1 Wahrscheinlichkeitstheorie, Teil I

1.1 Wahrscheinlichkeitsraum, Zufallsvariablen, charakteristische Funktion und Unabhängigkeit

1.2 Wahrscheinlichkeitsraum und Zufallsvariable

Definition 1.1 *Ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ heißt "Wahrscheinlichkeitsraum", wenn Ω eine Menge, $\mathcal{F}$ eine σ-Algebra von Teilmengen von Ω und $\mathbb{P}$ ein Maß auf $\mathcal{F}$ mit $\mathbb{P}(\Omega) = 1$ ist. Elemente $A \in \mathcal{F}$ werden auch "Ereignisse" genannt. Ein messbare Abbildung $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \in \mathbb{N}$, heißt "Zufallsvariable". Unter dem "Erwartungswert" $\mathbb{E}(X)$ einer Zufallsvariablen versteht man den Wert des Integrals*

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, \mathbb{P}(d\omega)\,.$$

Das Bildmaß von $\mathbb{P}$ unter der Abbildung X wird mit $\mathbb{P}_X$ abgekürzt und heißt "Verteilung von X". Allgemein nennt man ein Maß auf $\mathbb{R}^d$ mit Gesamtmasse Eins eine Verteilung. Für zwei Zufallsvariablen X und Y bezeichnet man mit

$$\mathrm{Cov}(X, Y) = \mathbb{E}\Big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^t\Big)$$

die "Kovarianzmatrix" und benutzt die Abkürzung $\mathrm{Cov}(X) = \mathrm{Cov}(X, X)$. Für eine reellwertige Zufallsvariable X bezeichnet man mit $Var(X) = \mathrm{Cov}(X)$ die "Varianz" von X. Die Größe $\mathbb{E}(X^k)$, $k \in \mathbb{N}$, wird das "k-te Moment" der Zufallsvariablen X bzw. der Verteilung $\mathbb{P}_X$ genannt.

Bemerkung: Wir werden Integratoren in der Form „$d\mathbb{P}$“ bzw. auch als „$\mathbb{P}(d\omega)$“.

Die folgenden Resultate sind der Integrationstheorie entnommen und in der Sprache der Wahrscheinlichkeitstheorie formuliert.

Satz 1.2 *Seien $(\Omega, \mathcal{F}, \mathbb{P})$ wie oben und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann gelten die folgenden Aussagen:*

(1.1) $X \geq 0\ \mathbb{P} -$ *fast sicher* $\quad \Rightarrow \quad \mathbb{P}(X \geq \lambda) \leq \dfrac{1}{\lambda}\, \mathbb{E}(X) \quad \forall\, \lambda > 0\,.$

(1.2) $g : \mathbb{R} \to \mathbb{R}$ *konvex und* $X, g(X) \in L^1(\Omega, \mathbb{P}) \quad \Rightarrow \quad g(\mathbb{E}(X)) \leq \mathbb{E}(g(X))\,.$

(1.3) *Zu jedem Maß* $\mu : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ *mit* $\mu(\mathbb{R}) = 1$ *existiert ein Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbb{P})$ *und eine Zufallsvariable* X *mit* $\mathbb{P}_X = \mu$.

(1.4) $f : \mathbb{R} \to \mathbb{R}$ *beschränkt oder nicht-negativ* $\Rightarrow \mathbb{E}(f(X)) = \int f(x)\, \mathbb{P}_X(dx)$.

(1.5) $X \geq 0\ \mathbb{P} -$ *fast sicher* $\Rightarrow \quad \mathbb{E}(X^p) = \int\limits_0^\infty p\lambda^{p-1}\mathbb{P}(X \geq \lambda)\, d\lambda \qquad \forall\, p \geq 1$.

Bemerkung: Die Gleichung $\mathbb{E}(f(X)) = \int f(x)\, \mathbb{P}_X(dx)$ wird ständig benutzt werden und gilt auch für $\mathbb{R}^d$-wertige Zufallsvariablen X.

Beweis: Die erste Ungleichung ist nach Chebyshev[1] benannt und gilt aufgrund von

$$\mathbb{P}(X \geq \lambda) = \mathbb{E}(\mathbb{1}_{\{X \geq \lambda\}}) \leq \int\limits_{\{X(\omega) \geq \lambda\}} \frac{X(\omega)}{\lambda}\, \mathbb{P}(d\omega) \leq \frac{1}{\lambda}\mathbb{E}(X).$$

Die zweiten Ungleichung heißt wird auch "Jensensche Ungleichung"[2] genannt. Da g konvex ist, gibt es zu $x_0 \in \mathbb{R}^d$ eine Konstante $c > 0$ mit $g\big(X(\omega)\big) \geq g(x_0) + c(X(\omega) - x_0)\ \forall \omega \in \Omega$. Nun wählt man $x_0 = \mathbb{E}(X)$ und bildet rechts und links Erwartungswerte.
Zum Nachweis der dritten Behauptung wählt man $\Omega = [0,1]$ versehen mit dem Lebesgue[3]-Maß $\mathbb{P} = \lambda^1$ und definiert dann $X : \Omega \to \mathbb{R}$ durch

$$X(\omega) = \inf\{t; \mu\big((-\infty, t]\big) \geq \omega\}.$$

Für alle $a \in \mathbb{R}$ gilt dann $P_X((-\infty, a]) = \mathbb{P}(X \leq a) = \mu(-\infty, a]$ und somit folgt die Behauptung.
Die Gleichung $\mathbb{E}(f(X)) = \int f(x)\, \mathbb{P}_X(dx)$ beweist man zunächst für Funktionen des Typs $f(x) = \mathbb{1}_{\mathbb{M}}(x), M \in \mathcal{B}(\mathbb{R}^d)$, und verwendet dann ein Abschließungsargument.
Schließlich gilt, da X nicht-negativ ist,

$$\int\limits_\Omega X(\omega)^p\, \mathbb{P}(d\omega) = \int\limits_0^\infty \int\limits_\Omega \mathbb{1}_{\{\mathbb{X}(\omega)^{\mathbb{p}} \geq \mathbb{t}\}}\, \mathbb{P}(d\omega)\, dt = p\int\limits_0^\infty \lambda^{p-1} \int\limits_\Omega \mathbb{1}_{\{\mathbb{X}(\omega) \geq \lambda\}}\, \mathbb{P}(d\omega)\, d\lambda,$$

woraus auch die letzte Aussage folgt. ∎

[1]Pafnuty Lvovich Chebyshev, 1821 (Borovsk, Russland)-1894 (St. Petersburg)
[2]Johan Ludwig William Valdemar Jensen, 1859 (Nakskov, Dänemark) - 1925 (Kopenhagen)
[3]Henri Léon Lebesgue, 1875 (Beauvais, Picardie/Frankreich) - 1941 (Paris)

Definition 1.3 *Seien* $(\Omega, \mathcal{F}, \mathbb{P})$ *und* X *wie in Definition (1.1). Dann bezeichnet man mit*

$$\sigma(X) = \big\{(X \in A); A \in \mathcal{B}(\mathbb{R})\big\} = \big\{\{\omega; X(\omega) \in A\}; A \in \mathcal{B}(\mathbb{R})\big\}$$

die von X *erzeugte* σ*-Algebra.*

Man überlegt sich leicht, dass $\sigma(X)$ tatsächlich eine $\sigma-$Algebra ist.

1.3 Unabhängigkeit

Ein mächtiges Konzept der Wahrscheinlichkeitstheorie ist das Konzept der Unabhängigkeit. Dieses stellt die erste große Hürde für Mathematiker ohne Vorkenntnisse der Wahrscheinlichkeitstheorie dar. Sei $\mathcal{F}$ eine σ-Algebra. Wir nennen die Ereignisse $A, B \in \mathcal{F}$ von einander "unabhängig", wenn die Wahrscheinlichkeit, dass A eintritt genauso groß ist wie die Wahrscheinlichkeit, dass A unter der zusätzlichen Annahme "B tritt ein" eintritt[4]. Genauer gilt:

Definition 1.4 *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *wie oben. Seien* $A, B \in \mathcal{F}$*. Dann heißt die Größe* $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ *die "bedingte Wahrscheinlichkeit von* A *gegeben* B*". Man nennt* A *und* B *voneinander unabhängig, wenn*

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad \textit{oder hierzu gleichbedeutend} \qquad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

gilt. Zwei σ*-Algebren* $\mathcal{F}$ *und* $\mathcal{B}$ *heißen von einander unabhängig, wenn* A *und* B *von einander unabhängig sind für alle* $A \in \mathcal{F}$ *und alle* $B \in \mathcal{B}$*. Eine Zufallsvariable* X *und eine* $\sigma-$ *Algebra* $\mathcal{F}$ *heißen von einander unabhängig, wenn* $\sigma(X)$ *und* $\mathcal{F}$ *von einander unabhängig sind.*

Beispiel: Sei $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1]^2, \mathcal{B}([0,1]^2), dx \times dy)$. Dann sind die beiden Mengen $A = \{(x,y) \in \Omega, x \leq \frac{1}{2}\}$ und $B = \{(x,y) \in \Omega, \frac{1}{4} \leq x \leq \frac{3}{4}, \frac{1}{4} \leq y \leq \frac{3}{4}, \}$ unabhängig voneinander. Stellen Sie sich vor, dass Sie sich im Wald Ω verlaufen haben. Sie wollen gerne wissen, wie groß die Wahrscheinlichkeit $\mathbb{P}(A)$ ist, dass Sie im Waldgebiet A sind, in dem es keine Bären gibt. Die Zusatzinformation, dass Sie im Waldgebiet B

[4] Auf gut Deutsch: Für die Wahrscheinlichkeit des Ereignisses A ist es schnurzpiepewurscht, ob B eintritt oder nicht.

sind, wo Steinpilze wachsen, hilft Ihnen leider nicht bei der Berechnung von $\mathbb{P}(A)$. Die Ereignisse A und B sind unabhängig.

Beispiel: Sei $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1]^2, \mathcal{B}([0,1]^2), dx \times dy)$. Seien $f, g : \mathbb{R} \to \mathbb{R}$ messbar. Dann sind die Zufallsvariablen $X(\omega) = X((x,y)) = f(x)$ und $Y(\omega) = Y((x,y)) = g(y)$ von einander unabhängig.

Wenn $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X, Y von einander unabhängige Zufallsvariablen und $f, g : \mathbb{R} \to \mathbb{R}$ messbar sind, so sind ebenfalls $f \circ X$ und $g \circ Y$ von einander unabhängig. Wir formulieren diesen Zusammenhang etwas allgemeiner in einem Lemma.

Lemma 1.5 *Seien $X_1, \ldots, X_N : \Omega \to \mathbb{R}^d$ unabhängig und $f_j : \mathbb{R}^{d \times m_j} \to \mathbb{R}$, $j = 1, \ldots, n$ messbare Funktionen mit $m_1 + \ldots + m_n = N$. Wir definieren*

$$Y_1 = (X_1, \ldots, X_{m_1}), \quad Y_2 = (X_{m_1+1}, \ldots, X_{m_1+m_2}), \quad \ldots, Y_n = (X_{N-m_n+1}, \ldots, X_N)\,.$$

Dann sind die Zufallsvariablen $(f_j \circ Y_j)$, $j = 1, \ldots, n$, unabhängig.

Man nennt eine Familie von Ereignissen $\{A_k\}$ unabhängig, wenn für jede endliche Indexmenge $\{k_1, k_2, \ldots, k_N\} \subset \mathbb{N}$ $\mathbb{P}(A_{k_1} \cap A_{k_1} \cap \ldots \cap A_{k_N}) = \prod_{j=1}^{N} \mathbb{P}(A_{k_j})$ gilt. Analoges gilt für Zufallsvariablen. Beachte, dass Unabhängigkeit einer Folge die paarweise Unabhängigkeit impliziert aber nicht umgekehrt.

Formal bedeutet die obige Definition von Unabhängigkeit von Zufallsvariablen folgendes: Die $\mathbb{R}^d$-wertigen Zufallsvariablen $X_1, \ldots, X_N$ sind unabhängig genau dann, wenn

$$\mathbb{P}_{(X_1,\ldots,X_N)} = \prod_{j=1}^{N} \mathbb{P}_{X_j} \quad \text{gilt}\,. \tag{1.6}$$

Hierbei bezeichnet $\mathbb{P}_{(X_1,\ldots,X_N)}$ die Verteilung der $(\mathbb{R}^d)^N$-wertigen Zufallsvariablen $(X_1, \ldots, X_N)$ und $\prod_{j=1}^{N} \mathbb{P}_{X_j}$ das Produktmaß. Seien $B_1, \ldots, B_N \in \mathcal{B}(\mathbb{R}^d)$. Dann folgt die behauptete Äquivalenz aus den beiden Beobachtungen

$$\mathbb{P}_{(X_1,\ldots,X_N)}(B_1 \times \ldots \times B_N) = \mathbb{P}\big(X^{-1}(B_1) \cap \ldots \cap X^{-N}(B_N)\big)\,,$$
$$\Big(\prod_{j=1}^{N} \mathbb{P}_{X_j}\Big)(B_1 \times \ldots \times B_N) = \prod_{j=1}^{N} \big(\mathbb{P}(X_j^{-1}(B_j))\big)\,.$$

Um eine weitere wichtige Eigenschaft von unabhängigen Zufallsvariablen angeben zu können, erinnern wir kurz an die Faltung von Maßen. Seien μ und ν Maße auf $\mathbb{R}^d$. Dann definiert die Faltung

$$(\mu * \nu)(A) = \int\limits_{\mathbb{R}^d} \nu(A-y)\,\mu(dy) = \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} \mathbb{1}_A(x+y)\nu(dx)\,\mu(dy), \quad A \in \mathcal{B}(\mathbb{R}^d)$$

ein weiteres Maß auf $\mathbb{R}^d$. Diese Definition verträgt sich mit der bekannten Definition der Faltung zweier Funktionen f und g im folgenden Sinne. Wenn μ und ν Dichten f und g bezüglich des Lebesgue-Maßes haben, so hat das Maß $\mu * \nu$ die Dichte $h = f * g$, denn, wie man leicht nachrechnet, es gilt für beschränkte Funktionen $h : \mathbb{R}^d \to \mathbb{R}$

$$\int\limits_{\mathbb{R}^d} h(z)\,(\mu * \nu)(dz) = \int\limits_{\mathbb{R}^d} h(z)(f * g)(z)\,dz\,.$$

Lemma 1.6 *Für unabhängige Zufallsvariablen $X_1, \ldots, X_N$ gilt*

$$\mathbb{P}_{X_1} * \mathbb{P}_{X_2} * \ldots * \mathbb{P}_{X_N} = \mathbb{P}_{X_1+X_2+\ldots+X_N}\,.$$

Beweis: *Für $M \in \mathcal{B}(\mathbb{R}^d)$ gilt*

$$\begin{aligned}
\big(\mathbb{P}_{X_1} * \mathbb{P}_{X_2} * \ldots * \mathbb{P}_{X_N}\big)(M) &= \int \ldots \int \mathbb{1}_M(x_1 + x_2 + \ldots + x_N)\,\mathbb{P}_{X_1} \ldots \mathbb{P}_{X_N} \\
&= \int \ldots \int \mathbb{1}_M(x_1 + x_2 + \ldots + x_N)\,\mathbb{P}_{(X_1,\ldots,X_N)} \\
&= \mathbb{E}\big(\mathbb{1}_M(X_1 + \ldots + X_N)\big) = \mathbb{P}\big(X_1 + \ldots + X_N \in M\big) \\
&= \mathbb{P}_{X_1+X_2+\ldots+X_N}(M)\,.
\end{aligned}$$

■

Für zwei Funktionen $f, g : \Omega = (0,1) \to \mathbb{R}$ gilt die Aussage $f \in L^1(\Omega)$, $g \in L^1(\Omega)$ $\Rightarrow (fg) \in L^1(\Omega)$ natürlich im Allgemeinen nicht. Für unabhängige Zufallsvariablen ist ein solcher Schluss hingegen zulässig und charakteristisch.

Lemma 1.7 *Für unabhängige reellwertige Zufallsvariablen $X_1, \ldots, X_N$ mit $X_i \in L^1(\Omega)\,\forall i$ gelten*

$$\prod_{i=1}^N |X_i| \in L^1(\Omega) \quad \textit{und} \quad \mathbb{E}\Big(\prod_{i=1}^N X_i\Big) = \prod_{i=1}^N \mathbb{E}(X_i)\,.$$

Beweis: Es sei eine Funktion $f : \mathbb{R}^N \to \mathbb{R}$ durch $f(x_1, \ldots, x_N) = |x_1| \cdot \ldots \cdot |x_n|$ definiert. Mit Hilfe des Satzes von Fubini[5] schließen wir dann

$$\mathbb{E}\Big(\prod_{i=1}^{N} |X_i|\Big) = \int_{\mathbb{R}^n} f \, \mathbb{P}_{(X_1,\ldots,X_N)} = \prod_{i=1}^{N} \Big(\int_{\mathbb{R}} |x_i| \mathbb{P}_{X_i}(dx_i)\Big) = \prod_{i=1}^{N} \mathbb{E}(|X_i|)\,.$$

Also gilt $\prod_{i=1}^N |X_i| \in L^1(\Omega)$. Die zweite Aussage beweist man, indem man den obigen Beweis wiederholt, nun aber die Betragsstriche fortlässt. ■

Die eben bewiesene Aussage sagt viel über die Struktur unabhängiger Zufallsvariablen aus. Seien X und Y reellwertig und unabhängig mit $\mathbb{E}(X) = 0$. Dann gilt für alle $f : \mathbb{R}^d \to \mathbb{R}$ mit $f \circ Y \in L^1(\Omega)$

$$\mathbb{E}(\big(X(f \circ Y)\big) = \mathbb{E}(X)\mathbb{E}(f \circ Y) = 0\,.$$

Das bedeutet, dass $f \circ Y$ im L^2-Sinn orthogonal zu X ist für alle zugelassenen f. Es ist also gar nicht so einfach, eine Folge unabhängiger Zufallsvariablen anzugeben.

Auf dem Wahrscheinlichkeitsraum $([0,1], \mathcal{B}([0,1]), \lambda^1)$ ist durch

$$A_0 = [0,1), \quad A_1 = [0, \tfrac{1}{2}), \quad A_2 = [0, \tfrac{1}{4}) \cup [\tfrac{2}{4}, \tfrac{3}{4}), \quad \ldots, \quad A_n = \bigcup_{k=0}^{(2^{n-1}-1)} \Big[\frac{2k}{2^n}, \frac{2k+1}{2^n}\Big)$$

eine Folge (A_k) unabhängiger Ereignisse definiert. Die Folge $(\mathbb{1}_{A_k})$ ist eine Folge unabhängiger Zufallsvariaben. Sehr häufig konstruiert man Folgen unabhängiger Zufallsvariablen, indem man Zufallsvariablen $X_j : \Omega_j \to \mathbb{R}^d$ betrachtet und dann die Folge auf dem unendlichen Poduktraum $\Omega = \Omega_1 \times \Omega_2 \times \ldots$ definiert. Die erforderliche Orthogonalität gilt dann automatisch.

Lemma 1.8 *Für unabhängige reellwertige Zufallsvariablen $X_1, \ldots, X_N$ mit $X_i \in L^2(\Omega)\,\forall i$ gilt*

$$\mathrm{Var}\,\Big(\sum_{j=1}^{N} X_j\Big) = \sum_{j=1}^{N} \mathrm{Var}(X_j)\,.$$

Beweis: Setze $Y_j = X_j - \mathbb{E}(X_j)$. Da dann die Y_j unabhängig sind und $\mathbb{E}(Y_j) = 0\;\forall\, j$ gilt, folgt mit obigem Lemma $\mathbb{E}(Y_j Y_k) = 0$ falls $j \neq k$. Somit können wir schließen

$$\mathrm{Var}(X_1 + \ldots + X_N) = \mathbb{E}\big((Y_1 + \ldots + Y_N)^2\big) = \sum_{j,k=1}^{N} \mathbb{E}(Y_j Y_k) = \sum_{j=1}^{N} \mathbb{E}(Y_j^2)\,,$$

[5] Guido Fubini, 1879 (Venedig) - 1943 (New York City)

woraus die Behauptung folgt. ■

Wir beweisen nun ein Hilfsresultat, das sich in vielen Situationen als sehr nützlich erweisen wird.

Definition 1.9 *Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und (A_k) eine Folge von Ereignissen. Dann bezeichnet man mit*

$$"A_k \text{ i.o.}" = \limsup_{k\to\infty} A_k = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j = \{\omega; \omega \in A_k \text{ für unendlich viele } k\}$$

das sogenannte "terminale Ereignis".

Lemma 1.10 (Borel[6]-Cantelli[7]) *Sei (A_k) eine Folge von Ereignissen in $(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt*

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty \quad \Longrightarrow \quad \mathbb{P}("A_k \text{ i.o.}") = 0 \,. \tag{1.7}$$

Ist die Folge (A_k) unabhängig, so gilt auch die Umkehrung, d.h.

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty \quad \Longrightarrow \quad \mathbb{P}("A_k \text{ i.o.}") = 1 \,. \tag{1.8}$$

Dass (1.8) wirklich die Umkehrung von (1.7) ist, wird durch Lemma 1.25 klar.

Beweis: Die Folge $B_k = \bigcup\limits_{j=k}^{\infty} A_j$ ist absteigend. Da $\mathbb{P}$ ein Maß ist, folgt nach Voraussetzung für jedes k

$$\mathbb{P}("A_k \text{ i.o.}") = \mathbb{P}\big(\bigcap_{l=1}^{\infty} B_l\big) \le \mathbb{P}\big(\bigcup_{j=k}^{\infty} A_j\big) \le \sum_{j=k}^{\infty} \mathbb{P}(A_j)\,,$$

womit (1.7) folgt wegen $\sum\limits_{j=k}^{\infty} \mathbb{P}(A_j) \to 0$ für $k \to \infty$. Sei nun die Folge (A_k) unabhängig [8]. Unter Verwendung der Voraussetzung und mehrfacher Komplementbil-

[7] Félix Édouard Justin Émile Borel, 1871 (Saint-Affrique)-1956 (Paris), Mitbegründer der modernen Maßtheorie.

[7] Francesco Paolo Cantelli, 1875 (Palermo) - 1966 (Rom).

[8] Man kann die Annahme mit einem etwas aufwendigeren Beweis darauf reduzieren, dass die A_k nur paarweise unabhängig sind.

dung folgt

$$\mathbb{P}\Big(\bigcup_{k\geq n} A_k\Big) = 1 - \mathbb{P}\Big(\bigcap_{k\geq n}(A_k)^c\Big) = 1 - \prod_{k\geq n}\mathbb{P}\big((A_k)^c\big) = 1 - \prod_{k\geq n}\big(1-\mathbb{P}(A_k)\big)$$
$$\geq 1 - \prod_{k\geq n} e^{-\mathbb{P}(A_k)} = 1 - e^{-\sum\limits_{k\geq n}\mathbb{P}(A_k)} = 1\,,$$

was den Beweis abschließt. ■

1.4 Die charakteristische Funktion

Definition 1.11 *Sei $\mathbb{P}_X$ die Verteilung einer Zufallsvariablen $X:\Omega\to\mathbb{R}^d$. Dann heißt die wie folgt definierte Abbildung $\varphi_X:\mathbb{R}^d\to\mathbb{C}$*

$$\varphi_X(X) = \mathbb{E}\big(e^{i\langle y,X\rangle}\big) = \int\limits_{\mathbb{R}^d} e^{i\langle y,x\rangle}\mathbb{P}_X(dx)$$

die "charakteristische Funktion" der Zufallsvariablen X bzw. der Verteilung $\mathbb{P}_X$.

Die obige Definition führt u.a. dazu, dass man Funktionen $\Omega\to\mathbb{R}$, $\omega\mapsto\mathbb{1}_M(\omega)$ für $M\subset\Omega$ nicht, wie in der Analysis üblich, als charakteristische Funktion sondern eher als "Indikatorfunktion" bezeichnet. Wir nennen solche Funktionen auch "Stufenfunktion", da Funktionen der Form $\omega\mapsto\sum\limits_{j=1}^{N} c_j\mathbb{1}_{M_j}(\omega)$, $c_j\in\mathbb{R}, M_j\subset\Omega$, häufig "Treppenfunktionen" heißen.

Lemma 1.12 *Seien $X_1,\ldots,X_N:\Omega\to\mathbb{R}^d$ unabhängige Zufallsvariablen und $S_N = \sum\limits_{j=1}^{N} X_j$. Dann gilt*

$$\varphi_{S_N}(y) = \prod_{j=1}^{N}\varphi_{X_j}(y) \qquad \forall\, y\in\mathbb{R}^d\,.$$

Bemerkung: Insbesondere gilt für unabhängige, identisch verteilte Zufallsvariablen $X_1,\ldots,X_N:\Omega\to\mathbb{R}^d$ die Gleichung $\varphi_{S_N}(y) = \Big(\varphi_{X_1}(y)\Big)^N$ $\forall\, y\in\mathbb{R}^d$.

Beweis: Da $\varphi_{S_N}(y) = \mathbb{E}\big(\exp(i\langle y, S_N\rangle)\big) = \mathbb{E}\big(\prod_{j=1}^{N} \exp(i\langle y, X_j\rangle)\big)$ gilt, folgt die Aussage aufgrund der Unabhängigkeit. ■

Lemma 1.13 *Sei $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable mit $\mathbb{E}(|X|^k) < \infty$ für $k \in \mathbb{N}$. Dann gilt $\varphi_X \in C^k(\mathbb{R})$ mit*

$$\frac{d^j \varphi_X}{dy^j}(y) = \int_{\mathbb{R}} (ix)^j e^{ixy}\, \mathbb{P}_X(dx) \quad \forall\, y \in \mathbb{R}, \quad j = 0, 1, \ldots, k\,. \tag{1.9}$$

Beweis: Wir beweisen die Aussage zuerst für $k = 1$. Zunächst zeigt man, dass für alle $y \in \mathbb{R}$ der Grenzwert $\lim_{h\to 0} \frac{\varphi_X(y+h)-\varphi_X(y)}{|h|}$ existiert. Er existiert wegen

$$\begin{aligned} \Big|\frac{\varphi_X(y+h) - \varphi_X(y)}{|h|}\Big| &= |h|^{-1}\Big|\int_{\mathbb{R}} e^{ixy}\Big(e^{ixh} - 1\Big)\, \mathbb{P}_X(dx)\Big| \\ &\le |h|^{-1}\Big|\int_{\mathbb{R}} |e^{ixy}||xh|\, \mathbb{P}_X(dx)\Big| \end{aligned}$$

und dem Satz von der majorisierten Konvergenz. Setzt man $(\varphi_X)'(y) = \int_{\mathbb{R}} (ix)e^{ixy}\, \mathbb{P}_X(dx)$ so sieht man, dass für alle $y \in \mathbb{R}$ $\lim_{h\to 0} \frac{\varphi_X(y+h)-\varphi_X(y)}{|h|} = (\varphi_X)'(y)$ gilt. Die Funktion $y \mapsto (\varphi_X)'(y)$ ist zudem stetig, was man durch eine weitere Anwendung des Satzes der majorisierten Konvergenz beweist. Den gesamten Schluss iteriert man nun und beweist so $\varphi_X \in C^k(\mathbb{R})$ und Gleichung (1.9) für $k \in \mathbb{N}$. ■

1.5 Gaußsche Zufallsvariablen

Definition 1.14 *Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt "normal verteilt" oder "Gaußsche Zufallsvariable"* [9]*, wenn für alle $M \in \mathcal{B}(\mathbb{R})$*

$$\mathbb{P}(X \in M) = \mathbb{P}_X(M) = \int_M \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(y-m)^2}{2\sigma^2}}\Big)\, dy$$

mit $m, \sigma \in \mathbb{R}$ und $\sigma > 0$ gilt. Die Verteilung $\mathbb{P}_X$ wird in diesem Fall oft mit $\mathcal{N}(m, \sigma^2)$ bezeichnet und man schreibt: "X ist $\mathcal{N}(m, \sigma^2)$-verteilt" bzw. "$X \sim \mathcal{N}(m, \sigma^2)$". Im Fall $m = 0, \sigma = 1$ nennt man X "standardisiert normal verteilt".

Dass die obige Definitionen sinnvoll sind, wird durch folgendes Lemma erläutert.

Lemma 1.15 *Seien* $m, \sigma \in \mathbb{R}$ *mit* $\sigma > 0$. *Sei* $f : \mathbb{R} \to \mathbb{R}$ *durch* $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-m)^2}{2\sigma^2}}$ *definiert. Sei* $X \sim \mathcal{N}(m, \sigma^2)$. *Dann gilt*

$$\int_{\mathbb{R}} f(x)\, dx = 1\,,$$

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\, \mathcal{N}(m, \sigma^2)(dx) = m\,,$$

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x-m)^2\, \mathcal{N}(m, \sigma^2)(dx) = \sigma^2\,,$$

Der Beweis der drei Behauptungen beruht auf Standardtechniken wie der Substitution von Variablen und der Transformation auf Polarkoordinaten.

Lemma 1.16 *Sei* $X \sim \mathcal{N}(0,1)$. *Dann gilt für alle* $x > 0$

$$\frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \le \mathbb{P}(X > x) \le \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\,.$$

Beweis: Wir beweisen zunächst die Abschätzung nach oben. Es gilt

$$\mathbb{P}(X > x) \le \int_x^\infty \frac{s}{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-s^2}{2}}\, ds = \frac{1}{x} \frac{1}{\sqrt{2\pi}} \Big[- e^{\frac{-s^2}{2}} \Big]_{s=x}^\infty\,,$$

woraus die gewünschte Abschätzung folgt. Sei nun $g : [0, \infty) \to \mathbb{R}$ definiert durch

$$g(x) = xe^{-x^2/2} - (x^2+1) \int_x^\infty e^{-s^2/2}\, ds\,.$$

Die Abschätzung nach unten ist bewiesen, wenn man gezeigt hat, dass $g(x) \le 0\, \forall\, x \ge 0$ gilt. Zunächst gelten $g(0) < 0$ und $\lim_{x\to\infty} g(x) = 0$. Ausreichend ist demnach zu zeigen $g'(x) \ge 0\ \forall x \ge 0$. Es gilt

$$g'(x) = 2e^{-x^2/2} - 2x \int_x^\infty e^{-s^2/2}\, ds = 2x \Big(\frac{1}{x} e^{-x^2/2} - \int_x^\infty e^{-s^2/2}\, ds \Big)\,,$$

[9] Carl Friedrich Gauß, 1777 (Braunschweig) - 1855 (Göttingen), Mathematischer Universalgelehrter

woraus die gewünschte Ungleichung folgt, wenn man die bereits bewiesene Abschätzung nach oben benutzt. ■

Lemma 1.17 *Für die charakteristische Funktion φ_X einer $\mathcal{N}(m, \sigma^2)$-verteilten Zufallsvariablen X gilt*

$$\varphi_X(y) = e^{imy - \frac{\sigma^2 y^2}{2}} .$$

Wir betrachten nun mehrdimensionale Gaußsche Zufallsvariablen.

Definition 1.18 *Eine Zufallsvariable $X : \Omega \to \mathbb{R}^d$ heißt "standardisiert normal verteilt" bzw. "standardisierte Gaußsche Zufallsvariable", wenn unabhängige Zufallsvariablen $X_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, d$, existieren mit $X(\omega) = (X_1(\omega), \ldots, X_d(\omega))^t$ $\forall\, \omega \in \Omega$. Eine Zufallsvariable $X : \Omega \to \mathbb{R}^d$ heißt "normal verteilt" bzw. "Gaußsche Zufallsvariable", wenn es einen Vektor $m \in \mathbb{R}^d$, eine $d \times n$-Matrix A, $n \in \mathbb{N}$, und eine standardisiert normal verteilte Zufallsvariable $Y : \Omega \to \mathbb{R}^n$ gibt mit*

$$X(\omega) = A(Y(\omega)) + m \qquad \forall\, \omega \in \Omega\,.$$

Ist $X : \Omega \to \mathbb{R}^d$ eine Gaußsche Zufallsvariable mit $\mathrm{Cov}(X) = S$ und $\mathbb{E}(X) = m$, so schreibt man $X \sim \mathcal{N}(m, S)$, wobei die Verteilung $\mathcal{N}(m, S)$ das Bildmaß auf $\mathbb{R}^d$ bezeichnet. Im Fall der obigen Definition gelten $\mathrm{Cov}(X) = AA^t$ und $\mathbb{E}(X) = m$. Die $d \times d$-Einheitsmatrix kürzen wir mit I_d ab und schreiben $X \sim \mathcal{N}(0, I_d)$ an Stelle von $X \sim \mathcal{N}(0, S)$ im Falle $S = I_d$. Wie wir weiter unten sehen sind $\mathcal{N}(0, I_d)$ verteilte Zufallsvariablen genau die standardisiert normal verteilte Zufallsvariablen $X : \Omega \to \mathbb{R}^d$.

Lemma 1.19 *Seien Θ eine orthogonale $d \times d$-Matrix und $X \sim \mathcal{N}(0, I_d)$. Dann gilt auch $\Theta X \sim \mathcal{N}(0, I_d)$.*

Beweis: Zunächst prüft man nach, dass die Dichte von $\mathbb{P}_X = \mathcal{N}(0, I_d)$ durch $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{|x|^2}{2}}$ gegeben ist. Hierbei verwendet man die Unabhängigkeit der einzelnen Komponenten von X. Mit Hilfe der Variablensubstitution $y = \Theta^{-1} x$ sieht man, dass die Dichte von $\mathbb{P}_{\Theta X}$ ebenfalls durch f gegeben ist. Die Aussage ist damit bewiesen. ■

Korollar 1.20 *Seien $X \sim \mathcal{N}(0, \sigma^2)$, $Y \sim \mathcal{N}(0, \sigma^2)$ unabhängig. Dann sind die beiden Zufallsvariablen $X+Y$ und $X-Y$ unabhängig mit $\mathbb{P}_{X+Y} = \mathbb{P}_{X-Y} = \mathcal{N}(0, 2\sigma^2)$.*

Beweis: Die Matrix $M = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ definiert eine orthogonale Abbildung $\Theta : \mathbb{R}^2 \to \mathbb{R}^2$ mit $\Theta\big((x, y)^t\big) = \frac{1}{\sqrt{2}}(x + y, x - y)^t$. Die Behauptung folgt mit Lemma 1.19. ∎

Das folgende Lemma rechtfertigt die Einschränkung der Notation $\mathcal{N}(m, S)$ auf zwei Parameter. Diese beiden reichen nämlich aus, um normal verteilte Zufallsvariablen festzulegen.

Lemma 1.21 *Seien $X, Y : \Omega \to \mathbb{R}^d$ normal verteilte Zufallsvariablen mit $\mathbb{E}(X) = \mathbb{E}(Y)$ und $\mathrm{Cov}(X) = \mathrm{Cov}(Y)$. Dann gilt $\mathbb{P}_X = \mathbb{P}_Y$.*

Beweis: Es ist ausreichend, den Fall $\mathbb{E}(X) = \mathbb{E}(Y)$ zu betrachten. Es existieren also eine Zufallsvariable $\widetilde{X} : \Omega \to \mathbb{R}^{n_1}$, eine $d \times n_1$-Matrix A, eine Zufallsvariable $\widetilde{Y} : \Omega \to \mathbb{R}^{n_2}$ und eine $d \times n_2$-Matrix B mit

$$X(\omega) = A(\widetilde{X}(\omega)), \quad Y(\omega) = B(\widetilde{Y}(\omega)) \qquad \forall\, \omega \in \Omega\,.$$

Sei $k = \max\{n_1, n_2\}$. Durch Hinzufügen von Nullen können wir annehmen, dass die Matrizen A, B beide k Spalten und d Zeilen haben und die $\widetilde{X}$ sowie $\widetilde{Y}$ Werte in $\mathbb{R}^k$ annehmen. Seien nun V_A und V_B die Vektorräume, welche von den Zeilenvektoren A_j, B_j, $j = 1, \ldots, k$, der Matrizen A und B aufgespannt werden. Wir wählen $l \in \mathbb{N}$ so, dass die Vektoren $A_1, A_2, \ldots, A_l$ eine Basis von V_A bilden. Wir definieren nun eine Abbildung $\Theta : V_A \to V_B$ durch $\Theta(A_j) = B_j$, $j = 1, \ldots, l$. Wir wollen zeigen, dass Θ ein Isomorphismus ist. Zunächst zeigen wir, dass Θ injektiv ist. Sei $v \in V_A$ mit $\Theta(v) = 0$, d.h. es gilt

$$\Theta(v_1 A_1 + \ldots + v_l A_l) = 0\,.$$

Sei nun $\widetilde{v}$ definiert durch $\widetilde{v}_j = v_j$ für $j = 1, \ldots, l$ und $\widetilde{v}_j = 0$ für $j = l + 1, \ldots, k$. Dann bedeutet die obige Zeile $\widetilde{v}^t B = 0$, und es folgt

$$|\widetilde{v}^t A|^2 = \widetilde{v}^t A A^t \widetilde{v}^t = \widetilde{v}^t B B^t \widetilde{v}^t = 0\,.$$

Also gilt $\widetilde{v}^t A = 0$, woraus wir $v = 0$ schließen. Die Abbildung Θ ist also injektiv, und es gilt $\dim V_A \leq \dim V_B$. Wenn wir die Rollen von A und B anfangs vertauschen,

erhalten wir $\dim V_A \geq \dim V_B$, also ist Θ ein Isomorphismus.
Für alle $v, w \in V_A$ gilt außerdem $\langle v, w\rangle = \langle A^t v, A^t w\rangle = \langle B^t v, B^t w\rangle = \langle \Theta(v)^t, \Theta(w)^t\rangle$, also ist Θ auch eine orthogonale Abbildung von V_A nach V_B. Man setzt nun Θ wie üblich fort zu einer orthogonalen Abbildung $\widetilde{\Theta}$ von $\mathbb{R}^k$ nach $\mathbb{R}^k$, indem man das orthogonale Komplement von V_A orthogonal auf das orthogonale Komplement von V_B abbildet. Damit gilt nun

$$X(\omega) = A(\widetilde{X}(\omega)), \quad Y(\omega) = B(\widetilde{Y}(\omega)) = A(\Theta(\widetilde{Y}(\omega)) \qquad \forall\, \omega \in \Omega\,.$$

und die Aussage folgt nach Lemma 1.19, da $\widetilde{X} \sim \mathcal{N}(0, I_k)$ und $\Theta(\widetilde{Y}) \sim \mathcal{N}(0, I_k)$. ■

Mit Hilfe von Lemma 1.21 können wir $\mathcal{N}(0, I_d)$-verteilte Zufallsvariablen mit standardisierten Gaußschen Zufallsvariablen identifizieren.

Korollar 1.22 *Sei $X : \Omega \to \mathbb{R}^d$ $\mathcal{N}(m, S)$-verteilt. Dann sind die folgenden Aussagen äquivalent:*

(i) Die Komponenten $X_1, \ldots, X_d$ sind unabhängig.

(ii) S ist eine Diagonalmatrix.

(iii) Die Komponenten $X_1, \ldots, X_d$ sind paarweise unkorreliert, d.h. es gilt $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$ für $1 \neq j$.

Beweis: Die Schlüsse (iii)→(ii) und (i)→(iii) gelten klarerweise. Der einzige beweisbedürftige Schluss ist (ii)→(i). Da man mit Hilfe von Lemma 1.21 [10] eine Zufallsvariable $Y : \Omega \to \mathbb{R}^d$ mit $Y \sim \mathcal{N}(m, S)$ und unabhängigen Komponenten $Y_1, \ldots, Y_d$ konstruieren kann, muss ebenfalls X unabhängige Komponenten haben. Denn zwei vektorwertige Zufallsvariablen mit derselben Verteilung haben entweder beide unabhängige Komponenten oder beide nicht. ■

[10] Seien $s_j^2; j = 1, \ldots, d$, die Einträge von S auf der Diagonalen. Sei $\widetilde{Y} : \Omega \to \mathbb{R}^d$ Gaußsch mit unabhängiggen Komponenten $\widetilde{Y}_1, \ldots, \widetilde{Y}_d$ und $\mathbb{E}(\widetilde{Y}_j) = \frac{1}{s_j}\mathbb{E}(X_j)$, $\mathrm{Var}(\widetilde{Y}) = I_d$. Dann gilt für die Zufallsvariable $Y = (s_1\widetilde{Y}_1, \ldots, s_d\widetilde{Y}_d)$ einerseits $\mathbb{E}(Y) = \mathbb{E}(X)$ und andererseits $\mathrm{Cov}(Y) = \mathrm{Cov}(X)$. Da Y Gaußsch ist, gilt nach Lemma 1.21, dass Y und X dieselbe Verteilung besitzen.

1.6 Stochastische Prozesse

Definition 1.23 *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei $I = [0, \infty), I = [0, b)$, $I = [0, b]$ mit $b \in \mathbb{R}_+$ oder $I = \mathbb{N}$. Eine Familie von Zufallsvariablen $\{X_t; t \in I, X_t : \Omega \to \mathbb{R}^d\}$ nennt man "stochastischer Prozess". Wir schreiben schlicht X [11] für einen stochastischen Prozess. Die Abbildung $t \mapsto X_t(\omega)$ heißt Pfad[12] , und in diesem Sinne kann man Ω als Raum der Pfade auffassen.*

Seien X und Y zwei stochastische Prozesse auf ein und demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Oftmals ist die Frage wichtig, wann man X und Y miteinander identifizieren kann bzw. wann und in welchem Sinne sie gleich sind. Gerade wenn man die Theorie des Lebesgue-Integrals verinnerlicht hat, ist man geneigt, die Bedeutung dieser Frage bei stochastischen Prozessen zu unterschätzen.

Definition 1.24 (Identifikation von stochastischen Prozessen) .
1) Wenn für alle $N \in \mathbb{N}$ und reelle Zahlen $0 \le t_1 < t_2 < t_3 \ldots < t_N$, $t_k \in I$ und alle Mengen $A \in \mathcal{B}(\mathbb{R}^d)$ gilt

$$\mathbb{P}\big((X_{t_1}, X_{t_2}, \ldots, X_{t_N}) \in A\big) = \mathbb{P}\big((Y_{t_1}, Y_{t_2}, \ldots, Y_{t_N}) \in A\big)\,,$$

so sagt man: "X und Y haben dieselben endlich-dimensionalen Verteilungen".
2) X und Y heißen "Modifikationen" voneinander, wenn für alle $t \in I$

$$\mathbb{P}(X_t = Y_t) = 1\,.$$

3) X und Y heißen "ununterscheidbar", falls

$$\mathbb{P}(X_t = Y_t \;\forall\, t \in I) = 1\,.$$

Wenn X und Y ununterscheidbar sind, so sind sie trivialerweise Modifikationen voneinander. Wenn sie Modifikationen voneinander sind, so haben sie dieselben endlich-dimensionalen Verteilungen. Folgendes Beispiel illustiert, dass letztere Eigenschaft nicht die anderen impliziert.

Sei X ein stochastischer Prozess auf $(\Omega, \mathcal{F}, \mathbb{P})$ und dem Zeitintervall $I = [0, 1]$. Sei $T : \Omega \to [0, 1]$ eine Zufallsvariable. Wir stellen uns T als Wecker vor, der jedem

[11] Die Schreibweise X_t für einen stochastischen Prozess ist weit verbreitet und genauso gut bzw. schlecht wie die Verwendung von $f(x)$ für eine Funktion $f : \mathbb{R} \to \mathbb{R}$.

[12] Man kann auch den Graphen $\{(t, X_t(\omega)); t \ge 0\}$ als Pfad bezeichnen.

Pfad einen Uhrzeit zuweist. Definiere nun einen neuen stochastischen Prozess Y auf $(\Omega, \mathcal{F}, \mathbb{P})$ durch

$$Y_t(\omega) = \begin{cases} X_t(\omega) & , t \neq T(\omega)\,, \\ 0 & , t = T(\omega)\,. \end{cases}$$

Dann besitzen X und Y dieselben endlich-dimensionalen Verteilungen, sind aber nicht ununterscheidbar. Wenn fast alle Pfade von X stetig und nicht trivial sind, so sind fast alle Pfade von Y unstetig. So etwas kann bei ununterscheidbaren Prozessen natürlich nicht auftreten. Wenn man im obigen Beispiel annimmt, dass T eine Verteilung mit einer stetigen Dichte hat, so sind X und Y Modifikationen, denn $\mathbb{P}(X_t = Y_t) = \mathbb{P}(T \neq t) = 1\ \forall\, t$.

1.7 Ein Beispiel

1.A Anhang 1: Weitere Resultate

Eine Richtung in Lemma 1.10 wird noch besser verständlich, wenn man folgendes grundlegendes Resultat kennt.

Lemma 1.25 (Null-Eins-Gesetz von É. Borel) *Sei (A_k) eine Folge paarweise unabhängiger Ereignisse in $\mathcal{F}$. Dann gilt*

$$\mathbb{P}(A_k \text{ i.o.}) = 1 \qquad \text{oder} \qquad \mathbb{P}(A_k \text{ i.o.}) = 0\,.$$

In späteren Versionen zu ergänzen: Eindeutigkeit der charakteristischen Funktion.

1.B Anhang 2: Folgen unabhängiger Zufallsvariablen und Grenzwertsätze

In späteren Versionen zu ergänzen: Sätze über die Konvergenz von Verteilungen, siehe das Buch von Dudley.

Lemma 1.26 (Lévy) *Sei $(X_n)_n$ eine Folge unabhängiger Zufallsvariablen. Dann sind die folgen Aussagen äquivalent:*

1. $\sum_{n=1}^{\infty} X_n$ *konvergiert* $\mathbb{P}$*-fast sicher.*

2. $\sum_{n=1}^{\infty} X_n$ *konvergiert dem Maß nach.*

3. $\sum_{n=1}^{\infty} X_n$ *konvergiert der Verteilung nach, d.h. die Verteilung* $\mu_k = \mathbb{P}_{\sum_{n=1}^{k} X_n}$ *konvergiert gegen ein Maß* μ *für* $k \to \infty$*.*