2}} + 1$, and therefore we find

$$\|g\|_\infty \leq \max\left(\|C_\lambda \circ f\|_\infty, \|C_\lambda \circ f_{P,\lambda}\|_\infty\right) \leq \varepsilon + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{\varepsilon}{\lambda}\Big)^{\frac{\alpha}{2}} + 2\,,$$

where in the last step we used $a(\lambda) \leq 1$. Now, $f \in \lambda^{-1/2} B_H$ implies that $\|f\|_\infty \leq \lambda^{-1/2}$ and an easy calculation shows that $2 + \lambda^{-\alpha/2} \leq 3\lambda^{-\frac{\alpha}{2-\alpha}}$. Therefore we obtain

$$\varepsilon \leq \mathbb{E}_P C_\lambda \circ f = \lambda\|f\|_H^2 + \mathcal{R}_{L,P}(f) \leq 2 + \|f\|_\infty^\alpha \leq 2 + \lambda^{-\frac{\alpha}{2}} \leq 3\lambda^{-\frac{\alpha}{2-\alpha}}\,.$$

From this we easily obtain $\varepsilon \leq 3^{1-\frac{\alpha}{2}}(\frac{\varepsilon}{\lambda})^{\frac{\alpha}{2}} \leq 2(\frac{\varepsilon}{\lambda})^{\frac{\alpha}{2}}$, which gives the assertion. □

We now prove that a variance bound of the form (5) assumed in Theorem 2.1 implies a variance bound of the form (7) assumed in Theorem 3.1:

Lemma 4.2. *Let P be a distribution on $X \times Y$ and suppose that there exist constants $v \geq 0$, $c \geq 1$, and $\vartheta \in [0,1]$ such that the variance bound assumption (5) is satisfied for some $0 < \lambda < 1$ and all $f \in \lambda^{-\frac{1}{2}} B_H$. Then for all $g \in \mathcal{G}(\lambda)$ we have*

$$\mathbb{E}_P g^2 \leq 16c\bigg(\Big(\frac{\mathbb{E}_P g}{\lambda}\Big)^{\frac{1}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{1/2} + 1\bigg)^v \Big((\mathbb{E}_P g)^\vartheta + 2a^\vartheta(\lambda)\Big)\,.$$

Proof. We use the shorthand notation $\mathbb{E}$ for $\mathbb{E}_P$. For $g \in \mathcal{G}(\lambda)$ pick an $f \in \lambda^{-\frac{1}{2}} B_H$ such that $g = C_\lambda \circ f - C_\lambda \circ f_{P,\lambda}$. Now observe that

$$\begin{aligned}\mathbb{E}g^2 &= \mathbb{E}\big(C_\lambda \circ f - C_\lambda \circ f_{P,\lambda}\big)^2 \\ &= \mathbb{E}\big(\lambda\|f\|^2 - \lambda\|f_{P,\lambda}\|^2 + L \circ f - L \circ f_{P,\lambda}\big)^2 \\ &\leq 2\mathbb{E}\big(\lambda\|f\|^2 - \lambda\|f_{P,\lambda}\|^2)^2 + 2\mathbb{E}\big(L \circ f - L \circ f_{P,\lambda}\big)^2 \\ &\leq 2\lambda^2\|f\|^4 + 2\lambda^2\|f_{P,\lambda}\|^4 + 2\mathbb{E}\big(L \circ f - L \circ f_{P,\lambda}\big)^2 \\ &\leq 4\mathbb{E}\big(L \circ f - L \circ f^*_{L,P}\big)^2 + 4\mathbb{E}\big(L \circ f^*_{L,P} - L \circ f_{P,\lambda}\big)^2 + 2\lambda^2\|f\|^4 + 2\lambda^2\|f_{P,\lambda}\|^4.\end{aligned}$$

Denote $C := \max\Big(\|f\|_\infty + 1, \|f_{P,\lambda}\|_\infty + 1\Big)$. Then the assumption (5) and $a^\vartheta + b^\vartheta \le 2(a+b)^\vartheta$ for all $a, b \ge 0$, imply that

$$\begin{aligned}\mathbb{E}\big(L\circ f - L\circ f^*_{L,P}\big)^2 + \mathbb{E}\big(L\circ f^*_{L,P} - L\circ f_{P,\lambda}\big)^2 \\ \le 2cC^v\Big(\mathbb{E}\big(L\circ f - L\circ f^*_{L,P}\big) + \mathbb{E}\big(L\circ f_{P,\lambda} - L\circ f^*_{L,P}\big)\Big)^\vartheta .\end{aligned}$$

Since $\lambda^2\|f\|^4 \le 1$ and $\lambda^2\|f_{P,\lambda}\|^4 \le 1$ we hence obtain

$$\begin{aligned}\mathbb{E}g^2 &\le 8cC^v\Big(\mathbb{E}\big(L\circ f - L\circ f^*_{L,P}\big) + \mathbb{E}\big(L\circ f_{P,\lambda} - L\circ f^*_{L,P}\big)\Big)^\vartheta + 2\lambda^2\|f\|^4 + 2\lambda^2\|f_{P,\lambda}\|^4 \\ &\le 8cC^v\Big(\mathbb{E}\big(L\circ f - L\circ f^*_{L,P}\big) + \mathbb{E}\big(L\circ f_{P,\lambda} - L\circ f^*_{L,P}\big)\Big)^\vartheta + 4\Big(\lambda^2\|f\|^4 + \lambda^2\|f_{P,\lambda}|^4\Big)^\vartheta \\ &\le 16cC^v\Big(\mathbb{E}\big(L\circ f - L\circ f^*_{L,P}\big) + \mathbb{E}\big(L\circ f_{P,\lambda} - L\circ f^*_{L,P}\big) + \lambda^2\|f\|^4 + \lambda^2\|f_{P,\lambda}\|^4\Big)^\vartheta \\ &= 16cC^v\Big(\mathbb{E}g + 2\mathbb{E}\big(L\circ f_{P,\lambda} - L\circ f^*_{L,P}\big) + 2\lambda\|f_{P,\lambda}\|^2\Big)^\vartheta \\ &\le 16cC^v\Big((\mathbb{E}g)^\vartheta + 2a^\vartheta(\lambda)\Big).\end{aligned}$$

What is left is to bound C in the right hand side of this inequality. To that end observe that Lemma 4.1 implies

$$\|f\|_\infty \le \|f\|_H \le \Big(\frac{a(\lambda) + \mathbb{E}g}{\lambda}\Big)^{1/2}$$

and

$$\|f_{P,\lambda}\|_\infty \le \|f_{P,\lambda}\|_H \le \Big(\frac{a(\lambda)}{\lambda}\Big)^{1/2} \le \Big(\frac{a(\lambda) + \mathbb{E}g}{\lambda}\Big)^{1/2}$$

so that we can bound

$$\begin{aligned}C &= \max\Big(\|f\|_\infty + 1, \|f_{P,\lambda}\|_\infty + 1\Big) \\ &\le \Big(\frac{a(\lambda) + \mathbb{E}g}{\lambda}\Big)^{1/2} + 1 \le \Big(\frac{\mathbb{E}g}{\lambda}\Big)^{1/2} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{1/2} + 1.\end{aligned}$$

□

The following lemma relates the covering numbers of B_H with $\omega_{P,n}(\mathcal{G}(\lambda), \varepsilon)$:

Lemma 4.3. *Let $n \in \mathbb{N}$, and assume that there are constants $a \ge 1$ and $p \in (0,2)$ such that for all $\delta > 0$, we have*

$$\sup_{T\in Z^n} \log \mathcal{N}\big(B_H, \delta, L_2(T_X)\big) \;\le\; a\delta^{-p}.$$

Then there is a constant $c_{L,p} > 0$ depending only on L and p such that for all distributions P on $X \times Y$, and all $\lambda \in (0,1]$, $\varepsilon > 0$ we have

$$\omega_{P,n}(\mathcal{G}(\lambda), \varepsilon) \le c_{L,p}\max\bigg\{\Big(\frac{a(\lambda)+\varepsilon}{\lambda} + 1\Big)^{\frac{\alpha p}{4}} \tau_\varepsilon^{\frac{2-p}{4}} \Big(\frac{a}{n}\Big)^{\frac{1}{2}}, \Big(\frac{a(\lambda)+\varepsilon}{\lambda} + 1\Big)^{\frac{\alpha}{2}} \Big(\frac{a}{n}\Big)^{\frac{2}{2+p}}\bigg\},$$

where $\tau_\varepsilon \ge \sup_{g\in\mathcal{G}_\varepsilon} \mathbb{E}_P g^2$ and $\mathcal{G}_\varepsilon := \{g \subset \mathcal{G}(\lambda) : \mathbb{E}_P g \le \varepsilon\}$.

Proof. Our first goal is to bound the covering numbers of $\mathcal{G}_\varepsilon$. To this end recall that for $g := C_\lambda \circ f - C_\lambda \circ f_{P,\lambda} \in \mathcal{G}_\varepsilon$, Lemma 4.1 shows that $\|f\|_H \le \big(\frac{a(\lambda)+\varepsilon}{\lambda}\big)^{1/2} =: \Lambda$.

With the help of the auxiliary sets $\hat{\mathcal{G}}_\varepsilon := \{C_\lambda \circ f : f \in \Lambda B_H\}$ and $\mathcal{H} := \{L \circ f : f \in \Lambda B_H\}$ we thus obtain

$$\begin{aligned}\log \mathcal{N}\big(\mathcal{G}_\varepsilon, 2\delta, L_2(T)\big) &\leq \log \mathcal{N}\big(\hat{\mathcal{G}}_\varepsilon, 2\delta, L_2(T)\big)\\ &\leq \log\Big(\frac{1}{\delta}+1\Big) + \log \mathcal{N}\big(\mathcal{H}, \delta, L_2(T)\big)\\ &\leq \log\Big(\frac{1}{\delta}+1\Big) + \log \mathcal{N}\Big(\Lambda B_H, \frac{\delta}{|L_{|[-\Lambda,\Lambda]}|_1}, L_2(T_X)\Big).\end{aligned}$$

Furthermore, the Lipschitz assumption (1) implies the right hand side is bounded by

$$\log\Big(\frac{1}{\delta}+1\Big) + \log \mathcal{N}\Big(B_H, \frac{\delta}{c_L \Lambda^\alpha}, L_2(T_X)\Big).$$

Consequently, there is a constant $\tilde{c}_{L,p} > 0$ depending only on L and p such that for all $\delta > 0$ we have

$$\sup_{T \in Z^n} \log \mathcal{N}\big(\mathcal{G}_\varepsilon, \delta, L_2(T)\big) \leq a\, \tilde{c}_{L,p} \Big(\frac{a(\lambda)+\varepsilon}{\lambda}\Big)^{\frac{\alpha p}{2}} \delta^{-p} \leq a\, \tilde{c}_{L,p} \Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac{\alpha p}{2}} \delta^{-p}.$$

By symmetrization, and the proofs of [9, Lem. 2.5] and [12, Prop. 5.7] we thus find

$$\omega_{P,n}(\mathcal{G}(\lambda), \varepsilon) \leq c_{L,p} \max\Big\{ \Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac{\alpha p}{4}} \tau_\varepsilon^{\frac{2-p}{4}} \Big(\frac{a}{n}\Big)^{\frac{1}{2}}, \Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac{\alpha}{2}} \Big(\frac{a}{n}\Big)^{\frac{2}{2+p}} \Big\}.$$

□

Proof of Theorem 2.1. Let $g := C_\lambda \circ f - C_\lambda \circ f_{P,\lambda}$ for some $f \in \lambda^{-1/2} B_H$. Lemma 4.1 implies that we have a supremum bound

$$\|g\|_\infty \;\leq\; 3\Big(\frac{\mathbb{E}_P g}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + 2.$$

Because of the variance bound assumption (5), Lemma 4.2 implies we have a variance bound of the form

$$\begin{aligned}\mathbb{E}_P g^2 &\leq 16c\Big(\Big(\frac{\mathbb{E}_P g}{\lambda}\Big)^{\frac{1}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{1/2} + 1\Big)^v \Big((\mathbb{E}_P g)^\vartheta + 2a^\vartheta(\lambda)\Big)\\ &\leq 48c\Big(\Big(\frac{\mathbb{E}_P g}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + 1\Big)^{\frac{v}{\alpha}} \Big((\mathbb{E}_P g)^\vartheta + 2a^\vartheta(\lambda)\Big)\\ &\leq \Big(3\Big(\frac{\mathbb{E}_P g}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + 2\Big)^{\frac{v}{\alpha}} \Big(48c(\mathbb{E}_P g)^\vartheta + 96ca^\vartheta(\lambda)\Big).\end{aligned}$$

Therefore we have variance and supremum bounds of the form (7) and (6) with the values $b = 3\lambda^{-\frac{\alpha}{2}}$, $\beta = \frac{\alpha}{2}$, $B = (\frac{a(\lambda)}{\lambda})^{\frac{\alpha}{2}} + 2$, $w = 48c$, $\nu = \frac{v}{\alpha}$, and $W = 96ca^\vartheta(\lambda)$. Denote $\tau_\varepsilon := 3^4 2^6 c \lambda^\vartheta (\frac{a(\lambda)+\epsilon}{\lambda}+1)^{\vartheta+\frac{v}{2}}$. Then for $g \in \mathcal{G}(\lambda)$ with $\mathbb{E}_P g \leq \varepsilon$ we obtain

$$\begin{aligned}\mathbb{E}_P g^2 &\leq (b\varepsilon^\beta + B)^\nu (w\varepsilon^\vartheta + W)\\ &= 48c\Big(3\Big(\frac{\varepsilon}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + 2\Big)^{\frac{v}{\alpha}} \Big(\varepsilon^\vartheta + 2a^\vartheta(\lambda)\Big)\\ &\leq 96 \cdot 9c\Big(\Big(\frac{\varepsilon}{\lambda}\Big)^{\frac{\alpha}{2}} + \Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}} + 1\Big)^{\frac{v}{\alpha}} \Big(\varepsilon^\vartheta + a^\vartheta(\lambda)\Big)\\ &\leq 96 \cdot 9 \cdot 3 \cdot 2c\Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac{v}{2}} \Big(a(\lambda)+\varepsilon\Big)^\vartheta\\ &\leq \tau_\varepsilon.\end{aligned}$$

Consequently we can apply Lemma 4.3 to obtain that

$$\begin{aligned}
&\omega_{P,n}(\mathcal{G}(\lambda),\varepsilon)\\
&\quad\le c_{L,p}\max\left\{\Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac{p}{4}}\tau_\varepsilon^{\frac{2-p}{4}}\Big(\frac{a}{n}\Big)^{\frac12},\Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac12}\Big(\frac{a}{n}\Big)^{\frac{2}{2+p}}\right\}\\
&\quad\le c_{L,p}\sqrt{c}\,\max\left\{\lambda^{\frac{\vartheta(2-p)}{4}}\Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac{2\alpha p+(v+2\vartheta)(2-p)}{8}}\Big(\frac{a}{n}\Big)^{\frac12},\Big(\frac{a(\lambda)+\varepsilon}{\lambda}+1\Big)^{\frac{\alpha}{2}}\Big(\frac{a}{n}\Big)^{\frac{2}{2+p}}\right\}.
\end{aligned}$$

We also bound the terms

$$\sqrt{\frac{2x(b\varepsilon^\beta+B)^\nu(w\varepsilon^\vartheta+W)}{n}}\le\sqrt{\frac{2x\tau_\varepsilon}{n}}=384\sqrt{2c}\lambda^{\frac{\vartheta}{2}}\left(\frac{a(\lambda)+\varepsilon}{\lambda}+1\right)^{\frac{\vartheta}{2}+\frac{v}{4}}\sqrt{\frac{x}{n}}$$

and

$$\frac{2x\left(b\varepsilon^\beta+B\right)}{n}=\frac{2x}{n}\left(3\Big(\frac{\epsilon}{\lambda}\Big)^{\frac{\alpha}{2}}+\Big(\frac{a(\lambda)}{\lambda}\Big)^{\frac{\alpha}{2}}+2\right)\le\frac{24x}{n}\Big(\frac{a(\lambda)+\epsilon}{\lambda}+1\Big)^{\frac{\alpha}{2}}$$

and then observe that Theorem 3.1 implies that there is a constant $K\ge 1$ such that

$$\Pr{}^*\Big(T\in(X\times Y)^n:\mathcal{R}_{C_\lambda,P}(f_{T,\lambda})<\mathcal{R}_{C_\lambda,P}(f_{P,\lambda})+\tilde\varepsilon\Big)\ \ge\ 1-e^{-x}\,,$$

whenever

$$\begin{aligned}
\tilde\varepsilon\ge K\max\Bigg\{&\lambda^{\frac{\vartheta(2-p)}{4}}\Big(\frac{a(\lambda)+\tilde\varepsilon}{\lambda}+1\Big)^{\frac{2\alpha p+(v+2\vartheta)(2-p)}{8}}\Big(\frac{a}{n}\Big)^{\frac12},\Big(\frac{a(\lambda)+\tilde\varepsilon}{\lambda}+1\Big)^{\frac{\alpha}{2}}\Big(\frac{a}{n}\Big)^{\frac{2}{2+p}},\\
&\lambda^{\frac{\vartheta}{2}}\Big(\frac{a(\lambda)+\tilde\varepsilon}{\lambda}+1\Big)^{\frac{\vartheta}{2}+\frac{v}{4}}\Big(\frac{x}{n}\Big)^{\frac12},\Big(\frac{a(\lambda)+\tilde\varepsilon}{\lambda}+1\Big)^{\frac{\alpha}{2}}\frac{x}{n}\Bigg\}.
\end{aligned}$$

If we further constrain by $\tilde\varepsilon\ge a(\lambda)+\lambda$ we find that it is sufficient to satisfy

$$\begin{aligned}
\tilde\varepsilon>\max\Bigg\{&a(\lambda)+\lambda,K\lambda^{\frac{\vartheta(2-p)}{4}}\Big(\frac{\tilde\varepsilon}{\lambda}\Big)^{\frac{2\alpha p+(v+2\vartheta)(2-p)}{8}}\Big(\frac{a}{n}\Big)^{\frac12},K\Big(\frac{\tilde\varepsilon}{\lambda}\Big)^{\frac{\alpha}{2}}\Big(\frac{a}{n}\Big)^{\frac{2}{2+p}},\\
&K\lambda^{\frac{\vartheta}{2}}\Big(\frac{\tilde\varepsilon}{\lambda}\Big)^{\frac{\vartheta}{2}+\frac{v}{4}}\Big(\frac{x}{n}\Big)^{\frac12},K\Big(\frac{\tilde\varepsilon}{\lambda}\Big)^{\frac{\alpha}{2}}\frac{x}{n}\Bigg\}.
\end{aligned}$$

Since $\vartheta\in(0,1]$ and $v\in[0,2]$ it follows that $0<v+2\vartheta\le 4$ which implies that $\frac{2\alpha p+(v+2\vartheta)(2-p)}{8}\le 1$ and $\frac{\vartheta}{2}+\frac{v}{4}\le 1$. Consequently we find that it is sufficient to satisfy

$$\begin{aligned}
\tilde\varepsilon\ge\max\Bigg\{&a(\lambda)+\lambda,\left(\frac{K^2a}{\lambda^{\frac{2\alpha p+v(2-p)}{4}}n}\right)^{\frac{4}{8-2\alpha p-(v+2\vartheta)(2-p)}},\left(\frac{K^{\frac{2+p}{2}}a}{\lambda^{\frac{\alpha(2+p)}{4}}n}\right)^{\frac{4}{(2+p)(2-\alpha)}},\\
&\left(\frac{K^2x}{\lambda^{\frac{v}{2}}n}\right)^{\frac{2}{4-(v+2\vartheta)}},\left(\frac{Kx}{\lambda^{\frac{\alpha}{2}}n}\right)^{\frac{2}{2-\alpha}}\Bigg\}.
\end{aligned}$$

Therefore we find that (with a change in the value of the constant K) if

$$\begin{aligned}
\varepsilon\ge\max\Bigg\{&a(\lambda)+\lambda,\left(\frac{Ka}{\lambda^{\frac{2\alpha p+v(2-p)}{4}}n}\right)^{\frac{4}{8-2\alpha p-(v+2\vartheta)(2-p)}},\left(\frac{Ka}{\lambda^{\frac{\alpha(2+p)}{4}}n}\right)^{\frac{4}{(2+p)(2-\alpha)}},\\
&\left(\frac{Kx}{\lambda^{\frac{v}{2}}n}\right)^{\frac{2}{4-(v+2\vartheta)}},\left(\frac{Kx}{\lambda^{\frac{\alpha}{2}}n}\right)^{\frac{2}{2-\alpha}}\Bigg\}
\end{aligned}$$

then

$$\mathcal{R}_{L,P}(f_{T,\lambda}) \le \lambda\|f_{T,\lambda}\|_H^2 + \mathcal{R}_{L,P}(f_{T,\lambda}) = \mathcal{R}_{C_\lambda,P}(f_{T,\lambda}) < \mathcal{R}_{C_\lambda,P}(f_{P,\lambda}) + \varepsilon$$
$$= a(\lambda) + \mathcal{R}^*_{L,P} + \varepsilon$$

holds with probability not less than $1 - e^{-x}$. □

5. Examples

Here we perform the analysis mentioned in Examples 2.2 and 2.3. Let us first apply the oracle inequality to bound $\mathcal{R}_{L_\alpha,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_\alpha,P}$ with high probability. To that end we now derive some variance bounds. First observe that [11, Table 3] shows that the modulus of convexity $\delta_{\psi_\alpha|[-B,B]}(\varepsilon)$ of the function $\psi_\alpha : t \mapsto |t|^\alpha$ restricted to the interval $[-B, B]$ satisfies

$$\delta_{\psi_\alpha|[-B,B]}(\varepsilon) \ge \frac{\alpha(\alpha-1)}{8} B^{\alpha-2}\varepsilon^2 \tag{8}$$

Consequently [2, Lemma 15] implies that modulus of convexity of $\mathcal{R}_{L_\alpha,P}$ for functions satisfying $\|f\|_\infty \le B$ is bounded below by $\frac{\alpha(\alpha-1)}{8}2^{\alpha-2}B^{\alpha-2}\varepsilon^2 \ge \frac{\alpha(\alpha-1)}{16} \times B^{\alpha-2}\varepsilon^2$. Moreover, the mean value theorem implies that

$$\Big||t_1-y|^\alpha - |t_2-y|^\alpha\Big| \le \alpha\Big(\max(t_1+1, t_2+1)\Big)^{\alpha-1}|t_1-t_2|$$

so that the loss function $f \mapsto L_\alpha(y, f(x))$ has a Lipschitz constant less than $\alpha\big(\max\{\|f_1\|_\infty, \|f_2\|_\infty\}+1\big)^{\alpha-1}$. Now let

$$f^*_{L_\alpha,P} \in \arg\min\big\{\mathcal{R}_{L_\alpha,P}(f)\big|f : X \to \mathbb{R} \text{ measurable}\big\}$$

and define $g_f(x,y) := |f(x)-y|^\alpha - |f^*_{L_\alpha,P}(x)-y|^\alpha$. Then the extension mentioned after the statement of [2, Lemma 14] to non-margin loss functions implies that we have the variance bound

$$\mathbb{E}g_f^2 \le \frac{8\alpha}{(\alpha-1)} \cdot \frac{\big(\max\{\|f\|_\infty, \|f^*_{L_\alpha,P}\|_\infty\}+1\big)^{2\alpha-2}}{\big(\max\{\|f\|_\infty, \|f^*_{L_\alpha,P}\|_\infty\}\big)^{\alpha-2}}\mathbb{E}g_f$$
$$\le \frac{8\alpha}{(\alpha-1)}\big(\max\{\|f\|_\infty, \|f^*_{L_\alpha,P}|_\infty\}+1\big)^\alpha \mathbb{E}g_f..$$

Observe that the right hand side of these bounds goes to ∞ as $\alpha \to 1$ since ψ_1 is not strictly convex. Also note that such a bound, but with different constants, follows directly from [11, Equation 28]. Since $\|f^*_{L_\alpha,P}\|_\infty \le 1$ we then obtain

$$\mathbb{E}g_f^2 \le \frac{8\alpha}{(\alpha-1)}\big(\|f\|_\infty + 2\big)^\alpha\, \mathbb{E}g_f\,..$$

Therefore we can apply Theorem 2.1 with $v = \alpha$ and $\vartheta = 1$ to obtain that there exists a constant $K_\alpha \ge 1$ such that for all $0 < \lambda \le 1$, $\varepsilon > 0$, $x \ge 1$ satisfying

$$\varepsilon \ge \max\left\{a_\alpha(\lambda) + \lambda, \left(\frac{K_\alpha a}{\lambda^{\frac{\alpha(2+p)}{4}} n}\right)^{\frac{4}{(2+p)(2-\alpha)}}, \left(\frac{K_\alpha x}{\lambda^{\frac{\alpha}{2}} n}\right)^{\frac{2}{2-\alpha}}\right\}, \tag{9}$$

we have

$$\Pr^*\Big(T \in Z^n : \mathcal{R}_{L_\alpha,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_\alpha,P} < a_\alpha(\lambda) + \varepsilon\Big) \;\geq\; 1 - e^{-x} \tag{10}$$

where $a_\alpha(\cdot)$ is the approximation error function defined with respect to the risk $\mathcal{R}_{L_\alpha,P}$.

In [13] it was shown that the assumption $f^*_{L_\alpha,P} \in H$ implies that $a_\alpha(\lambda) \leq \lambda\|f^*_{L_\alpha,P}\|^2_H$ for all $\lambda > 0$. We assume without loss of generality that $a_\alpha(\lambda) \leq \lambda$. Let us first consider when $\alpha = 2$. If we now assume (λ_n) is a strictly positive null-sequence with $\lambda_n^{1+p/2} n \to \infty$ then it is easy from the convention (3) applied to the inequality (9) that our learning rate is of the form λ_n thus finishing the proof for Example 2.2. Now consider the case $1 < \alpha < 2$. Then (9) becomes

$$\varepsilon \geq \max\left\{a_\alpha(\lambda) + \lambda, \lambda^{-\frac{\alpha}{2-\alpha}}\left(\frac{K_\alpha a}{n}\right)^{\frac{4}{(2+p)(2-\alpha)}}, \lambda^{-\frac{\alpha}{2-\alpha}}\left(\frac{K_\alpha x}{n}\right)^{\frac{2}{2-\alpha}}\right\}. \tag{11}$$

Moreover when $n \geq K_\alpha a$ elementary calculations show that it is sufficient to satisfy

$$\varepsilon \geq a_\alpha(\lambda) + \lambda + \lambda^{-\frac{\alpha}{2-\alpha}} x^{\frac{2}{2-\alpha}}\left(\frac{K_\alpha a}{n}\right)^{\frac{4}{(2+p)(2-\alpha)}}.. \tag{12}$$

If we now assume $\lambda = n^{-\kappa}$. Then elementary calculations show that we obtain the rate $n^{-\kappa}$ independently of the value α when $\kappa \leq \frac{2}{2+p}$ and when $\kappa > \frac{2}{2+p}$ we obtain the rate $n^{-\frac{2}{2+p}+\frac{\alpha}{2-\alpha}(\kappa-\frac{2}{2+p})}$.

Let us now assume that the conditional distributions $P(y|x)$ are symmetric. We now proceed to derive a calibration inequality

$$\mathcal{R}_{L_2,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_2,P} \leq \Psi(\mathcal{R}_{L_\alpha,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_\alpha,P})$$

so that we can apply the bounds on $\mathcal{R}_{L_\alpha,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_\alpha,P})$ defined by (10) and (12) to obtain bounds on $\mathcal{R}_{L_2,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_2,P}$ in terms of α. Since we will need results and notations from [11] we first give a brief outline of its content. Consider a loss function L and a measure Q on Y. Then the associated inner risk is defined as

$$\mathcal{C}_{L,Q}(t) = \int_Y L(y,t)dQ(y), \quad t \in \mathbb{R},$$

and can be used to compute the risk

$$\mathcal{R}_{L,P}(f) = \int_X \mathcal{C}_{L,P(\cdot|x)}(f(x))dP_X(x).$$

The minimal inner risk is defined as $\mathcal{C}^*_{L,Q} := \inf_{t\in\mathbb{R}} \mathcal{C}_{L,Q}(t)$. Consider now another loss function $\acute{L}$. Then the calibration function $\delta_{\max,L,\acute{L}}(\varepsilon, Q)$ is defined as the largest function comparing the excess inner risks, i.e.

$$\delta_{\max,L,\acute{L}}(\mathcal{C}_{L,Q}(t) - \mathcal{C}^*_{L,Q}, Q) \leq \mathcal{C}_{\acute{L},Q}(t) - \mathcal{C}^*_{\acute{L},Q}.$$

We shall also find it convenient to consider the template loss L_{mean} introduced in [11] and defined by

$$L_{\text{mean}}(Q,t) := |\mathbb{E}Q - t|, \quad t \in \mathbb{R}$$

and its inner risk

$$\mathcal{C}_{L_{\text{mean}},Q}(t) = \int_Y |\mathbb{E}Q - t| dQ(y), \quad t \in \mathbb{R}.$$

We can now proceed to derive the appropriate calibration inequality function Ψ for comparing L_2 and L_α. Since $P(y|x)$ is symmetric for all x, [11, Theorem 3.23] implies that we have mean calibration with calibration function bounded below by

$$\delta_{\max, L_{\text{mean}}, L_\alpha}(\varepsilon, Q) \geq \delta_{\psi_\alpha|[-(2+\varepsilon),2+\varepsilon]}(2\varepsilon)$$

where $\delta_{\psi_\alpha|[-(2+\varepsilon),2+\varepsilon]}$ is the modulus of convexity of the function ψ_α restricted to the interval $[-(2+\varepsilon), 2+\varepsilon]$. By (8) we then obtain

$$\delta_{\max, L_{\text{mean}}, \psi_\alpha}(\varepsilon, Q) \geq \frac{\alpha(\alpha-1)}{2}(2+\varepsilon)^{\alpha-2}\varepsilon^2.$$

Since [11, Equation (38)] states $\delta_{\max, L_2, L_\alpha}(\varepsilon, Q) = \delta_{\max, L_{\text{mean}}, L_\alpha}(\sqrt{\varepsilon}, Q)$ we find

$$\delta_{\max, L_2, L_\alpha}(\varepsilon, Q) \geq \frac{\alpha(\alpha-1)}{2}(2+\sqrt{\varepsilon})^{\alpha-2}\varepsilon.$$

We now seek to apply [11, Theorem 2.13]. In that notation we bound

$$B_f = \sup_x \left|f(x) - \mathbb{E}(y|x)\right|^2 \leq \left|\|f\|_\infty + 1\right|^2.$$

Denote $\phi(\varepsilon) := \frac{\alpha(\alpha-1)}{2}(2+\sqrt{\varepsilon})^{\alpha-2}\varepsilon$. Then since

$$\frac{d}{d\varepsilon}\Big((2+\sqrt{\varepsilon})^{\alpha-2}\varepsilon\Big) = (2+\sqrt{\varepsilon})^{\alpha-3}(2+\frac{\alpha}{2}\sqrt{\varepsilon}) > 0$$

and

$$\frac{d^2}{d\varepsilon^2}\Big((2+\sqrt{\varepsilon})^{\alpha-2}\varepsilon\Big) = (\alpha-2)\varepsilon^{-\frac{1}{2}}\Big(\frac{3}{2}+\frac{\alpha}{4}\sqrt{\varepsilon}\Big)(2+\sqrt{\varepsilon})^{\alpha-4} \leq 0$$

we conclude that ϕ is strictly monotonically increasing and concave. It follows that

$$\phi^{**}_{B_f}(\varepsilon) \geq \phi^{**}_{|\|f\|_\infty+1|^2}(\varepsilon) = \frac{\phi(\left|\|f\|_\infty+1\right|^2)}{\left|\|f\|_\infty+1\right|^2}\varepsilon = \frac{\alpha(\alpha-1)}{2}(3+\|f\|_\infty)^{\alpha-2}\varepsilon$$

where ** denotes the Fenchel-Legendre bi-conjugate operation (see e.g. [10]). It then follows from [11, Theorem 2.13] that

$$\mathcal{R}_{L_2,P}(f) - \mathcal{R}^*_{L_2,P} \leq \frac{2}{\alpha(\alpha-1)}(3+\|f\|_\infty)^{2-\alpha}(\mathcal{R}_{L_\alpha,P}(f) - \mathcal{R}^*_{L_\alpha,P}) \tag{13}$$

for all bounded measurable functions f. Note that the constant in this inequality goes to ∞ as α goes to 1. The deeper reason for this behaviour is that ψ_α is strictly convex when $\alpha > 1$ but not strictly convex when $\alpha = 1$ as discussed in [11].

We conclude from inequalities (13) and (10) that whenever (12) is satisfied that with probability greater than 1 e^{-x} we have

$$\mathcal{R}_{L_2,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_2,P} \leq \frac{4}{\alpha(\alpha-1)}(3+\|f_{T,\lambda}\|_\infty)^{2-\alpha}\varepsilon.$$

However we also know from the last line of the proof of Theorem 2.1 that whenever (12) is satisfied that with probability greater than $1 - e^{-x}$

$$\|f_{T,\lambda}\|_\infty \leq \|f_{T,\lambda}\|_H \leq \sqrt{\frac{a_\alpha(\lambda)+\varepsilon}{\lambda}} \leq \sqrt{2}\sqrt{\frac{\varepsilon}{\lambda}}.$$

Now since $1 \leq \frac{\varepsilon}{\lambda}$ when (12) is satisfied it follows that

$$(3+\|f_{T,\lambda}\|_\infty)^{2-\alpha} \leq \left(3+\sqrt{2}\sqrt{\frac{\varepsilon}{\lambda}}\right)^{2-\alpha} \leq \left(3+\sqrt{2}\right)\left(\frac{\varepsilon}{\lambda}\right)^{1-\frac{\alpha}{2}}$$

so that with probability greater than $1-2e^{-x}$ we have

$$\mathcal{R}_{L_2,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_2,P} \leq \frac{4}{\alpha(\alpha-1)}(3+\sqrt{2})\lambda^{\frac{\alpha}{2}-1}\varepsilon^{2-\frac{\alpha}{2}}.$$

If we now apply the inequality $a_\alpha(\lambda) \leq \lambda$ and let

$$\varepsilon := 2\lambda + \lambda^{-\frac{\alpha}{2-\alpha}} x^{\frac{2}{2-\alpha}} \left(\frac{K_\alpha a}{n}\right)^{\frac{4}{(2+p)(2-\alpha)}},$$

then we see that with probability greater than $1-2e^{-x}$ we have

$$\begin{aligned}
&\mathcal{R}_{L_2,P}(f_{T,\lambda}) - \mathcal{R}^*_{L_2,P} \\
&\leq \frac{4}{\alpha(\alpha-1)}(3+\sqrt{2})\lambda^{\frac{\alpha}{2}-1}\left(2\lambda + \lambda^{-\frac{\alpha}{2-\alpha}} x^{\frac{2}{2-\alpha}} \left(\frac{K_\alpha a}{n}\right)^{\frac{4}{(2+p)(2-\alpha)}}\right)^{2-\frac{\alpha}{2}} \\
&\leq c_\alpha\left(\lambda + \lambda^{-\frac{2}{2-\alpha}} x^{\frac{4-\alpha}{2-\alpha}} \left(\frac{K_\alpha a}{n}\right)^{\frac{2}{2+p}\frac{4-\alpha}{2-\alpha}}\right)
\end{aligned}$$

for some constant c_α which depends only on α.

Now let us consider the case when $\lambda = n^{-\kappa}$. Then disregarding the constants the righthand side becomes

$$n^{-\kappa} + n^{(\kappa - \frac{2}{2+p})\frac{2}{2-\alpha} - \frac{2}{2+p}}$$

so that we obtain performance bounds of the form $n^{-\rho}$ with

$$\rho = \min\left(\kappa, \frac{2}{2+p} + (\frac{2}{2+p} - \kappa)\frac{2}{2-\alpha}\right).$$

Simple calculations show that when $\kappa \leq \frac{2}{2+p}$ then $\rho = \kappa$ independently of the value of α and when $\kappa > \frac{2}{2+p}$ then $\rho = \frac{2}{2+p} + (\frac{2}{2+p} - \kappa)\frac{2}{2-\alpha}$. In the latter case it is important to observe that the rates get worse as α increases towards 2. Indeed one can show that $\rho \leq 0$ in the interval

$$2 - (\kappa - \frac{2}{2+p})(2+p) \leq \alpha \leq 2.$$

Moreover one can see that smaller α minimizes the sensitivity to the degree to which κ is greater than $\frac{2}{2+p}$.

Acknowledgment

Ingo Steinwart thanks T. Zhang and P. Bartlett for the fruitful discussions in Eindhoven and Albuquerque that helped to develop the double localization argument. He further thanks A. Caponnetto and V. Temlyakov for the discussions on optimal rates for Sobolev spaces at the Toyota Technological Institute at Chicago.

References

[1] Bartlett, P. L., Bousquet, O. and Mendelson, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537.

[2] Bartlett, P. L., Jordan, M. I. and McAuliffe, J. D. (2006). Convexity, classification, and risk bounds. *J. Amer. Statist. Assoc.* **101** 138–156.

[3] Blanchard, G., Bousquet, O. and Massart, P. (2004). Statistical performance of support vector machines. *Ann. Statist.*, submitted.

[4] Bousquet, O. (2002). A Bennet concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris* **334** 495–500.

[5] Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press.

[6] Christmann, A. and Steinwart, I. (2005). Consistency and robustness of kernel based regression. Los Alamos National Laboratory Technical Report LA-UR-04-8797. *Bernoulli*, submitted.

[7] DeVore, R., Kerkyacharian, G., Picard, D. and Temlyakov, V. Mathematical methods for supervised learning. Technical report, University of South Carolina, 2005. Available at *http://www.math.sc.edu/ devore/publications/DKPT7.pdf*.

[8] Girosi, F., Jones, M. and Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation* **7** 219–269.

[9] Mendelson, S. (2002). Improving the sample complexity using global data *IEEE Transactions on Information Theory* **48** 1977–1991.

[10] Rockafellar, R. T. and Wets, R. J.-B. (1998). Variational Analysis Springer.

[11] Steinwart, I. (2005). How to compare different loss functions and their risks. *Constructive Approximation*, to appear. *http://www.c3.lanl.gov/ ml/pubs/2005_loss/paper.pdf*.

[12] Steinwart, I. and Scovel, C. (2004). Fast rates for support vector machines using Gaussian kernels. *Ann. Statist.*, to appear. *http://www.c3.lanl.gov/ ingo/publications/ann-04a.pdf*.

[13] Steinwart, I. and Scovel, C. Fast rates for support vector machines. *Lecture Notes in Computer Science* **3559**, 279–294. Springer-Verlag, 18th Annual Conference on Learning Theory, COLT 2005. Eds. Peter Auer, Ron Meir.

[14] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76.