Wahrscheinlichkeitsrechnung 1
Zusammenfassung der Vorlesung vom Sommersemester 1994

Friedrich Graef
Institut für Angewandte Mathematik
der Universität Erlangen-Nürnberg

28. April 1995

Inhaltsverzeichnis

1 Wahrscheinlichkeitsräume

Definition 1.1 *Eine Menge $\mathcal{A}$ von Teilmengen einer nichtleeren Menge Ω heißt eine* Mengenalgebra *(über Ω), wenn gilt:*

$$\Omega \in \mathcal{A}$$
$$A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$$
$$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$$

Definition 1.2 *Besitzt eine Mengenalgebra $\mathcal{A}$ zusätzlich noch die Eigenschaft $A_n \in \mathcal{A}$ für $n = 1, 2, \ldots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, so heißt sie eine σ-Algebra.*

Bemerkungen und Bezeichnungen

1. Ω heißt *das sichere Ereignis*, da es stets eintritt. Sein Komplement $\overline{\Omega} = \emptyset$, die leere Menge, tritt nie ein und heißt daher *das unmögliche Ereignis.*

2. $\{\omega\}$ heißt *Elementarereignis.*

3. Die *Morgan*schen Regeln zeigen, daß mit Ereignissen $A, B, A_1, A_2, \ldots$ auch ihre Durchschnitte Ereignisse sind:

$$A \cap B = \overline{\overline{A} \cup \overline{B}}$$
$$\bigcap_{n=1}^{\infty} A_n = \overline{\bigcup_{n=1}^{\infty} \overline{A_n}}$$

4. Anstelle von $A \cap B$ schreiben wir auch kurz AB.

5. Zwei Mengen A, B heißen *disjunkt*, falls $A \cap B = \emptyset$. In diesem Fall schreiben wir $A + B$ anstelle von $A \cup B$.

6. Mengen $A_1, A_2, A_3, \ldots$ heißen *paarweise disjunkt*, falls für alle i, j mit $i \neq j$ die Mengen A_i und A_j disjunkt sind. Wir schreiben dann $\sum_{n=1}^{\infty} A_n$ anstelle von $\bigcup_{n=1}^{\infty} A_n$.

Definition 1.3 *Eine Funktion $P : \mathcal{A} \longrightarrow \mathbf{R}$ auf einer σ-Algebra $\mathcal{A}$ von Teilmengen einer Menge Ω mit den Eigenschaften*

$$0 \leq P(A) \leq 1 \tag{1.1}$$

für alle $A \in \mathcal{A}$,

$$P(\Omega) = 1 \tag{1.2}$$

für das sichere Ereignis,

$$P(A + B) = P(A) + P(B) \tag{1.3}$$

für disjunkte Ereignisse $A, B \in \mathcal{A}$ und

$$P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) \tag{1.4}$$

für paarweise disjunkte Ereignisse $A_n \in \mathcal{A}$ heißt eine Wahrscheinlichkeitsverteilung *auf $\mathcal{A}$.*

Die vier in Definition 1.3 beschriebenen Forderungen an P sind die von *Kolmogoroff* aufgestellten *Axiome* der Wahrscheinlichkeitsrechnung.

Definition 1.4 *Ein Tripel* $(\Omega, \mathcal{A}, P)$*, wo* Ω *eine nichtleere Menge,* $\mathcal{A}$ *eine* σ*-Algebra von Teilmengen von* Ω *und* P *eine Wahrscheinlichkeitsverteilung auf* $\mathcal{A}$ *ist, heißt ein* Wahrscheinlichkeitsraum.

Einige Rechenregeln für Wahrscheinlichkeitsverteilungen. Wegen $A + \overline{A} = \Omega$ folgt aus den Axiomen (1.2) und (1.3), daß $P(A) + P(\overline{A}) = 1$ oder

$$P(\overline{A}) = 1 - P(A). \tag{1.5}$$

Speziell für $A = \Omega$ erhält man daraus

$$P(\emptyset) = 0. \tag{1.6}$$

Ist $A \subset B$, so läßt sich B als $A + \overline{A}B$ darstellen. Das ergibt $P(B) = P(A) + P(\overline{A}B)$. Da Wahrscheinlichkeiten stets nichtnegativ sind, folgt daraus

$$A \subset B \Rightarrow P(A) \leq P(B). \tag{1.7}$$

Durch iterierte Anwendung von Axiom (1.3) erhält man

$$P(\sum_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k) \tag{1.8}$$

für jede endliche Anzahl n von paarweise disjunkten Ereignissen.
Für beliebige nicht notwendig disjunkte Ereignisse A, B ergibt sich aus $A \cup B = A\overline{B} + AB + \overline{A}B$ die Beziehung

$$P(A \cup B) = P(A) + P(B) - P(AB) \tag{1.9}$$

$A_1, A_2, \ldots$ sei eine Folge von Ereignissen mit $A_n \subset A_{n+1}$ für alle $n = 1, 2, \ldots$ und

$$A := \bigcup_{n=1}^{\infty} A_n.$$

Wir schreiben für diese Situation kurz $A_n \nearrow A$.
Es gilt in diesem Fall

$$A_n \nearrow A \Rightarrow \lim_{n\to\infty} P(A_n) = P(A) \tag{1.10}$$

Beweis: Nach (1.7) ist die Folge der Zahlen $P(A_n)$ monoton steigend und nach (1.1) ist sie durch 1 nach oben beschränkt. Der Limes existiert also. Wir definieren eine Folge von Ereignissen B_n durch $B_1 := A_1$ und $B_n := \overline{A_{n-1}}A_n$ für $n > 1$. Die B_n sind paarweise disjunkt und es gilt

$$\begin{aligned} A_n &= \sum_{k=1}^{n} B_k \\ A &= \sum_{k=1}^{\infty} B_k. \end{aligned}$$

Daraus ergibt sich aus den bisher abgeleiteten Rechenregeln

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) = \sum_{k=1}^{\infty} P(B_k) = P(A).$$

Für eine Folge $A_1, A_2, \ldots$ von Ereignissen mit $A_n \supset A_{n+1}$ für alle n und

$$A := \bigcap_{n=1}^{\infty} A_n$$

schreiben wir kurz $A_n \searrow A$. In diesem Fall gilt

$$A_n \searrow A \Rightarrow \lim_{n\to\infty} P(A_n) = P(A). \tag{1.11}$$

Beweis: Die Mengen $C_n := \overline{A_n}$ erfüllen $C_n \nearrow C$ mit $C = \overline{A}$. Wegen $P(C_n) = 1 - P(A_n)$ und $P(C) = 1 - P(A)$ ergibt sich (1.11) aus (1.10), angewandt auf die C_n.

2 Laplace-Experimente

Definition 2.1 *Ein* Laplace-Experiment *ist ein Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$ *mit den folgenden Eigenschaften:*

- Ω *ist eine endliche Menge*
- $\mathcal{A} = 2^{\Omega}$
- $P(A) = |A|/|\Omega|$

Urnenmodelle
In einer Urne befinden sich N durchnumerierte Kugeln, die mit den Nummern $1, \ldots, K$ sind schwarz, der Rest weiß. Es werden zufällig und ohne Zurücklegen n Kugeln aus der Urne genommen.
Ergebnismenge bei Berücksichtigung der Reihenfolge, in der die Kugeln aus der Urne gezogen werden, ist

$$\begin{aligned} \Omega_1 \;=\; & \{\omega = (\omega_1, \omega_2, \ldots, \omega_n)\,; \\ & \omega_i \in \{1, 2, \ldots, N\} \text{ paarweise verschieden } \} \end{aligned}$$

wobei ω_i die Nummer der i-ten gezogenen Kugel ist. Ω_1 enthält

$$|\Omega_1| = N \cdot (N-1) \cdot (N-2) \cdots (N-n+1)$$

Elemente.
Ergebnismengen, wenn die Reihenfolge nicht berücksichtigt wird:

$$\begin{aligned} \Omega_2 \;=\; & \{\omega = \{\omega_1, \omega_2, \ldots, \omega_n\}\,; \\ & \omega_i \in \{1, 2, \ldots, N\} \text{ paarweise verschieden } \} \end{aligned}$$

wobei $\omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ die Menge der Nummern der gezogenen Kugeln ist. Jedem solchen ω entspricht genau ein Vektor $\delta = (\delta_1, \delta_2, \ldots, \delta_N)$ mit Komponenten $\delta_i \in \{0, 1\}$, wobei die n Komponenten $\delta_{\omega_i} = 1$ sind. Daher kann man als äquivalente Ergebnismenge auch

$$\Omega_3 = \{\delta = (\delta_1, \delta_2, \ldots, \delta_N)\,;\ \sum_{i=1}^{N} \delta_i \;=\; n\}$$

benutzen. Die Anzahl $|\Omega_3|$ der Elemente dieser Menge ist durch den *Binomialkoeffizienten* $\binom{N}{n}$ gegeben. Einzelheiten zu den Binomialkoeffizienten findet man in jedem Buch über Kombinatorik. Wir beschränken uns hier auf die wichtigsten Berechnungsformeln.

$$\begin{aligned}\binom{n}{0} &= \binom{n}{n} = 1\\ \binom{n+1}{k} &= \binom{n}{k-1} + \binom{n}{k}\\ \binom{n}{k} &= \binom{n}{n-k}\end{aligned}$$

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n!}{k!(n-k)!}$$

mit $0! = 1$ und $m! = 1 \cdot 2 \cdot 3 \cdots (m-1) \cdot m$ für $m = 1, 2, \ldots$.

Beispiel 2.1 *Es werden n Kugeln ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge gezogen. Wie groß ist die Wahrscheinlichkeit, daß sich unter den gezogenen Kugeln k schwarze befinden?*

Dieses Ereignis wird durch die Teilmenge

$$\begin{aligned}B_k = \{ \quad & (\delta_1, \delta_2, \ldots, \delta_K, \delta_{K+1}, \ldots, \delta_N) \in \Omega_3;\\ & \textstyle\sum_{i=1}^{K} \delta_i = k,\\ & \textstyle\sum_{i=K+1}^{N} \delta_i = n-k\}\end{aligned}$$

von Ω_3 beschrieben; denn einer schwarzen Kugel entspricht eine Nummer $1 \leq \omega_i \leq K$ und dies wiederum einer Komponente $\delta_{\omega_i} = 1$. Da es $\binom{K}{k}$ Vektorteilstücke $(\delta_1, \ldots, \delta_K|$ mit $\sum_{i=1}^{K} \delta_i = k$ und $\binom{N-K}{n-k}$ Teilstücke $|\delta_{K+1}, \ldots, \delta_N)$ mit $\sum_{i=K+1}^{N} \delta_i = n-k$ gibt und die Vektoren aus B_k durch beliebige Kombinationen solcher Teilstücke entstehen, ist $|B_k| = \binom{K}{k}\binom{N-K}{n-k}$ und daher

$$P(B_k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

3 Bedingte Wahrscheinlichkeiten.

Definition 3.1 *Seien $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ zwei Ereignisse, wobei $P(B) > 0$. Dann heißt*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3.1}$$

die bedingte Wahrscheinlichkeit *des Ereignisses A unter der Bedingung B.*

Rechenregeln für bedingte Wahrscheinlichkeiten.

Bedingte Wahrscheinlichkeiten erscheinen meist in der Form, wie sie sich aus der Multiplikation der Gleichung (3.1) mit $P(B)$ ergibt:

$$P(A \cap B) = P(A|B)P(B). \tag{3.2}$$

Iterierte Anwendung von (3.2) führt zu

$$\begin{aligned}P(A \cap B \cap C) = & \quad P(A|(B \cap C))P(B \cap C)\\ = & \quad P(A|(B \cap C))P(B|C)P(C),\end{aligned} \tag{3.3}$$

was sich in offensichtlicher Weise auf Durchschnitte $A_1 \cap A_2 \cap \ldots \cap A_n$ beliebiger endlicher Länge n erweitern läßt.
Aus

$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A)$$

erhält man für den Fall, daß $P(A)$ und $P(B)$ beide positiv sind, die Beziehung

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{3.4}$$

Sei J entweder die endliche Indexmenge $\{1, 2, \ldots, n\}$ mit einer natürlichen Zahl n oder die Menge aller natürlichen Zahlen.

Definition 3.2 *Eine Menge $\{B_j\ ;\ j \in J\}$ von Ereignissen $B_j \in \mathcal{A}$ heißt eine* Partition *von Ω, wenn die B_j paarweise disjunkt sind und außerdem*

$$\sum_{j \in J} B_j \ = \ \Omega$$

gilt.

Ist $A \in \mathcal{A}$ ein beliebiges Ereignis und $\{B_j\ ;\ j \in J\}$ eine Partition von Ω, so ist

$$A = A \cap \Omega = A \cap \sum_{j \in J} B_j = \sum_{j \in J} A \cap B_j.$$

Anwendung von P liefert $P(A) = \sum_{j \in J} P(A \cap B_j)$ und zusammen mit (4.2) erhält man die *Formel von der totalen Wahrscheinlichkeit*

$$P(A) \ = \ \sum_{j \in J} P(A|B_j)P(B_j). \tag{3.5}$$

Die Kombination von (4.4) und (3.5) schließlich ergibt die *Formel von Bayes*:

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{j \in J} P(A|B_j)P(B_j)}. \tag{3.6}$$

Stochastische Unabhängigkeit von zwei Ereignissen.

Definition 3.3 *Zwei Ereignisse A, B in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit der Eigenschaft $P(A \cap B) = P(A)P(B)$ heißen* stochastisch unabhängig.

Ausser der Additionsregel $P(A+B) = P(A)+P(B)$ für disjunkte Ereignisse haben wir damit auch noch eine Produktregel. Aber **Achtung!** "Disjunkt" und "stochastisch unabhängig" sind komplementäre Begriffe. Zwei Ereignisse A und B können im allgemeinen nicht gleichzeitig disjunkt und stochastisch unabhängig sein; denn bei $P(A) > 0$ und $P(B) > 0$ folgt aus $A \cap B = \emptyset$, daß $P(A \cap B) = 0 \neq P(A)P(B)$ und damit stochastische Abhängigkeit.

Lemma 3.1 *A, B seien zwei Ereignisse.*

1. *Mit A, B sind auch die Paare $\overline{A}, B$, $A, \overline{B}$ und $\overline{A}, \overline{B}$ stochastisch unabhängig.*
2. *Ist $P(A) = 0$ oder $P(A) = 1$, so sind A und B für beliebige Ereignisse B stochastisch unabhängig.*

Beweis

1. Aus $B = AB + \overline{A}B$ und der Unabhängigkeit folgt $P(B) = P(A)P(B) + P(\overline{A}B)$ oder
$$P(\overline{A}B) = (1 - P(A))\, P(B) = P(\overline{A})P(B).$$
Die anderen beiden Aussagen erhält man in ähnlicher Weise.
2. Ist $P(A) = 0$, so ist wegen $AB \subset A$ auch $P(AB) = 0$ und es gilt dann stets $P(AB) = 0 = P(A)P(B)$. Ist $P(A) = 1$ so ist $P(\overline{A}) = 0$, somit $\overline{A}$ und B unabhängig. Der Rest folgt aus Punkt 1.

Globale stochastische Unabhängigkeit von Ereignissen.

Seien $A_1, A_2, \ldots, A_n$ endlich viele Ereignisse. Unter globaler stochastischer Unabhängigkeit versteht man die Eigenschaft, daß das Eintreten eines beliebigen Ereignisses A_k unter diesen n unabhängig vom Eintreten bzw. Nichteintreten aller übrigen ist. Um diese Eigenschaft exakt zu formulieren, definieren wir zunächst die stochastische Unabhängigkeit von σ-Algebren:

Definition 3.4 *n Unter-σ-Algebren $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ der σ-Algebra $\mathcal{A}$ heißen stochastisch unabhängig, wenn für beliebige Ereignisse $B_k \in \mathcal{A}_k$, $k = 1, \ldots, n$ gilt:*
$$P(B_1 \cap B_2 \cap \ldots \cap B_n) = P(B_1)P(B_2)\cdots P(B_n)$$

Für die obigen Ereignisse A_k kann man dann formulieren:

Definition 3.5 *Die Ereignisse $A_1, A_2, \ldots, A_n$ heißen* global stochastisch unabhängig, *wenn die σ-Algebren*
$$\mathcal{A}_k = \{\emptyset, A_k, \overline{A}_k, \Omega\}$$
stochastisch unabhängig sind.

4 Zufallsgrößen und ihre Verteilungen.

Sei $X : \tilde{\Omega} \longrightarrow \Omega$ eine Abbildung von einer nichtleeren Menge $\tilde{\Omega}$ in eine nichtleere Menge Ω. Für eine Teilmenge $A \subset \Omega$ heißt $(X \in A) := \{\tilde{\omega} \in \tilde{\Omega}\ ;\ X(\tilde{\omega}) \in A)\}$ das *Urbild* der Menge A unter der Abbildung X. Für die Zuordnung $A \longmapsto (X \in A)$ gelten die folgenden Regeln:

$$(X \in \Omega) = \tilde{\Omega} \tag{4.1}$$
$$(X \in \overline{A}) = \overline{(X \in A)} \tag{4.2}$$

und für beliebige Indexmengen J

$$(X \in \bigcup_{j\in J} A_j) = \bigcup_{j\in J}(X \in A_j) \tag{4.3}$$
$$(X \in \bigcap_{j\in J} A_j) = \bigcap_{j\in J}(X \in A_j). \tag{4.4}$$

Sind A und B disjunkte Teilmengen von Ω, so sind wegen

$$\begin{aligned}(X \in A) \cap (X \in B) &= (X \in A \cap B) = (X \in \emptyset)\\ &= (X \in \overline{\Omega}) = \overline{\tilde{\Omega}} = \emptyset\end{aligned}$$

auch $(X \in A)$ und $(X \in B)$ disjunkt, so daß

$$(X \in \sum_{j\in J} A_j) = \sum_{j\in J}(X \in A_j) \tag{4.5}$$

Seien $\tilde{\mathcal{A}}$ und $\mathcal{A}$ σ-Algebren auf den Mengen $\tilde{\Omega}$ bzw. Ω.

Definition 4.1 *Eine Abbildung $X : \tilde{\Omega} \longrightarrow \Omega$ mit der Eigenschaft $(X \in A) \in \tilde{\mathcal{A}}$ für alle $A \in \mathcal{A}$ heißt $\tilde{\mathcal{A}}$-$\mathcal{A}$-meßbar.*

Bezeichnung. Ist auf $\tilde{\mathcal{A}}$ noch zusätzlich eine Wahrscheinlichkeitsverteilung $\tilde{P}$ und damit insgesamt ein Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ gegeben, so spricht man in der Wahrscheinlichkeitsrechnung im allgemeinen nicht von einer meßbaren Abbildung, sondern von einer *Zufallsgröße*. Ist Ω eine Menge von Zahlen, so bezeichnet man X üblicherweise als eine *Zufallsvariable*.

Bemerkung. Handelt es sich bei $\tilde{\mathcal{A}}$ und $\mathcal{A}$ jeweils um die Menge aller Teilmengen der betreffenden Grundmenge, so ist jede Abbildung X $\tilde{\mathcal{A}}$-$\mathcal{A}$-meßbar.

Satz 4.1 *Ist $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{A}$ eine σ-Algebra auf einer Menge Ω und $X : \tilde{\Omega} \longrightarrow \Omega$ eine Zufallsgröße, so ist die Abbildung $P^X : \mathcal{A} \longrightarrow \mathbf{R}$, definiert durch*
$$P^X(A) := \tilde{P}(X \in A) \ \text{ für } \ A \in \mathcal{A}$$
eine Wahrscheinlichkeitsverteilung.

Definition 4.2 *Die in Satz 4.1 definierte Wahrscheinlichkeitsverteilung P^X heißt die* Verteilung *der Zufallsgröße X (bzgl. der Wahrscheinlichkeit $\tilde{P}$).*

Diesen Zusammenhang werden wir im folgenden stets durch ein Diagramm der Form

$$(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}) \xrightarrow{X} (\Omega, \mathcal{A}, P^X) \tag{4.6}$$

darstellen.

Beispiel: Beim Ziehen von n Kugeln im Beispiel 3.1 wäre für die Frage nach der Wahrscheinlichkeit für k schwarze Kugeln unter den gezogenen die naheliegende Ergebnismenge $\Omega = \{0, 1, 2, \ldots, n\}$ mit $\mathcal{A} = 2^\Omega$ als σ-Algebra. Zur Berechnung der Wahrscheinlichkeiten der Elementarereignisse $\{k\}$ betrachteten wir nach Durchnumerieren der Kugeln die Ergebnismenge

$$\tilde{\Omega} = \{(\delta_1, \ldots, \delta_N); \delta_i \in \{0,1\} \sum_{i=1}^{N} \delta_i = n\}$$

und darauf die Menge aller Teilmengen als Ereignisalgebra. Die Anzahl der schwarzen Kugeln unter den n gezogenen ist bei dieser Darstellung die Zufallsvariable $X : \tilde{\Omega} \longrightarrow \Omega$, gegeben durch $X(\delta_1, \ldots, \delta_N) = \sum_{i=1}^{K} \delta_i$. Die Wahrscheinlichkeit, k schwarze Kugeln zu ziehen, ist gleich der Wahrscheinlichkeit, daß X einen Wert in der Menge $\{k\}$ annimmt, also

$$P^X\{k\} = \tilde{P}(X \in \{k\}) = \tilde{P}(B_k).$$

Wenn man davon ausgeht, daß jedes Ergebnis aus $\tilde{\Omega}$ die gleiche Realisierungschance besitzt, $\tilde{P}$ also eine Laplace-Wahrscheinlichkeit ist, erhält man für die Verteilung von X wie im Abschnitt 2 hergeleitet,

$$P^X\{k\} = \frac{\begin{pmatrix} K \\ k \end{pmatrix}\begin{pmatrix} N-K \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}. \tag{4.7}$$

Komposition von Zufallsgrößen.

Seien $\Omega_1, \Omega_2, \Omega_3$ nichtleere Mengen und $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ auf diesen Mengen gegebene σ-Algebren, sowie P eine Wahrscheinlichkeitsverteilung auf $\mathcal{A}_1$. Für zwei Abbildungen $X : \Omega_1 \longrightarrow \Omega_2$ und $Y : \Omega_2 \longrightarrow \Omega_3$ nennen wir die Abbildung $Z : \Omega_1 \longrightarrow \Omega_3$, definiert durch $Z(\omega_1) := Y(X(\omega_1))$, die *Komposition* von X und Y und schreiben dafür $Z = Y \circ X$.

Falls alle drei Abbildungen meßbar bezüglich der jeweiligen σ-Algebren sind, können wir sie als Zufallsgrößen betrachten:

$$\begin{array}{rcl} (\Omega_1, \mathcal{A}_1, P) & \xrightarrow{X} & (\Omega_2, \mathcal{A}_2, P^X) \\ (\Omega_2, \mathcal{A}_2, P^X) & \xrightarrow{Y} & (\Omega_3, \mathcal{A}_3, P^Y) \\ (\Omega_1, \mathcal{A}_1, P) & \xrightarrow{Z} & (\Omega_3, \mathcal{A}_3, P^Z). \end{array}$$

Aus dem nachstehenden Diagramm, das den Zusammenhang zwischen den drei Zufallsgrößen verdeutlicht, ergibt sich zwangsläufig eine Vermutung, wie sie durch den nachstehenden Satz ausgesprochen wird.

$$\begin{tikzcd} (\Omega_1, \mathcal{A}_1, P) \arrow[r, "X"] \arrow[dr, "Z = Y \circ X"'] & (\Omega_2, \mathcal{A}_2, P^X) \arrow[d, "Y"] \\ & (\Omega_3, \mathcal{A}_3, P^{\diamond}) \end{tikzcd}$$

Satz 4.2 *Ist X $\mathcal{A}_1$-$\mathcal{A}_2$-meßbar, Y $\mathcal{A}_2$-$\mathcal{A}_3$-meßbar und wird Y als Zufallsgröße auf dem Wahrscheinlichkeitsraum $(\Omega_2, \mathcal{A}_2, P^X)$ betrachtet, so ist $Z = Y \circ X$ $\mathcal{A}_1$-$\mathcal{A}_3$-meßbar und es gilt $P^Z = P^Y$.*

Beweis. Zu zeigen ist, daß für beliebige Ereignisse $B \in \mathcal{A}_3$ das Urbild $(Z \in B)$ in der σ-Algebra $\mathcal{A}_1$ liegt und $P^Z(B) = P^Y(B)$ gilt.

Sei ein solches B gegeben und $A := (Y \in B)$. Wegen der Meßbarkeit von Y ist $A \in \mathcal{A}_2$ und für beliebige $\omega_1 \in \Omega_1$ gilt

$$\begin{array}{rcl} \omega_1 \in (Z \in B) & \Leftrightarrow & Z(\omega_1) = Y(X(\omega_1)) \in B \\ & \Leftrightarrow & X(\omega_1) \in A \\ & \Leftrightarrow & \omega_1 \in (X \in A), \end{array}$$

d.h. $(Z \in B) = (X \in A)$. Da X meßbar ist, erhält man daraus $(Z \in B) = (X \in A) \in \mathcal{A}_1$, also die erste Behauptung. Weiter gilt

$$\begin{array}{rcl} P^Z(B) & = & P(Z \in B) = P(X \in A) = P^X(A) \\ & = & P^X(Y \in B) = P^Y(B), \end{array}$$

also ist auch die zweite Aussage richtig.

Stochastische Unabhängigkeit

Seien $X_1, X_2, \ldots, X_n$ Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in Mengen Ω_k ; $k = 1, 2, \ldots, n$; auf denen die σ-Algebren $\mathcal{A}_k$ gegeben sind.

Definition 4.3 *Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ heißen* stochastisch unabhängig, *wenn für beliebige Ereignisse $B_k \in \mathcal{A}_k$; $k = 1, 2, \ldots, n$; gilt*

$$\begin{array}{l} P\left((X_1 \in B_1) \cap (X_2 \in B_2) \cap \ldots \cap (X_n \in B_n)\right) \\ \quad = P(X_1 \in B_1) P(X_2 \in B_2) \cdots P(X_n \in B_n) \end{array} \tag{4.8}$$

5 Diskrete Wahrscheinlichkeitsräume.

Definition 5.1
Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ heißt diskret, *falls die Ergebnismenge Ω abzählbar und $\{\omega\} \in \mathcal{A}$ für alle $\omega \in \Omega$ ist. P heißt in diesem Fall eine* diskrete Wahrscheinlichkeitsverteilung.

Aus dieser Definition folgt, daß in einem diskreten Wahrscheinlichkeitsraum die σ-Algebra $\mathcal{A}$ die Menge

2^Ω aller Teilmengen von Ω ist; denn jede beliebige Teilmenge $A \subset \Omega$ ist abzählbar und läßt sich daher als eine abzählbare Vereinigung

$$A = \sum_{\omega \in A} \{\omega\}$$

von Elementen aus $\mathcal{A}$ darstellen.

Definition 5.2 *Für eine diskrete Wahrscheinlichkeitsverteilung P heißt die Funktion $f_P : \Omega \longrightarrow \mathbf{R}$, definiert durch $f_P(\omega) = P\{\omega\}$, die* Wahrscheinlichkeitsfunktion *oder* Zähldichte *von P.*

Mit Hilfe der Wahrscheinlichkeitsfunktion kann man die Wahrscheinlichkeit jedes beliebigen Ereignisses $A \subset \Omega$ berechnen, denn es gilt aufgrund von Axiom (1.4)

$$P(A) = P(\sum_{\omega \in A} \omega) = \sum_{\omega \in A} P\{\omega\} = \sum_{\omega \in A} f_P(\omega).$$

Daraus folgt, daß eine diskrete Wahrscheinlichkeitsverteilung vollständig durch ihre Wahrscheinlichkeitsfunktion charakterisiert ist.
Auskunft darüber, welche Funktionen $f : \Omega \longrightarrow \mathbf{R}$ Wahrscheinlichkeitsfunktionen sind, gibt der

Satz 5.1 *Sei $f : \Omega \longrightarrow \mathbf{R}$ eine Funktion auf der abzählbaren Menge Ω mit den Eigenschaften*

$$\begin{aligned} f(\omega) &\geq 0 \text{ für alle } \omega \in \Omega \text{ und} \\ \sum_{\omega \in \Omega} f(\omega) &= 1. \end{aligned}$$

Dann gibt es genau eine diskrete Wahrscheinlichkeitsverteilung auf 2^Ω mit f als Wahrscheinlichkeitsfunktion. Umgekehrt besitzt natürlich jede Wahrscheinlichkeitsfunktion diese beiden Eigenschaften.

Beweis Ist f eine Funktion mit den obigen Eigenschaften, so definiert man $P : 2^\Omega \longrightarrow \mathbf{R}$ durch

$$P(\emptyset) := 0 \text{ und } P(A) := \sum_{\omega \in A} f(\omega)$$

und prüft leicht nach, daß alle Axiome einer Wahrscheinlichkeitsverteilung erfüllt sind.
Ist $\tilde{P}$ eine weitere diskrete Wahrscheinlichkeitsverteilung mit f als Wahrscheinlichkeitsfunktion, so gilt für alle $\omega \in \Omega$ nach Definition

$$\tilde{P}\{\omega\} = f(\omega) = P\{\omega\},$$

woraus sich durch entsprechende Summenbildung $\tilde{P}(A) = P(A)$ für beliebige $A \subset \Omega$ bzw. $\tilde{P} = P$ ergibt.

Beispiel 5.1 *Die Wahrscheinlichkeitsverteilung mit $\Omega = \{0, 1, \ldots, \min\{K, n\}\}$ und der Wahrscheinlichkeitsfunktion*

$$f(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

heißt die hypergeometrische Verteilung *oder kurz $\mathcal{H}(N, K, n)$-Verteilung.*

Beispiel 5.2 *Auf $\Omega = \{0, 1, 2, \ldots\} = \mathbf{N}_0$ erfüllt die Funktion*

$$f(k) = e^{-\mu}\frac{\mu^k}{k!}$$

mit einer positiven rellen Zahl μ die Voraussetzungen für eine Wahrscheinlichkeitsfunktion. Die zugehörige Verteilung heißt die Poisson-Verteilung *mit Parameter μ oder $\mathcal{P}(\mu)$-Verteilung.*

Die Poisson-Verteilung findet Anwendung, wenn die Häufigkeit des Auftretens eines bestimmten Ereignisses innerhalb eines Zeitintervalls gezählt wird, z.B. Anrufe bei einer Telefonzentrale, Kunden an einer Kasse, PKWs, die eine Meßstelle passieren, oder die Anzahl der Elektronen, die von der Kathode einer Vakuumröhre pro Zeiteinheit emittiert werden ("shot noise").

Beispiel 5.3 *Die Wahrscheinlichkeitsverteilung mit der Wahrscheinlichkeitsfunktion*

$$f(k) = p(1-p)^{k-1}$$

für $k \in \Omega = \{1, 2, \ldots\} = \mathbf{N}$ und $0 < p < 1$ heißt die geometrische Verteilung *mit Parameter p oder $\mathcal{G}(p)$-Verteilung.*

Die geometrische Verteilung findet Anwendung bei der Analyse von Wartezeiten bis zum Eintreffen eines bestimmten Ereignisses bzw. von Lebensdauern von Geräten. Sie ist dadurch charakterisiert, daß für die Ereignisse

$$A_k = \{k+1, k+2, \ldots\},$$

daß man mehr als k Zeiteinheiten warten muß, die Gleichungen

$$P(A_{k+l}|A_l) = P(A_k)$$

für alle $k, l \in \mathbf{N}$ gelten, d.h., daß die schon abgelaufene Wartezeit keinen Einfluß auf die Wahrscheinlichkeiten der nachfolgenden Wartezeiten hat.

Momente einer Verteilung

Die folgenden Konzepte sind nur auf diskrete Wahrscheinlichkeitsräume $(\mathcal{X}, 2^\mathcal{X}, P)$ mit einer abzählbaren Teilmenge $\mathcal{X} \subset \mathbf{R}$ von reellen Zahlen als Ergebnismenge anwendbar.
Sei $f(x) = P\{x\}$ die Wahrscheinlichkeitsfunktion von P.

Definition 5.3 *Gilt für eine natürliche Zahl k, daß $\sum_{x \in \mathcal{X}} |x|^k f(x) < \infty$, so heißt*

$$m_k(P) = \sum_{x \in \mathcal{X}} x^k f(x)$$

das k-te Moment *der Verteilung P. Andernfalls sagt man, daß die Verteilung P kein k-tes Moment besitzt.*

Bezeichnung: $m_1(P)$ heißt auch der *Mittelwert* der Verteilung P.

Beispiel 5.4 *Mittelwert der geometrischen Verteilung.*

$$\begin{aligned} m_1(P) &= \sum_{n=0}^{\infty} npq^{n-1} = p\sum_{n=1}^{\infty} nq^{n-1} = \\ &= p\sum_{n=1}^{\infty} \frac{d}{dq}q^n = p\sum_{n=0}^{\infty} \frac{d}{dq}q^n = \\ &= p\frac{d}{dq}\sum_{n=0}^{\infty} q^n = p\frac{d}{dq}\frac{1}{1-q} = \\ &= p\frac{1}{(1-q)^2} = p\frac{1}{p^2} = \frac{1}{p} \end{aligned}$$

Liegt die Ergebnismenge $\mathcal{X}$ ganz im Bereich der nichtnegativen reellen Zahlen, so lassen sich die Momente der Verteilung P auch mittels der *momenterzeugenden Funktion*

$$M(t) = \sum_{x\in\mathcal{X}} e^{tx} f(x),$$

die auf jeden Fall für $t < 0$ wohldefiniert ist, berechnen. Es gilt für diese t nämlich

$$\frac{d^k}{dt^k}M(t) = \sum_{x\in\mathcal{X}} \frac{\partial^k}{\partial t^k}e^{tx} f(x) = \sum_{x\in\mathcal{X}} x^k e^{tx} f(x)$$

und daher

$$\lim_{t\nearrow 0} \frac{d^k}{dt^k}M(t) = \sum_{x\in\mathcal{X}} x^k \lim_{t\nearrow 0} e^{tx} f(x) = \sum_{x\in\mathcal{X}} x^k f(x) = m_k(P),$$

falls das k-te Moment existiert.

Beispiel 5.5 *Die Momente der Poisson-Verteilung*

$$\begin{aligned} M(t) &= \sum_{n=0}^{\infty} e^{tn} e^{-\mu}\frac{\mu^n}{n!} = e^{-\mu}\sum_{n=0}^{\infty} \frac{(e^t\mu)^n}{n!} = \\ &= e^{-\mu}e^{e^t\mu} = e^{\mu(e^t-1)}. \end{aligned}$$

Die ersten beiden Ableitungen sind $M^{(1)}(t) = M(t)\mu e^t$ und $M^{(2)}(t) = \mu(M^{(1)}(t)e^t + M(t)e^t)$, woraus $M(0) = 1$, $M^{(1)}(0) = m_1(P) = \mu$ und $M^{(2)}(0) = m_2(P) = \mu^2 + \mu$ folgt.

Definition 5.4 *Gilt für eine natürliche Zahl k, daß $\sum_{x\in\mathcal{X}} |x - m_1(P)|^k f(x) < \infty$, so heißt*

$$\hat{m}_k(P) = \sum_{x\in\mathcal{X}} (x - m_1(P))^k f(x)$$

das k-te zentrale Moment *der Verteilung P. Andernfalls sagt man, daß die Verteilung P kein k-tes zentrales Moment besitzt.*

Bezeichnung: $\hat{m}_2(P)$ heißt auch die *Varianz* der Verteilung P.

Satz 5.2 $\hat{m}_2(P) = m_2(P) - (m_1(P))^2$

Beweis

$$\begin{aligned} \hat{m}_2(P) &= \sum_{x\in\mathcal{X}} \left(x^2 - 2xm_1(P) + m_1(P)^2\right) f(x) \\ &= \sum_{x\in\mathcal{X}} x^2 f(x) \; - \; 2m_1(P)\sum_{x\in\mathcal{X}} xf(x) \\ &\quad + m_1(P)^2 \sum_{x\in\mathcal{X}} f(x) \\ &= m_2(P) \; - \; 2m_1(P)m_1(P) \; + \; m_1(P)^2 \cdot 1 \\ &= m_2(P) - (m_1(P))^2 \end{aligned}$$

Sei $\varepsilon > 0$ und $B_\varepsilon = \{x \in \mathcal{X} \; ; \; |x - m_1(P)| > \varepsilon\}$ das Ereignis, daß das Zufallsexperiment ein Ergebnis x mit einem Abstand von mindestens ε vom Mittelwert liefert. Dann gilt

Satz 5.3 *(*Ungleichung von Tschebischeff*)*

$$P(B_\varepsilon) \leq \frac{\hat{m}_2(P)}{\varepsilon^2}$$

Beweis

$$\begin{aligned} \hat{m}_2(P) &= \sum_{x\in B_\varepsilon} (x - m_1(P))^2 f(x) \; + \\ &\quad + \sum_{x\notin B_\varepsilon} (x - m_1(P))^2 f(x) \\ &\geq \sum_{x\in B_\varepsilon} \varepsilon^2 f(x) \; + 0 \; = \; \varepsilon^2 \sum_{x\in B_\varepsilon} f(x) \; = \; \varepsilon^2 P(B_\varepsilon) \end{aligned}$$

Die obige Ungleichung erhält man dann durch Division mit ε^2.

Der Erwartungswert einer Zufallsvariablen

Sei $(\Omega, 2^\Omega, P)$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \longrightarrow \mathbf{R}$ eine reellwertige Funktion mit der abzählbaren Bildmenge $\mathcal{X} = \{x \in \mathbf{R} \; ; \; \text{es gibt ein } \omega \text{ mit } X(\omega) = x\}$. Bezüglich dieses Wertebereichs ist X eine Zufallsvariable:

$$(\Omega, 2^\Omega, P) \xrightarrow{X} (\mathcal{X}, 2^{\mathcal{X}}, P^X)$$

Definition 5.5 *Falls $\sum_{\omega\in\Omega} |X(\omega)| P\{\omega\} < \infty$, heißt*

$$\mathcal{E}_P X \; = \; \sum_{\omega\in\Omega} X(\omega) P\{\omega\}$$

der Erwartungswert *der Zufallsvariablen X bez. der Wahrscheinlichkeitsverteilung P. Andernfalls sagt man, daß X keinen Erwartungswert besitzt.*

Der Index P wird meist weggelassen: $\mathcal{E}X = \mathcal{E}_P X$.

Definition 5.6 *Falls X einen Erwartungswert besitzt und die nachstehende Reihe konvergiert, heißt*

$$\mathrm{var}_P X \; = \; \sum_{\omega\in\Omega} (X(\omega) - \mathcal{E}_P X)^2 P\{\omega\}$$

die Varianz *der Zufallsvariablen X bez. der Wahrscheinlichkeitsverteilung P.*

Der Index P wird meist weggelassen: $\mathrm{var}\, X = \mathrm{var}_P X$.

Satz 5.4 *Der Erwartungswert (die Varianz) der Zufallsvariablen X existiert genau dann wenn der Mittelwert (die Varianz) der Verteilung P^X der Zufallsvariablen X existiert und es gilt $\mathcal{E}_P X = m_1(P^X)$ bzw. $\mathrm{var}_P X = \hat{m}_2(P^X)$.*

Beweis Für $x \in \mathcal{X}$ sei $A_x = (X = x) = \{\omega \in \Omega\ ;\ X(\omega) = x\}$. Dann ist $\{A_x\ ;\ x \in \mathcal{X}\}$ eine Partition von Ω. Ist weiterhin f^X die Wahrscheinlichkeitsfunktion der Verteilung von X, so gilt gemäß der Definitionen

$$P(A_x) = P(X = x) = P^X\{x\} = f^X(x).$$

Damit erhält man

$$\begin{aligned}
\mathcal{E}_P X &= \sum_{\omega\in\Omega} X(\omega)P\{\omega\} = \sum_{x\in\mathcal{X}}\sum_{\omega\in A_x} X(\omega)P\{\omega\} \\
&= \sum_{x\in\mathcal{X}}\sum_{\omega\in A_x} xP\{\omega\} = \sum_{x\in\mathcal{X}} x \sum_{\omega\in A_x} P\{\omega\} \\
&= \sum_{x\in\mathcal{X}} xP(A_x) = \sum_{x\in\mathcal{X}} xf^X(x) \\
&= m_1(P^X)
\end{aligned}$$

Der Beweis für die Varianzen verläuft analog.
Die Tschebischeff'sche Ungleichung

$$P^X(B_\varepsilon) \le \frac{\hat{m}_2(P^X)}{\varepsilon^2}$$

für die Verteilung der Zufallsvariablen X läßt sich mittels dieses Satzes auch noch etwas anschaulicher formulieren: Wegen $P^X(B_\varepsilon) = P(X \in B_\varepsilon)$ und $(X \in B_\varepsilon) = \{\omega \in \Omega\ ;\ |X(\omega) - m_1(P^X)| > \varepsilon\} = \{\omega \in \Omega\ ;\ |X(\omega) - \mathcal{E}X| > \varepsilon\} = (|X - \mathcal{E}X| > \varepsilon)$ lautet sie

$$P(|X - \mathcal{E}X| > \varepsilon) \le \frac{\mathrm{var}\, X}{\varepsilon^2}$$

6 Borelsche Mengen und Lebesguesches Maß

Hinsichtlich der Bezeichnungen in diesem und den folgenden Abschnitten beziehen wir uns auf den Abschnitt 3 "Integralrechnung" des Skripts zur Vorlesung Ingenieurmathematik III für Informatiker, WS 1989/90.
$\mathbf{R}^n = \{x = (x_1, \ldots, x_n)\ ;\ x_i \in \mathbf{R}\}$ sei die Menge der n-tupel reeller Zahlen und $\mathcal{I}_n$ die Menge der *Intervalle* im $\mathbf{R}^n$. Für das Maß oder den Inhalt $|I|$ eines Intervalls I lassen wir bei unbeschränkten Intervallen auch den Wert $|I| = \infty$ zu.

Definition 6.1 *Die kleinste σ-Algebra über dem $\mathbf{R}^n$, die die Menge $\mathcal{I}_n$ aller Intervalle enthält, heißt die σ-Algebra der* borelschen Mengen *des $\mathbf{R}^n$ und wird mit $\mathcal{B}_n$ bezeichnet.*

Für $\mathcal{B}_1$ schreibt man auch kurz $\mathcal{B}$.

Satz 6.1 *Es gibt genau eine Abbildung $\lambda : \mathcal{B}_n \longrightarrow \mathbf{R} \cup \{\infty\}$ mit den Eigenschaften*

$$\begin{aligned}
\lambda(B) &\ge 0 \\
\lambda(\sum_k B_k) &= \sum_k \lambda(B_k) \\
\lambda(I) &= |I| \quad \textit{für Intervalle } I
\end{aligned}$$

Definition 6.2 *Die Abbildung λ aus Satz 6.1 heißt das* Lebesguesche Maß *auf dem $\mathbf{R}^n$*

Den Zusammenhang zwischen dem Lebesgueschen Maß und dem Lebesgueschen Integral erhält man durch die *Indikatorfunktion* 1_A einer Teilmenge $A \subset \mathbf{R}^n$:

$$1_A(x) = \begin{cases} 1 & \text{für } x \in A \\ 0 & \text{sonst} \end{cases}$$

Satz 6.2 *Für beschränkte Teilmengen $B \subset \mathbf{R}^n$ gilt: B ist borelsche Menge genau dann, wenn die Indikatorfunktion 1_B integrierbar ist, wobei*

$$\lambda(B) = \int 1_B(x)\, dx = \int_B 1\, dx.$$

Zu jeder unbeschränkten borelschen Menge B gibt es eine monotone Folge $\{B_i\ ;\ i = 1, 2, \ldots\}$ von beschränkten borelschen Mengen mit $B_i \nearrow B$.

7 Verteilungen.

Wir beginnen in diesem Abschnitt mit der Betrachtung von Wahrscheinlichkeiten auf der Menge $\mathbf{R}$ der rellen Zahlen bzw. der Menge $\mathbf{R}^n$ der n-dimensionalen Vektoren $x = (x_1, x_2, \ldots, x_n)$ mit rellen Komponenten x_k als Ergebnismenge.

Definition 7.1 *Eine Wahrscheinlichkeitsverteilung $P : \mathcal{B}_n \longrightarrow \mathbf{R}$ auf den Borel'schen Mengen des $\mathbf{R}^n$ heißt eine* Verteilung auf dem $\mathbf{R}^n$.

Beispiel 7.1 *Sei $M \in \mathcal{B}_n$ eine Menge mit $0 < \lambda(M) < \infty$. Wie man mit den üblichen Rechenregeln und Konvergenzsätzen für das Lebesgue'sche Integral leicht überprüft, ist*

$$P(B) := \frac{\lambda(B \cap M)}{\lambda(M)}$$

eine Verteilung auf dem $\mathbf{R}^n$. Sie ist das kontinuierliche Analogon zu einem Laplace-Experiment, denn sie ist dadurch charakterisiert, daß bei der Durchführung des entsprechenden Zufallsexperiments jedes Ergebnis x aus der Menge M "gleich möglich" ist. P heißt die uniforme *oder* Gleich-Verteilung *auf der Menge M, abgekürzt $\mathcal{U}(M)$-Verteilung.*

Spezielle Typen von Verteilungen.

1. Diskrete Verteilungen.
Sei $\Omega \subset \mathbf{R}^n$ eine abzählbare Menge und $P : 2^\Omega \longrightarrow \mathbf{R}$ eine diskrete Wahrscheinlichkeitsverteilung mit Wahrscheinlichkeitsfunktion $f : \Omega \longrightarrow \mathbf{R}$. Eine solche Wahrscheinlichkeitsverteilung läßt sich als Verteilung $\tilde{P}$ auf dem $\mathbf{R}^n$ auffassen, wenn man
$$\tilde{P}(B) := P(B \cap \Omega) = \sum_{\omega \in B \cap \Omega} f(\omega)$$
setzt. Ein derartiges $\tilde{P}$ heißt *diskrete Verteilung mit Trägermenge* Ω.

2. Absolutstetige Verteilungen.
Im Zusammenhang mit Lebesgue'schen Integralen ist die folgende Sprechweise üblich: Eine Aussage über eine auf dem $\mathbf{R}^n$ definierte Funktion f gilt "fast überall", wenn es eine Menge N mit $\lambda(N) = 0$ – d.h. eine "Nullmenge"– gibt, so daß die Aussage für alle Argumente $x \in \mathbf{R}^n$ mit Ausnahme der x aus der Menge N richtig ist.

Definition 7.2 *Eine Lebesgue-integrierbare Funktion* $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ *mit*
$$\begin{aligned} f(x) &\geq 0 \ \textit{fast überall und} \\ \int f(x)\,dx &= 1 \end{aligned}$$
heißt eine Dichte, Dichtefunktion *oder* Wahrscheinlichkeitsdichte.

Derartige Funktionen liefern Verteilungen auf dem $\mathbf{R}^n$:

Satz 7.1 *Sei f eine Dichte. Dann ist die durch*
$$P(B) = \int 1_B(x) f(x)\,dx$$
definierte Funktion P auf $\mathcal{B}_n$ eine Verteilung auf dem $\mathbf{R}^n$.

Beweis: Die Kolmogoroff'schen Axiome folgen für dieses P aus den üblichen Rechenregeln und Konvergenzsätzen für das Lebesgue'sche Integral.
Bezeichnung: Eine Verteilung dieser Form heißt eine *absolutstetige* Verteilung mit Dichte f. Im Gegensatz zu diskreten Wahrscheinlichkeitsverteilungen und den zugehörigen Wahrscheinlichkeitsfunktionen ist eine Dichte durch die Verteilung nicht eindeutig bestimmt. Je zwei Dichten für die gleiche absolutstetige Verteilung unterscheiden sich jedoch höchstens auf einer Nullmenge.

Beispiel 7.2 *Die Funktion*
$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \quad \textit{für } t \in \mathbf{R}$$
ist eine Dichte. Die zugehörige Verteilung heißt die Gauss'sche Einheitsverteilung *oder $\mathcal{N}(0,1)$-Verteilung.*

Beispiel 7.3 *Die uniforme Verteilung auf der Menge M besitzt die Dichte*
$$f(x) = \frac{1}{\lambda(M)} 1_M(x)$$

8 Eindimensionale Verteilungen.

In diesem Abschnitt befassen wir uns mit der Charakterisierung von eindimensionalen Verteilungen, also Wahrscheinlichkeitsverteilungen $P : \mathcal{B} \longrightarrow \mathbf{R}$ auf der σ-Algebra $\mathcal{B}$ der Borelschen Teilmengen der rellen Zahlenachse.

Die Verteilungsfunktion.

Definition 8.1 *Die Funktion $F : \mathbf{R} \longrightarrow \mathbf{R}$, definiert durch*
$$F(t) := P(-\infty, t]$$
für alle $t \in \mathbf{R}$, heißt die Verteilungsfunktion *der Verteilung P.*

Beispiel 8.1 *Verteilungsfunktion der $\mathcal{U}[a,b]$ - Verteilung.*

Es ist
$$P(-\infty, t] = \frac{\lambda\left((-\infty, t] \cap [a,b]\right)}{\lambda\left([a,b]\right)}$$
mit $\lambda\left([a,b]\right) = b - a$. Wegen
$$(-\infty, t] \cap [a,b] = \begin{cases} \emptyset & \text{falls } t < a \\ [a,t] & \text{falls } a \leq t \leq b \\ [a,b] & \text{falls } t > b \end{cases}$$
ist
$$F(t) = \begin{cases} 0 & \text{falls } t < a \\ \frac{t-a}{b-a} & \text{falls } a \leq t \leq b \\ 1 & \text{falls } t > b \end{cases}$$

Satz 8.1 *Eine Verteilungsfunktion besitzt die folgenden 5 Eigenschaften:*
$$0 \leq F(t) \leq 1$$
$$\begin{array}{lll} s \leq t & \Rightarrow & F(s) \leq F(t) \ (F \textit{ monoton nichtfallend}) \\ t \nearrow \infty & \Rightarrow & F(t) \nearrow 1 \\ t \searrow -\infty & \Rightarrow & F(t) \searrow 0 \\ t \searrow t_0 & \Rightarrow & F(t) \searrow F(t_0) \ (F \textit{ ist rechtsstetig}) \end{array}$$

Diese fünf Eigenschaften charakterisieren eine Verteilungsfunktion bereits vollständig, denn es gilt der

Satz 8.2 *Zu jeder Funktion $F : \mathbf{R} \longrightarrow \mathbf{R}$ mit den in Satz 8.1 aufgeführten Eigenschaften gibt es genau eine Verteilung P auf $\mathbf{R}$ mit $P(-\infty, t] = F(t)$ für alle $t \in \mathbf{R}$.*

Beispiel 8.2 *Die Funktion*

$$F(t) = \begin{cases} 0 & \text{falls } t \leq 0 \\ 1 - e^{-\lambda t} & \text{falls } t > 0 \end{cases}$$

mit einer Konstanten $\lambda > 0$ erfüllt alle Eigenschaften einer Verteilungsfunktion. Die zugehörige Verteilung heißt die Exponentialverteilung mit Parameter λ *oder kurz die $\mathcal{E}(\lambda)$-Verteilung.*

Um einen der Anwendungsbereiche der Exponentialverteilung darzustellen, betrachten wir für $s > 0$ die Ereignisse $A_s := (s, \infty)$. Es ist

$$P(A_s) = 1 - P(\overline{A_s}) = 1 - P(-\infty, s] = 1 - F(s) = e^{-\lambda s}$$

woraus sich die Gleichung

$$P(A_{s+t}) = P(A_s)P(A_t)$$

für alle positiven rellen Zahlen s, t ergibt. Aus dieser Gleichung wiederum folgt – auf dem umgekehrten Weg wie bei der Herleitung der geometrischen Verteilung – die Identität

$$P(A_{s+t}|A_s) = P(A_t),$$

d.h., die Exponentialverteilung ist das kontinuierliche Gegenstück zur geometrischen Verteilung. Man benutzt sie beispielsweise zur Beschreibung der Lebensdauer nahezu verschleißfreier Geräte, wenn die Zeit nicht in diskreten Beobachtungsintervallen, sondern kontinuierlich gemessen wird.

Die Momente einer absolutstetigen Verteilung.

Definition 8.2 *Sei P eine Verteilung auf* $\mathbf{R}$ *mit Dichte f und k eine natürliche Zahl. Falls der Absolutbetrag der Funktion $t \longrightarrow t^k f(t)$ integrierbar ist, heißt das Integral*

$$m_k(P) \ = \ \int t^k f(t)dt$$

das k-te (absolute) Moment *der Verteilung P. Falls $t \longrightarrow |(t - m_1(P))^k f(t)|$ integrierbar ist, heißt*

$$\hat{m}_k(P) \ = \ \int (t - m_1(P))^k f(t)dt$$

das k-te zentrale Moment *der Verteilung P.*

Speziell $m_1(P)$ heißt der *Mittelwert* und $\hat{m}_2(P)$ die *Varianz* der Verteilung P. Interpretiert man physikalisch eine Dichte f als spezifische Dichte einer Belegung der reellen Zahlenachse mit "Wahrscheinlichkeitsmasse", so entspricht der Mittelwert dem Schwerpunkt und die Varianz dem Trägheitsmoment bezüglich einer Achse durch den Schwerpunkt.

Lemma 8.1 $\hat{m}_2(P) = m_2(P) - (m_1(P))^2$.

Beweis: Wie im Fall diskreter Verteilungen, wenn man dort die Summenzeichen durch Integrale ersetzt.

Beispiel 8.3 *Mittelwert und Varianz der $\mathcal{N}(0,1)$-Verteilung.*

Die Dichte $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$ der $\mathcal{N}(0,1)$-Verteilung ist eine gerade Funktion: $\phi(-t) = \phi(t)$; daher ist $t \longrightarrow t \cdot \phi(t)$ ungerade: $(-t) \cdot \phi(-t) = -(t \cdot \phi(t))$ und

$$m_1(P) = \int_{-\infty}^{\infty} t \cdot \phi(t)dt \ = \ 0.$$

Daraus folgt

$$\hat{m}_2(P) = m_2(P) = \int_{-\infty}^{\infty} t^2 \phi(t)dt.$$

Mittels partieller Integration erhält man

$$\begin{aligned} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt &= \int_{-\infty}^{\infty} t \left(t e^{-\frac{1}{2}t^2} \right) dt \\ &= \left[t \left(-e^{-\frac{1}{2}t^2} \right) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-\frac{1}{2}t^2} dt \\ &= 0 + \sqrt{2\pi} \end{aligned}$$

und damit

$$\hat{m}_2(P) = 1.$$

Zur Berechnung der Momente einer Verteilung P, deren Dichte $f(x) = 0$ für fast alle $x < 0$ — einer "nichtnegativen Verteilung" — gibt es das Hilfsmittel der *momenterzeugenden Funktion.*

Definition 8.3 *Sei P eine nichtnegative Verteilung auf* $\mathbf{R}$ *mit Dichte f. Dann heißt die für $t < 0$ definierte Funktion*

$$M(t) \ = \ \int e^{tx} f(x)dx$$

die momenterzeugende Funktion *der Verteilung P.*

Schreibt man $-s$ für t, so sieht man, daß es sich dabei im Fall einer absolutstetigen Verteilung um die *Laplace - Transformation* der Dichte handelt. Die momenterzeugende Funktion besitzt die folgenden Eigenschaften:

- $M(t)$ ist für alle $t < 0$ beliebig oft differenzierbar und es gilt

$$\begin{aligned} M^{(k)}(t) := & \frac{d^k}{dt^k} M(t) = \frac{d^k}{dt^k} \int_0^{\infty} e^{tx} f(x)dx \\ = & \int_0^{\infty} \frac{\partial^k}{\partial t^k} e^{tx} f(x)dx = \int_0^{\infty} x^k e^{tx} f(x)dx. \end{aligned}$$

- $M^{(k)}(t)$ ist in t monoton steigend und falls $\lim_{t \nearrow 0} M^{(k)}(t) < \infty$, dann gilt

$$\begin{aligned} \lim_{t \nearrow 0} M^{(k)}(t) = & \lim_{t \nearrow 0} \int_0^{\infty} x^k e^{tx} f(x)dx \\ = & \int_0^{\infty} x^k \lim_{t \nearrow 0} e^{tx} f(x)dx \\ = & \int_0^{\infty} x^k f(x)dx = m_k(P). \end{aligned}$$

Falls also die k-te Ableitung der momenterzeugenden Funktion an der Stelle $t = 0$ existiert, ist dieser Wert gleich dem k-ten Moment der Verteilung.

Beispiel 8.4 *Die Exponentialverteilung mit Parameter λ besitzt die Verteilungsfunktion $F(x) = (1 - e^{-\lambda x}) \cdot 1_{(0,\infty)}(x)$ und die Dichte $f(x) = \frac{d}{dx}F(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x)$. Ihre momenterzeugende Funktion ist*

$$M(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda \int_0^\infty e^{(t-\lambda)x}dx = \frac{\lambda}{\lambda - t}.$$

Wegen $M^{(1)}(t) = \frac{\lambda}{(\lambda-t)^2}$ und $M^{(2)}(t) = \frac{2\lambda}{(\lambda-t)^3}$ ist $m_1(P) = M^{(1)}(0) = \frac{1}{\lambda}$ und $m_2(P) = M^{(2)}(0) = \frac{2}{\lambda^2}$. Für die Varianz erhält man nach dem Lemma 8.1 schließlich den Wert $\hat{m}_2(P) = \frac{1}{\lambda^2}$.

Die Ungleichung von Tschebyscheff. Wie für diskrete Verteilungen beweist man:

Satz 8.3 (Ungleichung von Tschebyscheff.) *Für eine positive Zahl ε sei*

$$B_\varepsilon = \{x \in \mathbf{R} \; ; \; |x - m_1(P)| > \varepsilon\}.$$

Dann ist

$$P(B_\varepsilon) \le \frac{\hat{m}_2(P)}{\varepsilon^2}.$$

9 Verteilungen auf dem R^n.

In diesem Abschnitt befassen wir uns mit Wahrscheinlichkeitsverteilungen und Zufallsgrößen auf der Ergebnismenge $\mathbf{R}^n$ der Vektoren

$$x = (x_1, x_2, \ldots, x_n)$$

mit rellen Komponenten x_k. Als σ-Algebra wird dabei stets die Menge $\mathcal{B}_n$ der Borelschen Teilmengen des $\mathbf{R}^n$ verwendet. Bei den Verteilungen beschränken wir uns auf diejenigen, die durch eine Dichte charakterisiert sind:

$$P(B) = \int 1_B(x)f(x)dx.$$

Sei $(\Omega, \mathcal{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \longrightarrow \mathbf{R}^n$ eine Abbildung.

Definition 9.1 *Die Abbildung X heißt $\mathcal{A}$-$\mathcal{B}_n$-*meßbar *oder ein* Zufallsvektor*, wenn für alle Borelschen Teilmengen B des $\mathbf{R}^n$ gilt*

$$(X \in B) \in \mathcal{A}.$$

Da $X(\omega) \in \mathbf{R}^n$, läßt sich dieser Vektor auch in der Form

$$X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$$

schreiben. Die so gegebenen Funktionen $X_k : \Omega \longrightarrow \mathbf{R}$ bezeichnen wir als die *Komponenten* des Zufallsvektors X und schreiben für X auch $X = (X_1, X_2, \ldots, X_n)$.

Satz 9.1 *$X = (X_1, X_2, \ldots, X_n)$ ist Zufallsvektor genau dann, wenn alle seine Komponenten Zufallsvariable sind.*

Wir können also stets eine beliebige Anzahl n von Zufallsvariablen $Y_1, Y_2, \ldots Y_n$ zu einer vektorwertigen Funktion

$$Y = (Y_1, Y_2, \ldots, Y_n)$$

zusammenfassen und als Zufallsvektor behandeln.

n Zufallsvariable $X_1, X_2, \ldots, X_n$ seien gegeben und die Dichte f der Verteilung P^X des Zufallsvektors $X = (X_1, X_2, \ldots, X_n)$ bekannt. Weiter seien gegeben m Funktionen $G_j : \mathbf{R}^n \longrightarrow \mathbf{R} \; ; \; j = 1, \ldots, m$ mit denen wir die Funktionen

$$\begin{aligned} Y_1(\omega) &= G_1(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \\ &\vdots \\ Y_m(\omega) &= G_m(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \end{aligned}$$

bilden. Ziel ist die Berechnung der Verteilung P^Y des Zufallsvektors $Y = (Y_1, Y_2, \ldots, Y_m)$ aus der Verteilung P^X.

Wenn wir $G : \mathbf{R}^n \longrightarrow \mathbf{R}^m$ durch $G(x) := (G_1(x), \ldots, G_m(x))$ definieren, so ist mit den Bezeichnungen aus Abschnitt 4 der Zufallsvektor $Y = G \circ X$ die Komposition von X und G und wir können den Satz 4.2 anwenden: Falls G ein Zufallsvektor, ist auch Y einer und beide besitzen die gleiche Verteilung. Wir brauchen uns daher nur mit der Situation

$$(\mathbf{R}^n, \mathcal{B}_n, P^X) \xrightarrow{G} (\mathbf{R}^m, \mathcal{B}_m, P^G)$$

zu befassen.

Marginalverteilungen.

Definition 9.2 *Die Verteilung P^{X_k} der k-ten Komponente eines Zufallsvektors $X = (X_1, X_2, \ldots, X_n)$ heißt die k-te* Marginalverteilung *oder k-te Randverteilung von P^X.*

Satz 9.2 *Die k-te Marginalverteilung besitzt als Dichte die k-*te Marginaldichte

$$f_k(x_k) = \underbrace{\int_{\mathbf{R}} \cdots \int_{\mathbf{R}}}_{(n-1)\text{-}mal} f(x_1, x_2, \ldots, x_n)dx_1 \ldots dx_{k-1}dx_{k+1} \ldots dx_n.$$

Satz 9.3 *Sei $X = (X_1, X_2, \ldots, X_n)$ ein Zufallsvektor, dessen Verteilung P^X die Dichte f besitzt. Dann gilt: Die Komponenten $X_1, X_2, \ldots, X_n$ sind genau dann stochastisch unabhängig, wenn es Dichten $f_k : \mathbf{R} \longrightarrow \mathbf{R}$ gibt, so daß für fast alle Vektoren $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ gilt*

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n).$$

Die f_k sind dabei die Marginaldichten von f.

Der Transformationssatz für Dichten.

Satz 9.4 *Sei P eine absolutstetige Verteilung auf dem $\mathbf{R}^n$ mit der Dichte f und $G : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ eine $\mathcal{B}_n$ - $\mathcal{B}_n$ - meßbare Abbildung mit den folgenden Eigenschaften:*

- *Es gibt eine Zerlegung $\mathbf{R}^n = N+M_1+\ldots+M_K$ des $\mathbf{R}^n$ in offene, zusammenhängende Mengen M_k und eine Restmenge N mit $P(N) = \int 1_N(x)f(x)dx = 0$, so daß für die Funktion G folgendes gilt:*

- *Auf jeder der Mengen M_k sind die Komponentenfunktionen der Abbildung G stetig partiell differenzierbar und die Funktionaldeterminante $J_G(x)$ ist für alle $x \in M_k$ von Null verschieden.*

- *Für jedes k besitzt G auf der Bildmenge $M_k^* = G(M_k)$ von M_k eine Umkehrfunktion $G_k^* : M_k^* \longrightarrow M_k$.*

Dann ist die Verteilung P^G der Zufallsgröße G bezüglich des Wahrscheinlichkeitsraums $(\mathbf{R}^n, \mathcal{B}_n, P)$ absolutstetig und besitzt die Dichte

$$g(y) \;=\; \sum_{k=1}^{K} f(G_k^*(y)) \frac{1}{|J_G(G_k^*(y))|} 1_{M_k^*}(y).$$

Beispiel 9.1 *X_1 und X_2 seien stochastisch unabhängige und im Intervall $(0,1)$ uniform verteilte Zufallsvariable. Welche Verteilung besitzt dann die Zufallsvariable $Y = X_1 X_2$?*

Es ist $Y = G_1(X_1, X_2)$ mit $G_1(x_1, x_2) = x_1 x_2$. Wie eingangs dieses Abschnitts ausgeführt, ist die Verteilung P^Y der Zufallsvariablen Y gleich der Verteilung P^{G_1}, wenn man die Funktion G_1 als Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\mathbf{R}^2, \mathcal{B}_2, P^X)$ mit der Verteilung P^X des Zufallsvektors $X = (X_1, X_2)$ betrachtet. Um den Transformationssatz für Dichten anwenden zu können, muß zunächst die Funktion G_1 durch eine weitere Funktion $G_2 : \mathbf{R}^2 \longrightarrow \mathbf{R}$ zu einer Abbildung $G = (G_1, G_2)$ des $\mathbf{R}^2$ in sich ergänzt werden, die alle Voraussetzungen dieses Satzes erfüllt. Man erhält dann die Dichte $g(y_1, y_2)$ der Verteilung P^G. Die gesuchte Verteilung P^{G_1} ist gemäß dieser Konstruktion gerade die erste Marginalverteilung von P^G. Ihre Dichte erhält man durch das Auswerten des Integrals

$$g_1(y_1) = \int g(y_1, y_2) dy_2.$$

Wegen der Unabhängigkeit der beiden Variablen X_1 und X_2 besitzt P^X die Dichte $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, wobei die Marginaldichten $f_i(t) = 1_{(0,1)}(t)$ jeweils die Dichten der $\mathcal{U}(0,1)$-Verteilung sind. Mit

$$M_1 \;=\; \{x = (x_1, x_2) \in \mathbf{R}^2 \;;\; 0 < x_1 < 1 \;,\; 0 < x_2 < 1\,\}$$

ist also

$$f(x) \;=\; \begin{cases} 1 & x \in M_1 \\ 0 & \text{sonst} \end{cases}$$

Die Restmenge $N = \overline{M_1}$ besitzt die Wahrscheinlichkeit $P^X(N) = 0$, denn auf ihr ist die Dichte f identisch Null.

Wir ergänzen die Funktion $y_1 = G_1(x_1, x_2) = x_1 x_2$ durch $y_2 = G_2(x_1, x_2) = x_2$. Auflösung dieser beiden Gleichungen nach den y_i ergibt für $x_2 = y_2 \neq 0$ die Umkehrfunktion G_1^*:

$$\begin{aligned} x_1 &= \tfrac{y_1}{y_2} \\ x_2 &= y_2, \end{aligned}$$

wobei die Menge M_1 umkehrbar eindeutig auf die Menge

$$M_1^* \;=\; \{y = (y_1, y_2) \;;\; 0 < y_1 < y_2 < 1\}$$

abgebildet wird. Die Funktionaldeterminante der Abbildung G ist

$$J_G(x) = det \begin{pmatrix} x_2 & x_1 \\ 0 & 1 \end{pmatrix} = x_2$$

und Einsetzen ergibt $J_G(G_1^*(y)) = y_2$, wodurch man die Dichte

$$g(y_1, y_2) = \begin{cases} 1 \cdot \frac{1}{y_2} & \text{für } 0 < y_1 < y_2 < 1 \\ 0 & \text{sonst} \end{cases}$$

für die Verteilung P^G erhält.
Für $y_1 \leq 0$ und $y_1 \geq 1$ ist $g(y_1, y_2) = 0$ und damit auch die erste Marginaldichte $g_1(y_1) = \int g(y_1, y_2) dy_2 = 0$. Für $0 < y_1 < 1$ erhält man

$$g_1(y_1) = \int_{y_1}^{1} \frac{1}{y_2} dy_2 = -\log y_1.$$

Beispiel 9.2 *X_1 und X_2 seien stochastisch unabhängige und $\mathcal{N}(0,1)$-verteilte Zufallsvariable. Welche Verteilung besitzt dann die Zufallsvariable $Y = X_1^2 + X_2^2$?*

Hier ist die Dichte der Verteilung P^X gleich

$$f(x_1, x_2) \;=\; \phi(x_1)\phi(x_2) \;=\; \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}.$$

Wenn man die Funktion $y_1 = G_1(x_1, x_2) = x_1^2 + x_2^2$ durch $y_2 = G_2(x_1, x_2) = x_2^2$ ergänzt, so ist die Funktionaldeterminante $J_G(x) = 4 x_1 x_2$ gleich Null auf der Menge

$$N \;=\; \{x = (x_1, x_2) \in \mathbf{R}^2 \;;\; x_1 = 0 \text{ oder } x_2 = 0\,\}.$$

Dies ist aber eine Lebesguesche Nullmenge des $\mathbf{R}^2$, so daß insbesondere auch $P^X(N) = \int 1_N(x) f(x) dx = 0$. Außerdem ist die Abbildung G nur noch umkehrbar, wenn man sie auf einen der vier offenen Quadranten M_1, M_2, M_3, M_4 einschränkt. Die Umkehrfunktionen G_k^* sind durch die Gleichungen

$$\begin{aligned} x_1 &= \pm\sqrt{y_1 - y_2} \\ x_2 &= \pm\sqrt{y_2} \end{aligned}$$

mit entsprechender Wahl der Vorzeichen gegeben und die Bildmengen der M_k sind für alle k gleich:

$$\begin{aligned} &M_1^* = M_2^* = M_3^* = M_4^* = M^* \\ &= \{y = (y_1, y_2) \;;\; 0 < y_2 < y_1\} \end{aligned}$$

Wegen $f(G_k^*(y)) = \frac{1}{2\pi}e^{-\frac{1}{2}y_1}$ und

$$|J_G(G_k^*(y))| = 4\sqrt{y_1 - y_2}\sqrt{y_2}$$

sind alle vier Summanden in der Formel des Satzes 10.4 gleich, so daß die Dichte von P^G gleich

$$g(y_1, y_2) = 4 \cdot \frac{1}{2\pi}e^{-\frac{1}{2}y_1} \cdot \frac{1}{4\sqrt{y_1 y_2 - y_2^2}} \cdot 1_{M^*}(y_1, y_2)$$

ist.
Für $y_1 \leq 0$ ist $g(y_1, y_2)$ und damit auch die Marginaldichte $g_1(y_1)$ Null. Für $y_1 > 0$ ergibt sich

$$g_1(y_1) = \frac{1}{2\pi}e^{-\frac{1}{2}y_1} \int_0^{y_1} \frac{1}{\sqrt{y_1 y_2 - y_2^2}} dy_2$$

und durch Variablensubstitution $s = 2\frac{y_2}{y_1} - 1$ im Integral erhält man

$$\begin{aligned} g_1(y_1) &= \tfrac{1}{2\pi}e^{-\frac{1}{2}y_1} \int_{-1}^{1} \tfrac{1}{\sqrt{1-s^2}} ds \\ &= \tfrac{1}{2}e^{-\frac{1}{2}y_1}. \end{aligned}$$

P^{G_1} bzw. P^Y ist also die Exponentialverteilung mit dem Parameter $\lambda = \frac{1}{2}$.

Dieser Lösungsweg sollte lediglich die Anwendung des Transformationssatzes für den Fall verdeutlichen, daß mehrere Mengen M_k eingeführt werden müssen, denn die Verteilung P^{G_1} läßt sich einfacher über ihre Verteilungsfunktion $F^{G_1}(t)$ bestimmen:

$$\begin{aligned} F^{G_1}(t) &= P^{G_1}(-\infty, t] \\ &= P^X(G_1 \leq t) \\ &= P^X\{(x_1, x_2)\ ;\ x_1^2 + x_2^2 \leq t\} \\ &= \int\int 1_{\{x_1^2+x_2^2 \leq t\}} \tfrac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)} dx_1 dx_2 \end{aligned}$$

Für $t < 0$ ist die Menge $\{x_1^2 + x_2^2 \leq t\}$ leer und somit ihre Wahrscheinlichkeit $F^{G_1}(t) = 0$. Für $t \geq 0$ geht man im Integral zu Polarkoordinaten über und erhält

$$\begin{aligned} F^{G_1}(t) = &\int_0^{\sqrt{t}} \int_0^{2\pi} \tfrac{1}{2\pi} r e^{-\frac{1}{2}r^2} dr d\phi \\ = &\quad \cdots \\ = &\quad 1 - e^{-\frac{1}{2}t} \end{aligned}$$

Faltungen.
Zur Berechnung der Verteilung einer Summe $Y = X_1 + X_2 + \cdots + X_n$ von Zufallsvariablen ergänzt man die Funktion $y_1 = G_1(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$ durch $y_2 = x_2, y_3 = x_3, \ldots, y_n = x_n$ zu einer Abbildung G, die auf dem ganzen $\mathbf{R}^n$ eineindeutig mit Umkehrfunktion

$$\begin{aligned} x_1 &= y_1 - y_2 - \cdots - y_n \\ x_2 &= y_2 \\ \vdots\ &\ \vdots\ \ \vdots \\ x_n &= y_n \end{aligned}$$

und der Funktionaldeterminante $J_G(x) = 1$ ist. Besitzt die Verteilung P^X des Zufallsvektors $X = (X_1, X_2, \ldots, X_n)$ die Dichte $f(x_1, x_2, \ldots, x_n)$, so ist die Dichte der Verteilung P^Y dementsprechend durch

$$g_1(y_1) = \int \cdots \int \quad f(y_1 - y_2 - \cdots - y_n, y_2, \ldots, y_n) \, dy_2 dy_3 \cdots dy_n$$

gegeben.
Speziell im Fall $n = 2$ und unabhängigen Variablen X_1, X_2 bzw. $f(x_1, x_2) = f_1(x_1) f_2(x_2)$ ist

$$g_1(y_1) = \int f_1(y_1 - y_2) f_2(y_2) dy_2$$

das Faltungsintegral der Dichten f_1 und f_2.

10 Erwartungswert, Varianz, Kovarianz.

Der Erwartungswert einer Treppenfunktion. Seien $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\{A_j\}$ eine Partition von Ω mit j aus einer endlichen oder abzählbaren Indexmenge J und x_j reelle Zahlen.

Definition 10.1 *Eine Funktion $X : \Omega \longrightarrow \mathbf{R}$ der Form*

$$X(\omega) = \sum_{j \in J} x_j 1_{A_j}(\omega) \tag{10.1}$$

heißt eine Treppenfunktion.

Definition 10.2 *Die Treppenfunktion X heißt P-integrabel, wenn*

$$\sum_j |x_j| P(A_j) \ < \ \infty$$

gilt. In diesem Fall heißt der Wert der Summe

$$\sum_j x_j P(A_j) \tag{10.2}$$

der Erwartungswert *von X bezüglich P Er wird mit $\mathcal{E}_P X$ bezeichnet.*

Der Wert $\mathcal{E}_P X$ ist unabhängig von der speziellen Darstellung 10.1 der Treppenfunktion.
Der Erwartungswert einer Zufallsvariablen. Sei nunmehr $X : \Omega \longrightarrow \mathbf{R}$ eine beliebige Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Zur Definition des Erwartungswerts von X approximieren wir X durch eine Folge $\{X_n\}$ von Treppenfunktionen, die wie folgt festgelegt werden:
Für ein festes $n \in \{1, 2, \ldots\}$ bilden die Mengen

$$A_k^{(n)} := \{\omega \in \Omega\ ;\ \frac{k}{2^n} < X(\omega) \leq \frac{k+1}{2^n}\}$$

mit $k \in \mathbf{Z} := \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ eine Partition von Ω und die Treppenfunktionen

$$X_n(\omega) := \sum_{k \in \mathbf{Z}} \frac{k}{2^n} 1_{A_k^{(n)}}(\omega)$$

konvergieren für $n \longrightarrow \infty$ wegen

$$X(\omega) - \frac{1}{2^n} \leq X_n(\omega) \leq X(\omega) \quad \text{für alle} \quad \omega$$

gleichmäßig gegen X :

$$\|X - X_n\|_\infty \overset{n\to\infty}{\longrightarrow} 0$$

Für das Verhalten der Erwartungswerte dieser Treppenfunktionen läßt sich der folgende Satz beweisen:

Satz 10.1 *Falls für ein n_0 der Erwartungswert der Treppenfunktion X_{n_0} existiert, dann existieren alle $\mathcal{E}_P X_n$ und die Folge dieser Erwartungswerte konvergiert für $n \longrightarrow \infty$ gegen eine Zahl e_X, wobei*

$$|\mathcal{E}_P X_n - e_X| \leq \|X_n - X\|_\infty.$$

Ist $\{Y_n\}$ eine weitere Folge von Treppenfunktionen mit der Eigenschaft $\|X - Y_n\|_\infty \overset{n\to\infty}{\longrightarrow} 0$, so existieren dann ebenfalls alle Erwartungswerte $\mathcal{E}_P Y_n$ und konvergieren gegen diese Zahl e_X.

Definition 10.3 *Der Limes e_X der Erwartungswerte der Treppenfunktionen X_n heißt der* Erwartungswert der Zufallsvariablen X *und wird mit $\mathcal{E}_P X$ bezeichnet.*

Bemerkungen: Wenn aus dem Zusammenhang klar ersichtlich ist, bezüglich welcher Wahrscheinlichkeit P der Erwartungswert gebildet wird, schreibt man meist nur $\mathcal{E}X$ anstelle von $\mathcal{E}_P X$. Anstelle der Formulierung "X ist P-integrabel" verwenden wir in der Folge häufig den Satz: "Der Erwartungswert von X existiert."

Satz 10.2 *Sind $X_1, X_2, \ldots, X_n$ Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, deren Erwartungswerte existieren, und $a_1, a_2, \ldots, a_n$ und b reelle Zahlen, so gilt*

$$\begin{aligned} &\mathcal{E}_P(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + b) \;= \\ &a_1 \mathcal{E}_P X_1 + a_2 \mathcal{E}_P X_2 + \cdots + a_n \mathcal{E}_P X_n + b. \end{aligned}$$

Sind die X_k außerdem stochastisch unabhängig, *so ist*

$$\mathcal{E}_P(X_1 \cdot X_2 \cdots X_n) \;=\; \mathcal{E}_P X_1 \cdot \mathcal{E}_P X_2 \cdots \mathcal{E}_P X_n.$$

Satz 10.3 *Die Verteilung P^X der Zufallsvariablen X sei absolut stetig oder diskret. Dann gilt: Der Erwartungswert $\mathcal{E}_P X$ existiert genau dann, wenn $m_1(P^X)$ existiert und es gilt*

$$\mathcal{E}_P X \;=\; m_1(P^X).$$

Berechnung von Erwartungswerten. Die in den Anwendung häufigste Problemstellung ist die Berechnung des Erwartungswerts einer Funktion $Y = G(X_1, X_2, \ldots, X_n)$ von Zufallsvariablen X_k, z.B. von $Y = X_1^2 + X_2^2 + \cdots + X_n^2$, bei bekannten Verteilungen der X_k. Nach Satz 4.1 ist $P^Y = P^G$, so daß wegen

$$\mathcal{E}_P Y = m_1(P^Y) = m_1(P^G) = \mathcal{E}_{P^X} G$$

diese Aufgabe darin besteht, den Erwartungswert einer Zufallsvariable $G : \mathbf{R}^n \longrightarrow \mathbf{R}$ in der Situation

$$(\mathbf{R}^n, \mathcal{B}_n, P) \xrightarrow{G} (\mathbf{R}, \mathcal{B}, P^G)$$

zu berechnen, wobei die Verteilung P $(= P^X)$ durch eine Dichte (oder Wahrscheinlichkeitsfunktion) gegeben ist: $P(B) = \int 1_B(x) f(x) dx$. Dafür gibt es im wesentlichen zwei verschiedene Lösungswege. Wegen $\mathcal{E}_P G = m_1(P^G)$ besteht eine Möglichkeit darin, die Dichte oder Wahrscheinlichkeitsfunktion $g(t)$ der Verteilung P^G und anschließend den Mittelwert $m_1(P^G) = \int t g(t) dt$ zu berechnen. Der andere Weg besteht in der Berechnung eines Mehrfachintegrals:

Satz 10.4 *In der oben beschriebenen Situation gilt*

$$\mathcal{E}_P G \;=\; \int G(x) f(x) dx.$$

Beispiel 10.1 *$X_1, X_2, \ldots, X_n$ seien $\mathcal{N}(0,1)$-verteilte Zufallsvariable. Welchen Erwartungswert besitzt $Y = X_1^2 + X_2^2 + \cdots + X_n^2$?*

Wir wenden als Erstes den Satz 10.2 an:

$$\begin{aligned} \mathcal{E}Y &= \mathcal{E}(X_1^2 + X_2^2 + \cdots + X_n^2) \\ &= \mathcal{E}(X_1^2) + \mathcal{E}(X_2^2) + \cdots + \mathcal{E}(X_n^2) \end{aligned}$$

und betrachten jeden Summanden für sich: P^{X_k} ist die $\mathcal{N}(0,1)$-Verteilung mit der Dichte $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$, so daß wegen $X^2(\omega) = G(X(\omega)$ mit $G(x) = x^2$

$$\begin{aligned} \mathcal{E}(X_k^2) = \mathcal{E}_{P^{X_k}} G &= \int G(x) \varphi(x) dx \\ &= \int x^2 \varphi(x) dx = m_2(\mathcal{N}(0,1)) = 1 \end{aligned}$$

und damit

$$\mathcal{E}Y = n$$

gilt.

Die Varianz einer Zufallsvariablen.
Definition 10.4

$$\mathrm{var}_P X = \mathcal{E}_P (X - \mathcal{E}_P X)^2$$

heißt die Varianz *der Zufallsvariablen X.*

Satz 10.5 *Die Verteilung P^X der Zufallsvariablen X sei absolut stetig oder diskret. Dann gilt:*

$$\mathrm{var}_P X = \hat{m}_2(P^X).$$

Bei den folgenden Rechenregeln lassen wir den Index P beim $\mathcal{E}$- und var-Operator weg:

$$\mathrm{var} X = \mathcal{E}(X^2) - (\mathcal{E}X)^2$$

Dies folgt aus

$$\begin{aligned} &\mathcal{E}(X - \mathcal{E}X)^2 \\ &= \mathcal{E}\left[X^2 - 2\mathcal{E}X \cdot X + (\mathcal{E}X)^2\right] \\ &= \mathcal{E}(X^2) + (-2\mathcal{E}X) \cdot \mathcal{E}X + (\mathcal{E}X)^2 \\ &= \mathcal{E}(X^2) - 2(\mathcal{E}X)^2 + (\mathcal{E}X)^2 \end{aligned}$$

wobei die Linearität des "$\mathcal{E}$"-Operators benutzt wurde. Für reelle Zahlen a, b gilt

$$\operatorname{var}(aX+b) = a^2 \operatorname{var} X; \qquad (10.3)$$

denn

$$\begin{aligned}\mathcal{E}[(aX+b)-\mathcal{E}(aX+b)]^2 &= \mathcal{E}[aX+b-a\mathcal{E}X-b]^2 \\ = \mathcal{E}\left[a^2(X-\mathcal{E}X)^2\right] &= a^2\mathcal{E}(X-\mathcal{E}X)^2.\end{aligned}$$

Für die Summe zweier Zufallsvariabler X und Y gilt schließlich

$$\begin{gathered}\operatorname{var}(X+Y) \\ = \mathcal{E}[(X+Y)-\mathcal{E}(X+Y)]^2 \\ = \mathcal{E}[(X-\mathcal{E}X)+(Y-\mathcal{E}Y)]^2 \\ = \mathcal{E}\left[(X-\mathcal{E}X)^2+2(X-\mathcal{E}X)(Y-\mathcal{E}Y)+(Y-\mathcal{E}Y)^2\right] \\ = \mathcal{E}(X-\mathcal{E}X)^2+2\mathcal{E}\left[(X-\mathcal{E}X)(Y-\mathcal{E}Y)\right]+\mathcal{E}(Y-\mathcal{E}Y)^2 \\ =: \operatorname{var}X + 2\operatorname{cov}(X,Y)+\operatorname{var}Y\end{gathered}$$

Definition 10.5 *Die Zahl*

$$\operatorname{cov}(X,Y) = \operatorname{cov}_P(X,Y) = \mathcal{E}\left[(X-\mathcal{E}X)(Y-\mathcal{E}Y)\right]$$

heißt die Kovarianz *der Zufallsvariablen X und Y.*

Für diesen Operator gilt

$$\begin{gathered}\operatorname{cov}(X,Y) \\ = \mathcal{E}\left[XY-(\mathcal{E}X)Y-(\mathcal{E}Y)X+(\mathcal{E}X)(\mathcal{E}Y)\right] \\ = \mathcal{E}(XY)-(\mathcal{E}X)\mathcal{E}Y-(\mathcal{E}Y)\mathcal{E}X+(\mathcal{E}X)(\mathcal{E}Y) \\ = \mathcal{E}(XY)-(\mathcal{E}X)(\mathcal{E}Y).\end{gathered}$$

Aus dieser Formel, dem Satz 10.2 und (10.4) folgt dann

Satz 10.6 *Sind die Zufallsvariablen X und Y* stochastisch unabhängig, *so ist* $\operatorname{cov}(X,Y)=0$ *und es gilt* $\operatorname{var}(X+Y)=\operatorname{var}X+\operatorname{var}Y$.

Zur Berechnung der Kovarianz muß gemäß (10.4) der Erwartungswert der Zufallsvariablen $X\cdot Y$ bestimmt werden. Aus dem obigen Diagramm ersieht man, daß

$$\begin{tikzcd}(\Omega,\mathcal{A},P) \arrow[r, "{(X,Y)}"] \arrow[dr, "XY"'] & (\mathbf{R},\mathcal{B},P^{(X,Y)}) \arrow[d, "{G(x,y)=xy}"] \\ & (\mathbf{R},\mathcal{B},P^G)\end{tikzcd}$$

dies durch die Auswertung des Integrals

$$\mathcal{E}(XY) = \mathcal{E}_{P(X,Y)}G = \int_{\mathbf{R}^2} xyf(x,y)d(x,y)$$

geschieht, wobei f die Dichte der Verteilung $P^{(X,Y)}$ ist. (Entsprechendes gilt für diskrete Verteilungen mit Wahrscheinlichkeitsfunktion f).

Direkt aus den Definitionen lassen sich die folgenden beiden Formeln ablesen:

$$\begin{aligned}\operatorname{cov}(X,X) &= \operatorname{var}X \\ \operatorname{cov}(Y,X) &= \operatorname{cov}(X,Y)\end{aligned}$$

und als wichtigste Rechenregel für den Kovarianzoperator notieren wir den

Satz 10.7 *Sind $X_1, X_2, \ldots, X_m$ und $Y_1, Y_2, \ldots, Y_n$ Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$, $a_1,a_2,\ldots,a_m$ und $b_1,b_2,\ldots,b_n$ reelle Zahlen, so gilt*

$$\begin{gathered}\operatorname{cov}(\textstyle\sum_{i=1}^m a_iX_i, \sum_{k=1}^n b_kY_k) = \\ \textstyle\sum_{i=1}^m\sum_{k=1}^n a_ib_k\operatorname{cov}(X_i,Y_k) = \\ a^T\begin{pmatrix}\operatorname{cov}(X_1,Y_1) & \ldots & \operatorname{cov}(X_1,Y_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_m,Y_1) & \ldots & \operatorname{cov}(X_m,Y_n)\end{pmatrix}b,\end{gathered}$$

wo a^T der Zeilenvektor mit den Zahlen a_i und b der Spaltenvektor mit den b_k ist.

Beweis: Sei $Z := \sum_{k=1}^n b_kY_k$. Dann ist

$$\begin{gathered}\operatorname{cov}(\textstyle\sum_{i=1}^m a_iX_i, Z) \\ = \mathcal{E}\left\{[\textstyle\sum_{i=1}^m a_iX_i - \mathcal{E}(\sum_{i=1}^m a_iX_i)](Z-\mathcal{E}Z)\right\} \\ = \mathcal{E}\left[\textstyle\sum_{i=1}^m a_i(X_i-\mathcal{E}X_i)(Z-\mathcal{E}Z)\right] \\ = \textstyle\sum_{i=1}^m a_i\mathcal{E}(X_i-\mathcal{E}X_i)(Z-\mathcal{E}Z) \\ = \textstyle\sum_{i=1}^m a_i\operatorname{cov}(X_i,Z)\end{gathered}$$

Auf die gleiche Weise erhält man für jeden einzelnen Summanden der letzten Zeile

$$\begin{aligned}\operatorname{cov}(X_i,Z) &= \operatorname{cov}(X_i,\textstyle\sum_{k=1}^n b_kY_k) \\ &= \textstyle\sum_{k=1}^n b_k\operatorname{cov}(X_i,Y_k)\end{aligned}$$

und damit die behauptete Formel.

Zum Abschluss dieses Abschnitts noch zwei Bezeichnungen: Für einen Zufallsvektor $X=(X_1,\ldots,X_n)$ heißt $\mathcal{E}X=(\mathcal{E}X_1,\ldots,\mathcal{E}X_n)$ der *Erwartungswert* und

$$\mathcal{C}_X = \begin{pmatrix}\operatorname{cov}(X_1,X_1) & \ldots & \operatorname{cov}(X_1,X_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n,X_1) & \ldots & \operatorname{cov}(X_n,X_n)\end{pmatrix}$$

die *Kovarianzmatrix* des Zufallsvektors X.

11 Die Normalverteilung

Die eindimensionale Normalverteilung

Die Verteilung P auf $\mathbf{R}$ mit der Dichte $\varphi(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$ heißt *gaussische Einheitsverteilung, standardisierte Normalverteilung oder* $\mathcal{N}(0,1)$*-Verteilung.* Eine Zufallsvariable G mit Verteilung $P^G=P$ heißt *gaussische Einheitsvariable.*

Sind μ und σ reelle Zahlen mit $\sigma\neq 0$, so besitzt die Zufallsvariable $X(t)=\mu+\sigma\,t$ auf $(\mathbf{R},\mathcal{B},P)$ die Verteilung P^X mit der Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad (11.1)$$

dem Mittelwert $m_1(P^X) = \mu$ und der Varianz $\hat{m}_2(P^X) = \sigma^2$.

Definition 11.1 *Die Verteilung mit der Dichte 11.1 heißt die* Normalverteilung mit Mittelwert μ und Varianz σ^2 oder $\mathcal{N}(\mu, \sigma^2)$-Verteilung.

Die mehrdimensionale Normalverteilung
Die Verteilung P auf dem $\mathbf{R}^n$ mit der Dichte

$$\begin{aligned} g(y) &= g(y_1, \ldots, y_n) = \varphi(y_1) \cdot \varphi(y_2) \cdots \varphi(y_n) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\|y\|^2} \end{aligned}$$

heißt *n-dimensionale gaussische Einheitsverteilung*. Sie ist die Verteilung eines *gaussischen Einheitsvektors*, d.h. eines Zufallsvektors $G = (G_1, \ldots, G_n)$, dessen Komponenten G_k stochastisch unabhängige gaussische Einheitsvariable sind.

Definition 11.2 *Sei $m \in \mathbf{R}^n$ und C eine symmetrische, positiv definite $n \times n$-Matrix. Dann heißt die Verteilung auf dem $\mathbf{R}^n$ mit der Dichte*

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{detC}} e^{-\frac{1}{2}(x-m)^T C^{-1}(x-m)} \quad (11.2)$$

die n-dimensionale Normalverteilung mit Mittelwert m und Kovarianzmatrix C oder $\mathcal{N}(m, C)$-Verteilung.

Der Grund für diese Bezeichnungen ergibt sich, wenn man den n-dimensionalen Zufallsvektor $X(y) = Ay+m$ auf dem Wahrscheinlichkeitsraum $(\mathbf{R}^n, \mathcal{B}_n, P)$ mit der n-dimensionalen gaussischen Einheitsverteilung P und einer $n \times n$-Matrix A mit der Eigenschaft $C = AA^T$ betrachtet. Dieser Zufallsvektor besitzt den Erwartungswert $\mathcal{E}X = m$, die Kovarianzmatrix $\mathcal{C}_X = C$ und als Verteilung—wie man mit dem Transformationssatz nachrechnet—die Normalverteilung mit der Dichte 11.2.
Daraus ergibt sich für die Komponenten $X_1, \ldots, X_n$ des Zufallsvektors X der

Satz 11.1 *Die $X_1, \ldots, X_n$ sind genau dann stochastisch unabhängig, wenn für alle $i \neq k$ gilt $cov(X_i, X_k) = 0$.*

Beweis Sind die Kovarianzen Null, so ist die Kovarianzmatrix eine Diagonalmatrix mit Diagonalelementen $\sigma_i^2 = \mathrm{var}X_i$ und die Dichte (11.2) hat die Form

$$f(x) = f_1(x_1) \cdot f_2(x_2) \cdots f_n(x_n),$$

wobei die f_k die Form (11.1) mit $\mu = m_k$ und $\sigma^2 = \sigma_k^2$ besitzen.

Satz 11.2 *Ist G ein gaussischer Einheitsvektor und U eine Orthogonalmatrix passender Dimension, so ist der Zufallsvektor $X = UG$ wieder ein gaussischer Einheitsvektor.*

Beweis folgt aus UU^T =Einheitsmatrix.

Satz 11.3 *Die Verteilungen P^{X_k} der Komponenten X_k des Zufallsvektors $X(y) = AG + m$ sind die eindimensionalen Normalverteilungen mit Mittelwerten m_k und Varianzen σ_k^2 = k-tem Diagonalelement von $C = AA^T$.*

Beweis für X_1:
Mit der euklidischen Norm

$$\|a\| = \sqrt{\sum_{k=1}^n a_{1k}^2} = \sqrt{\sigma_1^2}$$

der ersten Zeile der Matrix A und $u_{1k} := \frac{a_{1k}}{\|a\|}$ ist

$$\begin{aligned} X_1 &= \sum_{k=1}^n a_{1k} G_k + m_1 \\ &= \|a\| \sum_{k=1}^n u_{1k} G_k + m_1 \\ &= \|a\| H_1 + m_1 \qquad (11.3) \end{aligned}$$

Ist U eine Orthogonalmatrix mit $u_{11}, \ldots, u_{1n}$ als erster Zeile, so ist nach Satz 11.2 der Vektor $H = UG$ ein gaussischer Einheitsvektor, bzw. seine erste Komponente H_1 eine $\mathcal{N}(0,1)$-verteilte Zufallsvariable. Aus der Form (11.3) der Zufallsvariable X_1 und Übungsaufgabe 28 folgt dann die Behauptung.

12 Produkte von Wahrscheinlichkeitsräumen

Endliche Produkte.
Seien $(\Omega_k, \mathcal{A}_k, P_k)$, $k = 1, 2, \ldots, n$, n Wahrscheinlichkeitsräume,

$$\begin{aligned} \Omega &= \bigotimes_{k=1}^n \Omega_k \\ &= \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n \\ &= \{\omega = (\omega_1, \ldots, \omega_n)\ ;\ \omega_k \in \Omega_k\ ,\ k = 1, \ldots, n\} \end{aligned}$$

das kartesische Produkt der Ergebnismengen und $Z_k : \Omega \longrightarrow \Omega_k$ die durch

$$Z_k(\omega_1, \ldots, \omega_n) = \omega_k$$

definierten Projektionsabbildungen.

Definition 12.1 *Die kleinste σ-Algebra, die alle Urbilder $(Z_k \in A_k)$ mit $A_k \in \mathcal{A}_k$ und $k = 1, 2, \ldots, n$ enthält, heißt die Produkt-σ-Algebra der $\mathcal{A}_k$ und wird mit $\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \cdots \otimes \mathcal{A}_n$ oder*

$$\bigotimes_{k=1}^n \mathcal{A}_k$$

bezeichnet.

Satz 12.1 *Es gibt genau eine Wahrscheinlichkeit P auf der Produkt-σ-Algebra der $\mathcal{A}_k$ mit den Eigenschaften*

$$P(Z_k \in A_k) = P_k(A_k) \text{ für alle } k \text{ und } A_k \qquad (12.1)$$

und

$$P\left[\bigcap_{k=1}^{n}(Z_k \in A_k)\right] = \prod_{k=1}^{n} P(Z_k \in A_k) \qquad (12.2)$$

Wegen

$$\begin{aligned}\bigcap_{k=1}^{n}(Z_k \in A_k) &= \{\omega = (\omega_1, \ldots, \omega_n) \in \Omega \; ; \; \omega_k \in A_k\} \\ &= A_1 \times A_2 \times \cdots \times A_n\end{aligned}$$

lassen sich die beiden Bedingungen (12.1) und (12.2) zu

$$P(A_1 \times A_2 \times \cdots \times A_n) = P_1(A_1) \cdot P_2(A_2) \cdots P_n(A_n) \qquad (12.3)$$

zusammenfassen.

Definition 12.2 *Die Wahrscheinlichkeitsverteilung P aus Satz 12.1 heißt das Produkt der Wahrscheinlichkeitsverteilungen P_k und wird mit $P_1 \otimes P_2 \otimes \cdots \otimes P_n$ oder*

$$\bigotimes_{k=1}^{n} P_k$$

bezeichnet.

Definition 12.3 *Der Wahrscheinlichkeitsraum $(\bigotimes_{k=1}^{n}\Omega_k, \bigotimes_{k=1}^{n}\mathcal{A}_k, \bigotimes_{k=1}^{n}P_k)$ heißt das Produkt der Wahrscheinlichkeitsräume $(\Omega_k, \mathcal{A}_k, P_k)$.*

Es ist dadurch charakterisiert, daß die Ereignisse A_k aus den verschiedenen Einzelexperimenten im Rahmen des Gesamtexperiments stochastisch unabhängig sind und die gleichen Wahrscheinlichkeiten wie im ursprünglichen Experiment besitzen.

13 Aufgaben

bf 1) $\mathcal{A}$ sei die Menge der Teilmengen von natürlichen Zahlen, die entweder selbst endlich sind, oder deren Komplement endlich ist.
Zeigen Sie, daß $\mathcal{A}$ eine Mengenalgebra, aber keine σ-Algebra ist.
2) Für Tüftler: Gibt es eine Mengenoperation $\star$, so daß man eine Mengenalgebra als eine Menge $\mathcal{A}$ von Teilmengen von Ω mit den Eigenschaften

$$\begin{aligned}\Omega &\in \mathcal{A} \\ A, B \in \mathcal{A} &\Rightarrow A \star B \in \mathcal{A}\end{aligned}$$

definieren kann?
3)

a) Stellen Sie Wahrscheinlichkeitsräume für das Experiment "Werfen eines Paars regulärer Würfel" auf, wobei einmal die beiden Würfel unterscheidbar und zum andern nicht unterscheidbar sind.

b) Das Experiment a) wird zweimal durchgeführt. Berechnen Sie die Wahrscheinlichkeit dafür, daß dabei das gleiche Ergebnis zustandekommt.

4) Ein regulärer Würfel wird dreimal geworfen. Wie groß ist die Wahrscheinlichkeit, daß dabei die kleinste geworfene Augenzahl die 2 und die größte die 5 ist?
5) P sei eine Wahrscheinlichkeit auf einer σ-Algebra $\mathcal{A}$ über N, die die Mengen

$$A_m = \{k \in \mathrm{N} \; ; \; k \text{ ist durch } m \text{ teilbar}\}$$

für $m = 1, 2, \ldots$ enthält. Es sei $P(A_m) = 1/m$.
Berechnen Sie die Wahrscheinlichkeit der Menge der natürlichen Zahlen, die teilbar durch 3, nicht teilbar durch 5 und teilbar durch 4 oder 6 sind.
6) Ein regulärer Würfel wird dreimal nacheinander geworfen. Wie groß ist die Wahrscheinlichkeit, daß man dabei eine streng monotone Folge von Augenzahlen erhält?
7) Bei einer Lotterie werde eine vierstellige Losnummer auf die folgende Weise ermittelt: In einer Trommel befinden sich 40 mit den Ziffern $0, 1, \ldots, 9$ versehene Kugeln, wobei jeweils 4 Kugeln die gleiche Ziffer tragen. Es werden 4 Kugeln ohne Zurücklegen gezogen und in der Reihenfolge ihrer Ziehung nebeneinandergelegt. Die 4 Ziffern auf den gezogenen Kugeln ergeben dann die Losnummer.
Berechnen Sie die Wahrscheinlichkeiten der folgenden Ereignisse:

a) Die Losnummer 3494 wird gezogen.

b) Es wird eine Losnummer gezogen, in der nur die Ziffern 1 und 7 vorkommen. (Beide Ziffern sollen vorkommen.)

c) Es wird eine Losnummer gezogen, die aus 4 verschiedenen Ziffern besteht.

8) Auf wieviele Arten kann man K nicht unterscheidbare Kugeln auf N Schubladen verteilen, wenn in jede Schublade eine beliebige Anzahl (auch 0) von Kugeln gelegt werden darf?
9) Dem Zöllner NN ist zugespielt worden, daß unter den 40 Passagieren eines gerade ankommenden Fährschiffs zwei Schmuggelware mit sich führen. Wieviele der Passagiere muß er (zufällig) zur Kontrolle auswählen, um mit einer Wahrscheinlichkeit von mindestens 0,9 wenigstens einen der Schmuggler zu erwischen?
10) Zu einem Praktikum haben sich N Personen angemeldet, aber nur $M < N$ Plätze stehen zur Verfügung. Zur Auslosung der Personen, die Plätze bekommen, wird das folgende Verfahren angewandt: In eine Schachtel werden N Zettel gelegt; auf M von

ihnen steht „Ja“, auf dem Rest „Nein“. In der Reihenfolge ihres Eintreffens zieht jede der angemeldeten Personen einen Zettel. Steht auf dem Zettel „Ja“, erhält die Person einen Platz. Ist dieses Verfahren gerecht, d. h. hat jeder die gleiche Chance, einen Praktikumsplatz zu erhalten?

11) Gegeben sind n (numerierte) Urnen, die jeweils w weiße und s schwarze Kugeln enthalten. Für $k = 1, 2, \ldots, n-1$ wird nacheinander aus Urne k zufällig eine Kugel gezogen und in Urne $k+1$ gelegt.
Wie groß ist die Wahrscheinlichkeit, daß eine anschließend aus Urne n gezogene Kugel weiß ist?

12) Gegeben sind zwei Urnen U_1 und U_2. Urne U_1 enthält 5 rote und 3 schwarze, Urne U_2 4 rote und 4 schwarze Kugeln. Aus U_1 wird zufällig eine Kugel gezogen und in Urne U_2 gelegt, dann zieht man zufällig eine Kugel aus U_2 und legt sie in U_1. Wie groß ist die Wahrscheinlichkeit, daß man beim anschließenden Ziehen einer Kugel aus U_1 eine rote Kugel zieht?

13) In einer US-Fernsehshow kann man auf die folgende Weise ein Auto gewinnen: Der Kandidat hat die Wahl zwischen drei Türen, wobei sich hinter zweien je eine Ziege befindet, die beim Öffnen das Meckern anfangen wird und hinter einer das Auto, das man gewinnt, wenn man diese Tür gewählt hat.
Nachdem sich der Kandidat für eine Tür entschieden hat, öffnet der Showmaster eine der beiden anderen Türen und zwar eine, hinter der sich eine Ziege befindet.
Jetzt hat der Kandidat noch einmal die Möglichkeit, seine Entscheidung zu revidieren. Er kann bei seiner ursprünglichen Wahl bleiben oder sich für die andere der beiden noch geschlossenen Türen entscheiden.
Was würden Sie an Stelle des Kandidaten tun — bei der ersten Entscheidung bleiben oder Umdisponieren?

14) A und B seien zwei Ereignisse in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$und es sei $0 < P(B) < 1$. Beweisen Sie, daß die beiden Ereignisse genau dann stochastisch unabhängig sind, wenn $P(A|B) = P(A|\overline{B})$.

15) X und Y seien stochastisch unabhängige und mit Parametern λ bzw. μ Poisson-verteilte Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$.
Welche Verteilung besitzt die Zufallsvariable $Z = X + Y$, definiert durch $Z(\omega) = X(\omega) + Y(\omega)$ für $\omega \in \Omega$?

16) $X : \Omega \longrightarrow \mathbb{N}_0$ und $Y : \Omega \longrightarrow \mathbb{N}_0$ seien Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$mit folgenden Eigenschaften:
1. X ist Poisson-verteilt mit Parameter μ.
2. Für $n = 0, 1, 2, \ldots$ gilt

$$P[(Y = k)|(X = n)] = \begin{cases} \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} \\ \text{für } k = 0, 1, \ldots, n \\ 0 \\ \text{für } k > n \end{cases}$$

mit einer Zahl $0 < p < 1$.
Berechnen Sie die Verteilung der Zufallsvariablen Y.

Erzeugende Funktionen

Ist P eine diskrete Wahrscheinlichkeitsverteilung auf $\mathbb{N}_0$ oder einer Teilmenge von $\mathbb{N}_0$ mit der Wahrscheinlichkeitsfunktion $f(n)$, so heißt die Funktion

$$\hat{f}(z) = \sum_n f(n) z^n$$

für $0 \leq z \leq 1$ die **erzeugende Funktion** von P.

17) X_1 und X_2 seien stochastisch unabhängige Zufallsvariable mit Wertebereich $\mathbb{N}_0$, deren Verteilungen die erzeugenden Funktionen $\hat{f}_1(z)$ bzw. $\hat{f}_2(z)$ besitzen. Zeigen Sie, daß die Verteilung der Zufallsvariable $Y = X_1 + X_2$ die erzeugende Funktion

$$\hat{g}(z) = \hat{f}_1(z)\hat{f}_2(z)$$

besitzt.

18) Berechnen Sie die Wahrscheinlichkeitsfunktion der Verteilung mit der erzeugenden Funktion

$$\hat{f}(z) = e^{\lambda(z-1)}$$

19) $X_1, X_2, \ldots, X_n$ seien stochastisch unabhängige Zufallsvariable, die nur die Werte 0 und 1 annehmen und die alle die gleiche Verteilung mit der Wahrscheinlichkeitsfunktion $f(0) = q$ und $f(1) = p$ mit $p + q = 1$ besitzen. Berechnen Sie die Wahrscheinlichkeitsfunktion der Verteilung der Zufallsvariablen

$$S = \sum_{k=1}^{n} X_k$$

20) Beweisen Sie die folgende Aussage: Sind $X_i : \Omega \longrightarrow \Omega_i$ für $i = 1, \ldots, n$ stochastisch unabhängige Zufallsgrößen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$und $Y_i : \Omega_i \longrightarrow \hat{\Omega}_i$ beliebige meßbare Abbildungen, so sind die Zufallsgrößen $Z_i = Y_i \circ X_i$ ebenfalls stochastisch unabhängig. (Auf den Ω_i und $\hat{\Omega}_i$ seien jeweils σ-Algebren $\mathcal{A}_i$ bzw. $\hat{\mathcal{A}}_i$ vorgegeben.)

21) Lotto mit Zusatzzahl: Aus einer Urne mit 49 Kugeln, von denen 43 weiß und 6 schwarz sind, werden 7 Kugeln zufällig und ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, daß sich unter den ersten 6 gezogenen Kugeln

a) 6 schwarze

b) 3 schwarze

befinden und außerdem die siebte gezogene Kugel schwarz ist?

22) X und Y seien stochastisch unabhängige und mit Parameter p geometrisch verteilte Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Welche Verteilung besitzt die Zufallsvariable $Z = \min(X, Y)$, definiert durch $Z(\omega) = \min\{X(\omega), Y(\omega)\}$ für $\omega \in \Omega$?

23) Zeigen Sie, daß

$$f(k) = \frac{a^k}{(1+a)^{k+1}} \quad k = 0, 1, 2, \ldots$$

für $a > 0$ eine Wahrscheinlichkeitsfunktion zu einer Verteilung auf $\mathbf{N}_0$ ist und berechnen Sie den Mittelwert dieser Verteilung.

24) Zeigen Sie, daß für eine Zufallsvariable X mit Wertebereich $\mathbf{N}_0$ gilt

$$\mathcal{E}X = \sum_{n=1}^{\infty} P(X \geq n).$$

25) Beim Würfelspiel „Die verflixte Sechs“ kann man eine beliebige Anzahl von Würfeln vom Tisch nehmen und werfen. Ist unter den geworfenen Augenzahlen mindestens eine Sechs, so ist der Gewinn Null. Andernfalls erhält man die Summe der geworfenen Augenzahlen als Gewinn gutgeschrieben.
Wieviele Würfel sollte man nehmen?

26) Wie dick im Verhältnis zum Durchmesser müßte eine Münze sein, damit sie beim Wurf mit Wahrscheinlichkeit $\frac{1}{3}$ auf dem Rand stehen bleibt?

27) Zwei Zahlen werden zufällig aus dem Intervall $[0,1]$ ausgewählt.
Wie groß ist unter der Bedingung, daß das Maximum der beiden Zahlen größer als $\frac{1}{2}$ ist, die Wahrscheinlichkeit, daß ihr Minimum kleiner als $\frac{1}{4}$ ist?

28) Ein Stab wird zufällig an zwei Stellen auseinandergebrochen. Wie groß ist die Wahrscheinlichkeit, daß man aus den drei Bruchstücken ein Dreieck formen kann?

29) Ein Stab wird an einer zufällig ausgewählten Stelle auseinandergebrochen und das längere der beiden Stücke noch einmal zufällig geteilt. Wie groß ist die Wahrscheinlichkeit, daß man aus den drei Bruchstücken ein Dreieck formen kann?

30) Für welche Werte der Konstanten c sind die nachstehenden Funktionen Dichten zu eindimensionalen Verteilungen?

$$f_1(x) = \begin{cases} cx(1-x) & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

$$f_2(x) = \frac{c}{\alpha^2 + (x-\beta)^2}$$

mit Parametern $\alpha > 0$ und $\beta \in \mathbf{R}$.

31) Für welchen Wert des Paramaters c ist die Funktion $f : \mathbf{R}^2 \to \mathbf{R}$, definiert durch

$$f(x_1, x_2) = \begin{cases} ce^{-(2x_1+3x_2)} & \text{für } x_1 > 0 \text{ und } 0 < x_2 < x_1 \\ 0 & \text{sonst} \end{cases}$$

eine Wahrscheinlichkeitsdichte?

32) Zeigen Sie, daß die nachstehenden Funktionen eindimensionale Verteilungsfunktionen sind und berechnen Sie die Dichten, Mittelwerte und Varianzen der zugehörigen Verteilungen:

$$F_a(t) = e^{-e^{-(t-a)}} \quad , \quad t \in \mathbf{R}$$

mit Parameter $a \in \mathbf{R}$ und

$$G_\beta(t) = \begin{cases} 1 - e^{-\frac{1}{2}(\frac{t}{\beta})^2} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

mit Parameter $\beta > 0$.

33) Berechnen Sie die Mittelwerte und Varianzen der Verteilungen mit den Dichten aus Aufgabe 30.

34) Berechnen sie die Verteilung P^X in dem Schema

$$(\mathbf{R}, \mathcal{B}, P) \xrightarrow{X} (\mathbf{R}, \mathcal{B}, P^X) ,$$

wenn

a) P die $\mathcal{N}(0,1)$-Verteilung und $X(s) := as + b$ mit $a \neq 0$,

b) P die $\mathcal{E}(\lambda)$-Verteilung, $X(s) = -\log(s)$ für $s > 0$ und $X(s) = 0$ für $s \leq 0$ ist.

35) Berechnen Sie die Verteilungen von

a) $Y_1 = \max(X_1, X_2)$ und

b) $Y_2 = \min(X_1, X_2)$,

wenn die Zufallsvariablen X_1 und X_2 stochastisch unabhängig sind und ihre Verteilungen im Fall a) die Verteilungsfunktionen F_a, F_b und im Fall b) die Verteilungsfunktionen G_α, G_β aus Aufgabe 32 besitzen.

36) Die Zufallsvariablen X_1 und X_2 seien stochastisch unabhängig und ihre Verteilungen mögen die Dichten f_1 bzw. f_2 und die Verteilungsfunktionen F_1 bzw. F_2 besitzen.
Entwickeln Sie eine Formel zur Berechnung der Ereignisse $(X_1 > aX_2)$ mit positiven reellen Zahlen a.

37) Berechnen Sie die Wahrscheinlichkeit des Ereignisses $(X_1 > X_2)$ unter den Voraussetzungen von Aufgabe 36, wenn die Verteilungen der Zufallsvariablen X_1 und X_2

a) Exponentialverteilungen mit Parametern λ_1 bzw. λ_2,

b) Extremwertverteilungen mit Verteilungsfunktionen F_a bzw. F_b (s. Aufg. 32) sind.

38) Berechnen Sie die Marginaldichten zu der Dichte aus Aufgabe 31.

39) X_1 und X_2 seien stochastisch unabhängige Zufallsvariable. X_1 sei $\mathcal{N}(0,1)$-verteilt und X_2 exponentiell verteilt mit Parameter $\frac{1}{2}$. Welche Verteilung besitzt dann die Zufallsvariable

$$Y = \sqrt{2}\frac{X_1}{\sqrt{X_2}} \ ?$$

Hinweis: $\int_0^\infty \sqrt{t}e^{-t}\,dt = \Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$

40) X_1 und X_2 seien stochastisch unabhängig und $\mathcal{U}(0,1)$-verteilt. Berechnen Sie die Verteilungen von

$$Y_1 = \sqrt{-2\log X_1} \cdot \cos(2\pi X_2)$$

und

$$Y_2 = \sqrt{-2\log X_1} \cdot \sin(2\pi X_2).$$

41) Der Zufallsvektor $X = (X_1, X_2)$ sei uniform verteilt auf der Menge

$$M = \left\{(x_1, x_2) \in \mathbf{R}^2 \; ; \; x_1 > 0, x_2 > 0, x_1^2 + x_2^2 < 1\right\}.$$

Berechnen Sie die Verteilung des Zufallsvektors $Y = (Y_1, Y_2)$ mit

$$Y_1 = X_1^2 + X_2^2 \text{ und } Y_2 = \frac{X_1}{X_2}.$$

Sind Y_1 und Y_2 stochastisch unabhängig?

42) Der Zufallsvektor $X = (X_1, X_2)$ sei auf der Menge

$$\left\{(x_1, x_2) \in \mathbf{R}^2 \; ; \; |x_1| + |x_2| \leq 1\right\}$$

uniform verteilt. Berechnen Sie die Erwartungswerte, Varianzen und die Kovarianz der beiden Zufallsvariablen X_1 und X_2.

43) Die Zufallsvariablen X und Y seien stochastisch unabhängig und mit Parametern $\lambda_X = 1$ bzw. $\lambda_Y = 2$ exponentiell verteilt. Berechnen Sie die Kovarianz der Zufallsvariablen $U = 2X + 3Y$ und $V = 3X - Y$.

44) $X_1, \ldots, X_n$ seien stochastisch unabhängige und im Intervall $(-1, 1)$ uniform verteilte Zufallsvariable. Berechnen Sie den Erwartungswert der Zufallsvariablen

$$Y = \max\{X_1, \ldots, X_n\}.$$

45) Ein Eisdielenpächter wird jeden Morgen mit seinem Tagesvorrat an Eis beliefert. Er weiß, daß die Nachfrage nach Eis exponentiell verteilt mit einem Parameter $\lambda > 0$ ist. Einkaufspreis einer Mengeneinheit Eis sei p_1 DM und Verkaufspreis sei p_2 DM. Am Abend wird das nicht verkaufte Eis vernichtet. Wieviele Einheiten muß der Pächter jeden Tag bestellen, damit der zu erwartende Gewinn maximal wird?

46) Der Zufallsvektor $X = (X_1, X_2)$ besitze eine Verteilung mit der Dichte

$$f(x_1, x_2) \; = \; \frac{1}{6\pi} e^{-\frac{1}{18}(5x_1^2 + 2x_1x_2 + 2x_2^2)}$$

Für welche Werte des Parameters a sind die Zufallsvariablen $Y_1 = X_1 - aX_2$ und $Y_2 = 2X_1 + X_2$ stochastisch unabhängig?

14 Literatur

Ulrich Krengel: Einführung in die Wahrscheinlichkeitstheorie und Statistik. vieweg studium, Aufbaukurs Mathematik, 1988.

Rudolf Mathar, Dietmar Pfeifer: Stochastik für Informatiker. Teubner, Stuttgart 1990.

Pierre Brémaud: An Introduction to Probabilistic Modeling. Springer-Verlag, 1988.

William Feller: An Introduction to Probability Theory and its Applications. Vols. I, II, Wiley 1968.

Kai Lai Chung: Elementare Wahrscheinlichkeitstheorie und stochastische Prozesse. Springer Hochschultext, 1978.

Karl Bosch: Elementare Einführung in die Wahrscheinlichkeitsrechnung. Vieweg Basiswissen, 1979.

B. W. Gnedenko: Lehrbuch der Wahrscheinlichkeitsrechnung. Verlag Harri Deutsch, 1978.