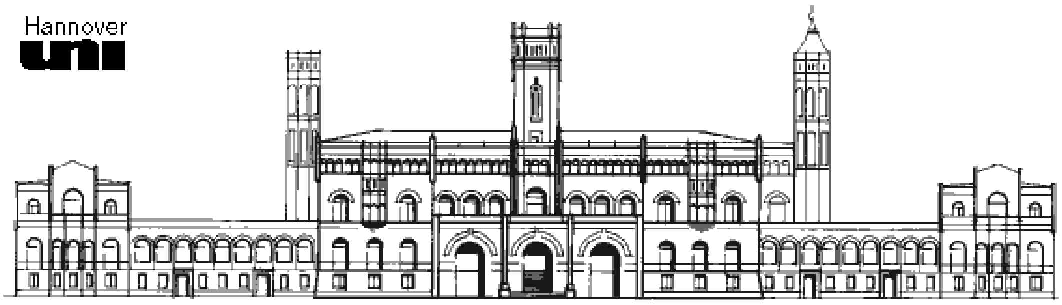


Vorlesungsmitschrift:

“Stochastik II”

Prof. Dr. R. Grübel

Institut für Mathematische Stochastik, Hannover

(WS 96/97)

24. November 1997

ge-TEX-ed von Sven Meyer
(email: `sm@stochastik.uni-hannover.de`)
(WWW: `www.stochastik.uni-hannover.de/~sm`)

Für diejenigen, die sich die Mühe machen, dieses Skript zu lesen und dabei auf Tippfehler stoßen: Bitte schickt mir, nachdem Ihr auf meiner Homepage nachgesehen und Euch vergewissert habt, daß dieser Fehler noch immer existiert, eine email, damit ich ihn berichtigen kann.

Mit bestem Dank, Sven.

Inhaltsverzeichnis

. . . und mit einem großen Dankeschön für die Mitarbeit an:

- Britta Kersten
- Luise Regier
- Astrid Vieweg

Kapitel 8

Integral und Erwartungswert

8.1 Einführung und Erinnerung

$(\Omega, \mathfrak{A}, P)$ WRaum, $X : \Omega \to \mathbb{R}$ eine Zufallsvariable (ZV) (d.h. X ist $(\mathfrak{A}, \mathfrak{B})$-meßbar).
EX war bisher nur für diskrete ZV definiert.
Allgemeine Situation wird mit dem Lebesgue-Integral behandelt.
Es sei $(\Omega, \mathfrak{A})$ ein meßbarer Raum und $\mu : \mathfrak{A} \to [0, \infty]$ ein Maß, d.h.

$\mu(\emptyset) = 0$

$\mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$, $A_1, A_2, \ldots \in \mathfrak{A}$ disjunkt (σ-Additivität).

Beachte: Wert ∞ ist zugelassen.

Rechenregeln: $a + \infty = \infty + a = \infty \quad \forall a \in \mathbb{R} \cup \{\infty\}$
$a \cdot \infty = \infty \cdot a = \infty \quad \forall a > 0$
$\infty - \infty$ nicht definiert.

Wichtiges Beispiel: Lebesgue-Maß l, festgelegt durch $l((a, b]) = b - a \quad \forall a, b \in \mathbb{R},\ a < b$.
WMaße sind Maße mit der Gesamtmasse 1 ($\mu(\Omega) = 1$)

Bemerkung 8.1

(i) "Allgemeine" Maße sind noch stets stetig von unten, aber i.A. nicht stetig von oben oder in $\emptyset$, z.B. $l((-\infty, n]) = \infty \quad \forall n \in \mathbb{N}$, aber $l(\bigcap_{n=1}^{-\infty}(-\infty, n]) = l(\emptyset) = 0$

(ii) l und $2l$ stimmen auf dem $\cap$-stabilen Erzeuger $\{(-\infty, x] : x \in \mathbb{R}\}$ überein, sind aber nicht mehr gleich: Satz 5.9 (Eindeutigkeitssatz) muß modifiziert werden (Übungsaufgabe)

(iii) Beziehungen wie $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ sind bei allgemeinen Maßen i.A. falsch (sinnlos), da $\infty - \infty$ auftreten kann.

8.2 Konstruktion des Lebesgue-Integrals

$(\Omega, \mathfrak{A}, \mu)$ sei ein Maßraum. Für eine (möglichst große) Klasse von $(\mathfrak{A}, \mathfrak{B})$-meßbaren Funktionen $f : \Omega \to \mathbb{R}$ soll $\int f d\mu \quad (= \int f(\omega)\mu d(\omega) = \int f(\omega) d\mu(\omega))$ definiert werden. Wir beginnen mit der Forderung, daß für Indikatorfunktionen $f = 1_A \quad \int f d\mu = \mu(A)$ gelten soll.
Beachte: $f = 1_A$ meßbar $\iff A \in \mathfrak{A}$.

Soll $f \to \int f d\mu$ linear sein, so ist damit $\int f d\mu$ für endliche Linearkombinatinen von meßbaren Indikatorfunktionen mit nicht-negativen Koeffizienten festgelegt.
Wir schreiben $Prim_+(\Omega, \mathfrak{A})$ für diese Klasse ("primitive Funktionen").
Offensichtlich gilt: $Prim_+(\Omega, \mathfrak{A}) = \{f : \Omega \to \mathbb{R} : f\ (\mathfrak{A}, \mathfrak{B})\text{-meßbar}, \#f(\Omega) < \infty^1\}$.

Definition 8.2 ("Stufe 1") Für $f \in Prim_+(\Omega, \mathfrak{A})$ sei $\int f d\mu := \sum_{x \in f(\Omega)} x\mu\ \underbrace{(f^{-1}(\{x\}))}_{\{\omega \in \Omega : f(\omega) = x\}}$.

Hierbei vereinbaren wir $0 \cdot \infty = \infty \cdot 0 = 0$; insbesondere gilt auch für endliche Maße $\mu \ \int c d\mu = c\mu(\Omega),\ c \geq 0$ (Integrale konstanter Funktionen, $\int 0 d\mu = 0$)

Lemma 8.3 Es seien $f, g \in Prim_+(\Omega, \mathfrak{A})$ und $c \geq 0$. Dann gilt:
$cf \in Prim_+(\Omega, \mathfrak{A}),\ f + g \in Prim_+(\Omega, \mathfrak{A})$, sowie
$\int cf d\mu = c \int f d\mu,\ \int (f + g) d\mu = \int f d\mu + \int g d\mu$ und $f \leq g \Longrightarrow \int f d\mu \leq \int g d\mu$.

Beweis der Monotonie (Linearität in den Übungen): Sei $f(\Omega) = \{a_1, \ldots, a_n\}$ und $g(\Omega) = \{b_1, \ldots, b_m\}$. Man setze

[1] Anzahl der Elemente des Wertebereiches von f ist kleiner ∞

$A_i := f^{-1}(\{a_i\})$ und $B_j := g^{-1}(\{b_j\})$, so gilt: $f = \sum_{i=1}^{n} a_i \cdot 1_{A_i}$, $g = \sum_{j=1}^{m} b_j \cdot 1_{B_j}$.
Aus $f \leq g$ folgt $A_i \cap B_j \neq \emptyset \implies a_i \leq b_j$, also:

$$\int f d\mu = \sum_{i=1}^{n} a_i \mu(A_i) =^2 \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mu(A_i \cap B_j) \leq \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} b_j \mu(B_j) = \int g d\mu. \qquad \square$$

Dieses Lemma zeigt, daß das bisher definierte Integral linear und isoton ist (auf $Prim_+$), insbesondere gilt für alle Funktionen $f = \sum_{i=1}^{n} a_i \cdot 1_{A_i}$, $a_i \geq 0$, $A_i \in \mathfrak{A}$ die Formel $\int f d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$, auch wenn die A_i's nicht disjunkt oder die a_i's nicht verschieden sind (alle "Darstellungen" liefern dasselbe Integral).

Lemma 8.4 Zu jeder $(\mathfrak{A}, \mathfrak{B})$-meßbaren, nicht-negativen Funktion $f : \Omega \to \mathbb{R}$ existiert eine Folge $(f_n)_{n \in \mathbb{N}} \subset Prim_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f$.

Beweis: $f_n := \sum_{i=0}^{2^n - 1} \frac{i}{2^n} 1_{A_{ni}}$; $A_{ni} := \{\omega : \frac{i}{2^n} \leq f(\omega) < \frac{i+1}{2^n}\}$. $\square$

Definition 8.5 ("Zweite Stufe") Das μ-Integral einer nicht-negativen, $(\mathfrak{A}, \mathfrak{B})$-meßbaren Funktion $f : \Omega \to \mathbb{R}^+$ wird definiert durch $\int f d\mu := \sup\{\int g d\mu : g \in Prim_+(\Omega, \mathfrak{A}), g \leq f\}$.
Wegen $g \equiv 0 \in Prim_+$, $g \leq f$ wird das Supremum über eine nicht-leere Menge gebildet.

Im Falle $f \in Prim_+$ taucht f selbst in dieser Menge auf; wegen der Monotonie auf $Prim_+$ (Lemma 8.4) impliziert dies, daß Def. 8.5 mit Def. 8.2 verträglich ist.

Lemma 8.6 Ist $(f_n)_{n \in \mathbb{N}} \subset Prim_+$ mit $f_n \uparrow f$, so gilt $\int f d\mu = \lim_{n \to \infty} \int f_n d\mu$.

Beweis: "$\geq$": folgt unmittelbar aus Def. 8.5.
"$\leq$": Sei nun $g \in Prim_+(\Omega, \mathfrak{A})$; $g = \sum_{i=1}^{k} a_i \cdot 1_{A_i}$ mit $g(\Omega) = \{a_1, \dots, a_k\}$, $g \leq f$.
Sei $0 < \eta < 1$ und $B_n := \{\omega \in \Omega : \eta g(\omega) \leq f_n(\omega)\}$.
Aus $\eta g < f$ auf $\{f > 0\}$ und $f_n = f$ auf $\{f = 0\}$, sowie $f_n \uparrow f$ folgt $B_n \uparrow \Omega$, also mit der Stetigkeit von unten von μ

$$\eta \int g d\mu = \eta \lim_{n \to \infty} \sum_{i=1}^{k} a_i \mu(A_i \cap B_n) = \lim_{n \to \infty} \int \eta g 1_{B_n} d\mu \leq \lim_{n \to \infty} \int f_n d\mu.$$

Da die rechte Seite nicht von η abhängt, folgt hieraus mit $\eta \uparrow 1$ $\quad \int g d\mu \leq \lim_{n \to \infty} \int f_n d\mu$.
Bildet man das Supremum über die zugelassenen g, so folgt die andere Ungleichung. $\square$

Lemma 8.7 Es seien $f, g : \Omega \to \mathbb{R}^+$ $(\mathfrak{A}, \mathfrak{B})$-meßbar und $c \geq 0$. Dann gilt:
$\int (cf) d\mu = c \int f d\mu$, $\int (f+g) d\mu = \int f d\mu + \int g d\mu$
$f \leq g \implies \int f d\mu \leq \int g d\mu$

Beweis: Lemma 8.4 liefert die Existenz von Folgen $(f_n)_{n \in \mathbb{N}}$, $(g_n)_{n \in \mathbb{N}} \subset Prim_+$ mit $f_n \uparrow f$, $g_n \uparrow g$.
Klar: $(f_n + g_n)_{n \in \mathbb{N}} \subset Prim_+$, $f_n + g_n \uparrow f + g$.

$$\int (f+g) d\mu \overset{\text{Lemma 8.6}}{=} \lim_{n \to \infty} \int (f_n + g_n) d\mu = \lim_{n \to \infty} \left(\int f_n d\mu + \int g_n d\mu\right) = \lim_{n \to \infty} \int f_n d\mu + \lim_{n \to \infty} \int g_n d\mu$$
$$\overset{\text{2xLemma 8.6}}{=} \int f d\mu + \int g d\mu$$

Analog: $\int (cf) d\mu = \lim_{n \to \infty} \int (cf_n) d\mu = c \lim_{n \to \infty} \int f_n d\mu = c \int f d\mu$
Isotonie folgt unmittelbar aus der Definition (es wird das Supremum über eine größere Menge gebildet.) $\square$

Wir definieren Positiv- und Negativteil einer Funktion f durch $f^+ := \max\{f, 0\}$, $f^- := -\min\{f, 0\}$
Klar: $f = f^+ - f^-$, $|f| = f^+ + f^-$
Mit f sind auch f^+ und f^- $(\mathfrak{A}, \mathfrak{B})$-meßbar.

Definition 8.8 ("Stufe 3") Eine $(\mathfrak{A}, \mathfrak{B})$-meßbare Funktion $f : \Omega \to \mathbb{R}$ heißt <u>integrierbar bezüglich μ</u>, wenn $\int f^+ d\mu < \infty$ und $\int f^- d\mu < \infty$.
Das <u>μ-Integral</u> einer solchen Funktion wird definiert durch $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$ (andere Schreibweisen: $\int f(\omega) \mu(d\omega)$, $\int f(\omega) d\mu(\omega)$, $\mu(f)$).
Die Menge aller μ-integrierbaren Funktionen auf $(\Omega, \mathfrak{A})$ wird mit $\mathfrak{L}^1(\Omega, \mathfrak{A}, \mu)$ bezeichnet.

Bemerkung 8.9 Gilt mindestens eine der Beziehung $\int f^+ d\mu < \infty$, $\int f^- d\mu < \infty$, so läßt sich $\int f d\mu$ noch sinnvoll als Differenz der beiden Werte definieren; solche Funktionen werden gelegentlich "quasi-integrierbar" oder "integrierbar im erweiterten Sinne" genannt.
Unmittelbar aus der Definition und $|f| = f^+ + f^-$ folgt:
(*) $\quad f$ μ-integrierbar $\iff |f|$ μ-integrierbar.
(**) $\quad |\int f d\mu| \leq \int |f| d\mu$

Satz 8.10 Es seien $f, g : \Omega \to \mathbb{R}$ μ-integrierbar und $c \in \mathbb{R}$. Dann sind auch cf und $f + g$ integrierbar und es gilt
$\int (cf) d\mu = c \int f d\mu$, $\int (f+g) d\mu = \int f d\mu + \int g d\mu$
$f \leq g \implies \int f d\mu \leq \int g d\mu$.

[2] denn $B_1, \dots, B_m$ bilden eine Partition von Ω

Beweis: Wegen $|f+g| \leq |f|+|g|$ folgt die Integrierbarkeit von $f+g$ aus (*)
$(f+g)^+$ ist i.a. verschieden von $f^+ + g^+$.
Es gilt jedoch $(f+g)^+ - (f+g)^- = f+g = f^+ - f^- + g^+ - g^-$, also $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$.
Damit (nach Lemma 8.7) $\int (f+g)^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int (f+g)^- d\mu + \int f^+ d\mu + \int g^+ d\mu$.
Also: $\int (f+g) d\mu = \int (f+g)^+ d\mu - \int (f+g)^- d\mu = \int f^+ d\mu + \int g^+ d\mu - \int f^- d\mu - \int g^- d\mu = \int f d\mu + \int g d\mu$.
Der Beweis $\int (cf) d\mu = c \int f d\mu$ ist einfach, wenn man $c \geq 0$ und $c < 0$ separat betrachtet.
Isotonie: Übungsaufgabe. □

Beispiel 8.11 (i) Bezeichnet δ_ω das Einpunktmaß in ω, also $\delta_\omega(A) = \begin{cases} 1 & , \quad \omega \in A \\ 0 & , \quad \text{sonst} \end{cases} = 1_A(\omega)$, so gilt $\int f d\delta_\omega = f(\omega)$ (Dirac-Maß)
(ii) Ein Maß μ von der Form $\mu = \sum\limits_{i=1}^{\infty} \alpha_i \delta_{\omega_i}$ (mit $\alpha_i \geq 0 \ \forall i$) heißt diskret.
Eine Funktion $f : \Omega \to \mathbb{R}$ ist genau dann μ-integrierbar, wenn $\sum\limits_{i=1}^{\infty} \alpha_i |f(\omega_i)| < \infty$, und man hat dann $\int f d\mu = \sum\limits_{i=1}^{\infty} \alpha_i f(\omega_i) \quad (= \sum\limits_{\omega \in \Omega} f(\omega)\mu(\{\omega\})\ )$
Im Spezialfall $\Omega = \mathbb{N}$, $\mathfrak{A} = \mathfrak{P}(\mathbb{N})$, $\mu = \sum\limits_{n=1}^{\infty} \delta_n$ (das abzählbare Maß; $\mu(A) = \#A$) erhält man Reihensummen als Integral.
(iii) Es seien $\Omega = [a,b]$, $a < b$, $\mathfrak{A}$ die Spur von $\mathfrak{B}$ auf Ω und μ die Restriktion von l auf $\mathfrak{A}$. Ist $f : \Omega \to \mathbb{R}$ Borel-meßbar und Riemannintegrierbar (Konvergenz der Ober- und Untersummen, etc.), so ist f auch μ-integrierbar und es gilt $\int f d\mu = \int\limits_a^b f(x) dx$.
Die Funktion $1_{\mathbb{Q} \cap [a,b]}$ ist ein Beispiel für eine μ-integrierbare Funktion, die nicht Riemannintegrierbar ist.
Grob: Das Integral bzgl. l (das Lebesgue-Integral) erweitert das (eigentliche) Riemann-Integral, so daß die bekannten Formeln für das Riemann-Integral weiterverwendet werden können (wir schreiben gelegentlich "dx" anstelle von "$l(dx)$".)

8.3 Transformationsformel, Nullmengen

$(\Omega, \mathfrak{A}, \mu)$ sei ein Maßraum, $(\Omega', \mathfrak{A}')$ ein meßbarer Raum, $T : \Omega \to \Omega'$ eine $(\mathfrak{A}, \mathfrak{A}')$-meßbare Funktion. Dann wird durch $\mu' : \mathfrak{A}' \to [0, \infty]$, $\mu'(A') := \mu(T^{-1}(A')) \quad \forall A' \in \mathfrak{A}'$ ein Maß auf $(\Omega', \mathfrak{A}')$ definiert, das Bildmaß von μ unter T (Schreibweise: μ^T.)
Dies verallgemeinert den aus Satz 5.4 bekannten Begriff der Verteilung einer Zufallsgröße.

Satz 8.12 (Integration bzgl. eines Bildmaßes, Transformationssatz)
Mit der oben gezeigten Beziehung gilt für alle $(\mathfrak{A}', \mathfrak{B})$-meßbaren $f : \Omega' \to \mathbb{R}$:
f ist genau dann μ^T-integrierbar, wenn $f \circ T$ μ-integrierbar ist und dann gilt:
$\int f d\mu^T = \int f \circ T d\mu$.

$$\begin{tikzcd} (\Omega, A, \mu) \arrow[r, "T"] \arrow[dr, "f \circ T"'] & (\Omega', A') \arrow[d, "f"] \\ & (R, B) \end{tikzcd}$$

Beweis: (mit der "üblichen Maschinerie", "algebraische Induktion")
Im Falle $f = 1_A$ hat man $\int f d\mu^T = \mu(T^{-1}(A)) = \int 1_{T^{-1}(A)} d\mu = \int 1_A \circ T d\mu$.
Die Beh. ist also für Indikatorfunktionen richtig. Da das Integral linear ist, ist sie damit auch auf $Prim_+(\Omega', \mathfrak{A}')$ richtig.
Ist $f \geq 0$, so existiert nach Lemma 8.4 und 8.6 eine Folge $(f_n)_{n \in \mathbb{N}} \subset Prim_+(\Omega', \mathfrak{A}')$ mit $f_n \uparrow f$ und $\int f d\mu^T = \lim\limits_{n \to \infty} \int f_n d\mu^T$.
Klar: $(f_n \circ T)_{n \in \mathbb{N}} \subset Prim_+(\Omega, \mathfrak{A})$, $f_n \circ T \uparrow f \circ T$.
Also folgt mit Lemma 8.6 $\int f d\mu^T = \lim\limits_{n \to \infty} \int f_n d\mu^T = \lim\limits_{n \to \infty} \int f_n \circ T d\mu = \int f \circ T d\mu$.
Ist schließlich $f : \Omega' \to \mathbb{R}$ eine beliebige $(\mathfrak{A}', \mathfrak{B})$-meßbare Abbildung, so erhält man mit dem bereits gezeigten:
$\int f^\pm d\mu^T < \infty \iff \int f^\pm \circ T d\mu < \infty$.
Da $(f \circ T)^\pm = f^\pm \circ T$, folgt die behauptete Integrabilitätsaussage; auch gilt dann
$\int f d\mu^T = \int f^+ d\mu^T - \int f^- d\mu^T = \int f^+ \circ T d\mu - \int f^- \circ T d\mu = \int (f \circ T)^+ d\mu - \int (f \circ T)^- d\mu = \int (f \circ T) d\mu$. □

Definition 8.13 $(\Omega, \mathfrak{A}, \mu)$ Maßraum. Für jedes $\omega \in \Omega$ sei $A(\omega)$ eine Aussage. Man sagt: "A gilt fast überall" (μ-f.ü.), wenn es eine μ-Nullmenge N (also ein $N \in \mathfrak{A}$ mit $\mu(N) = 0$) gibt mit: $A(\omega)$ ist wahr für alle $\omega \notin N$.

Satz 8.14 $(\Omega, \mathfrak{A}, \mu)$ Maßraum, $f, g : \Omega \to \mathbb{R}$ meßbar.
(i) Sei außerdem $f \geq 0$. Dann gilt: $\int f d\mu = 0 \iff f = 0$ μ-f.ü.
(ii) Ist f integrierbar und gilt $f = g$ μ-f.ü., so ist auch g integrierbar und es gilt $\int f d\mu = \int g d\mu$.
Beweis: (i) Setze $N = \{f \neq 0\}$.
- Angenommen es gilt $\int f d\mu = 0$. Setze $A_n = \{\omega \in \Omega : f(\omega) \geq \frac{1}{n}\}$. Dann $A_n \uparrow N$, also $\mu(N) = \lim\limits_{n \to \infty} \mu(A_n)$.
  $0 = \int f d\mu \geq \int \frac{1}{n} 1_{A_n} d\mu = \frac{1}{n}\mu(A_n) \geq 0$, also $\mu(A_n) = 0 \ \forall n$, und damit $\mu(N) = 0$.
- Ist umgekehrt N eine μ-Nullmenge mit $f(\omega) = 0 \ \forall \omega \in N^c$, und $g = \sum a_i 1_{A_i} \in Prim_+$ mit $g \leq f$, $\{a_1, \ldots, a_n\} = g(\Omega)$, so folgt $A_i \subseteq N \ \forall i$ mit $a_i > 0$ und damit $\int g d\mu = \sum\limits_{a_i > 0} a_i \mu(A_i) = 0$.

Hieraus folgt $\int f d\mu = 0$ mit Def. 8.5 (Supremumsbildung)

(ii) Es seien zunächst $f, g \geq 0$, $N := \{f \neq g\}$. Dann gilt
$\int f d\mu = \int f 1_N d\mu + \int f 1_{N^c} d\mu = 0 + \int g 1_{N^c} d\mu = \int g 1_N d\mu + \int g 1_{N^c} d\mu = \int g d\mu$.
Insbesondere $\int f d\mu = \infty \iff \int g d\mu = \infty$, hieraus folgt (bei beliebigen f, g) die Integrierbarkeitsaussage wie im Beweis des Transformationssatzes.
Es seien nun $f, g : \Omega \to \mathbb{R}$ meßbar mit $f = g$ μ-f.ü. Wegen $\{f^+ = g^+\} \supseteq \{f = g\} \subseteq \{f^- = g^-\}$ gilt dann auch $f^+ = g^+$ μ-f.ü., also $\int f d\mu = \int f^+ d\mu - \int f^- d\mu = \int g^+ d\mu - \int g^- d\mu = \int g d\mu$. □

8.4 Konvergenzsätze

Das Maß-Integral hat bei der Behandlung von Funktionenfolgen gegenüber dem Riemann-Integral einige Vorzüge.
$(\Omega, \mathfrak{A}, \mu)$ Maßraum, $f, f_1, f_2, \ldots : \Omega \to \mathbb{R}$ seien meßbar.

Satz 8.15 (Satz von Beppo Levi, Satz von der monotonen Konvergenz)
Sind $f, f_1, f_2, \ldots$ nicht-negativ mit $f_n \uparrow f$, so gilt $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$.

Beweis: Zu jedem f_n existiert eine Folge $(g_{nm})_{m\in\mathbb{N}} \subset Prim_+$ mit $g_{nm} \uparrow f_n$ für $m \to \infty$

Diagramm:
$$\begin{array}{ccccccccccc} g_{11} & \leq & g_{12} & \leq & g_{13} & \leq & g_{14} & \leq & \cdots & \to & f_1 \\ g_{21} & \leq & g_{22} & \leq & g_{23} & \leq & g_{24} & \leq & \cdots & \to & f_2 \\ g_{31} & \leq & g_{32} & \leq & g_{33} & \leq & g_{34} & \leq & \cdots & \to & f_3 \\ \vdots & & \vdots & & \vdots & & \vdots & & & & \downarrow \\ & & & & & & & & & & f \end{array}$$

Setze $h_m := \max\{g_{1m}, \ldots, g_{mm}\}$.
Klar: $h_m \in Prim_+ \ \forall m$, $h_m \uparrow$.
Für $n \leq m$ gilt $g_{nm} \leq h_m$, also folgt $f_n = \sup_m g_{nm} = \sup_{m\geq n} g_{nm} \leq \sup_m h_m$.
Außerdem $h_m \leq f_m \leq f$, also insgesamt $h_m \uparrow f$.
Nun folgt mit Lemma 8.6 $\int f d\mu = \lim_{m\to\infty} \int h_m d\mu \leq \lim_{m\to\infty} \int f_m d\mu \leq \int f d\mu$. □

Satz 8.16 (Lemma von Fatou)
Gilt $f_n \geq 0 \ \forall n$, so folgt $\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$.

Beweis: Setze $g_n := \inf_{m\geq n} f_m$, dann gilt $g_n \uparrow f := \liminf_{n\to\infty} f_n$.
$g_n \leq f_n$. Mit Satz 8.15 folgt $\int \liminf_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int g_n d\mu = \liminf_{n\to\infty} \int g_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$. □

Satz 8.17 (Satz von der majorisierten (dominierten) Konvergenz, Satz von Lebesgue)
Es gelte $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ für alle $\omega \in \Omega$. Existiert dann eine μ-integrierbare Funktion $g : \Omega \to \mathbb{R}$ mit $|f_n(\omega)| \leq g(\omega)$ für alle $\omega \in \Omega$, $n \in \mathbb{N}$, so folgt $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$. ($g$ heißt "integrierbare Majorante")

Beweis: Setze $g_n := |f_n - f|$, $h := |f| + g$.
Wegen $|h| \leq 2g$ ist h integrierbar. Außerdem $h - g_n = |f| + g - |f_n - f| \geq |f| + g - |f_n| - |f| = g - |f_n| \geq 0$.
Wegen $g_n \to 0$ gilt $h - g_n \to h$, also folgt mit Fatou $\int h d\mu \leq \liminf_{n\to\infty} \int (h - g_n) d\mu = \int h d\mu - \limsup_{n\to\infty} \int g_n d\mu$.
(beachte: hier geht Integrierbarkeit von g ein, sonst könnt hier u.U. $\infty - \infty$ stehen.)
also: $\limsup \int g_n d\mu = 0$.
Damit $\lim_{n\to\infty} |f_n - f| d\mu = \lim_{n\to\infty} \int g_n d\mu = 0$ und mit $|\int f_n d\mu - \int f d\mu| = |\int (f_n - f) d\mu| \leq \int |f_n - f| d\mu$ folgt hieraus $\int f_n d\mu \to \int f d\mu$. □

Bemerkung 8.18

(i) Man kann die Konstruktion des μ-Integrals auf $\overline{\mathbb{R}}$ ($:= [-\infty, \infty]$)-wertige Funktionen ausdehnen, muß dann aber bei den Linearitätsaussagen Einschränkungen machen ($f + g$ u.U. nicht überall definiert). Ist ein $\overline{\mathbb{R}}$-wertiges f in diesem Sinne integrierbar, so ist $N := \{\omega \in \Omega : |f(\omega)| = \infty\}$ eine μ-Nullmenge und es gilt $\int f d\mu = \int \underbrace{f 1_{N^c} d\mu}_{\mathbb{R}-\text{wertig}}$.

(ii) Die Erweiterung aus (i) wurde in den obigen Sätzen bereits implizit verwendet ($\liminf f_n$ beispielsweise kann durchaus den Wert ∞ annehmen.)

(iii) Zu den Sätzen 8.15-8.17 existieren auch "fast-überall"-Versionen. Hat man beispielsweise in Satz 8.17 nur $f_n \to f$ μ-f.ü., so gilt noch stets $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$:
Ist N eine μ-Nullmenge, außerhalb dieser Konvergenz gilt, so hat man $f_n 1_{N^c}(\omega) \to f 1_{N^c}(\omega)$ für <u>alle</u> $\omega \in \Omega$.
Auf $f_n 1_{N^c}$ läßt sich Satz 8.17 anwenden:
$\int f d\mu \overset{8.14 (ii)}{=} \int f 1_{N^c} d\mu \overset{8.17}{=} \lim_{n\to\infty} \int f_n 1_{N^c} d\mu \overset{8.14 (ii)}{=} \lim_{n\to\infty} \int f_n d\mu$.

8.5 Maße und Dichten

Abkürzende Schreibweise: $\int\limits_A f d\mu = \int f 1_A d\mu$.

Lemma 8.19 $(\Omega, \mathfrak{A}, \mu)$ Maßraum, $f : \Omega \to \mathbb{R}_+$ meßbar. Dann wird durch $\nu : \mathfrak{A} \to [0, \infty]$, $\nu(A) := \int\limits_A f d\mu$ ein Maß auf $(\Omega, \mathfrak{A})$ definiert.
Beweis: $\nu(\emptyset) = \int\limits_\emptyset f d\mu = \int f 1_\emptyset d\mu = \int 0 d\mu = 0$.
Seien $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkt.
$\nu(\sum\limits_{n=1}^{\infty} A_n) = \int f 1_A d\mu =^3 \lim\limits_{n\to\infty} \int f 1_{\bigcup\limits_{i=1}^{n} A_i} d\mu =^4 \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} \int f 1_{A_i} d\mu =^5 \lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} \nu(A_i) = \sum\limits_{i=1}^{\infty} \nu(A_i)$. □

Definition 8.20 μ, ν seien Maße auf $(\Omega, \mathfrak{A})$. Existiert eine nicht-negative, meßbare Funktion $f : \Omega \to \mathbb{R}$ mit $\nu(A) = \int\limits_A f d\mu$ für alle $A \in \mathfrak{A}$, so heißt f eine μ-Dichte von ν. Schreibweise: $f = \frac{d\nu}{d\mu}$.

Aus Satz 8.14 (ii) folgt, daß mit f auch jedes meßbare g mit $f = g$ μ-f.ü. eine μ-Dichte von ν ist. Dichten sind i.A. nicht eindeutig.
Hat ν eine Dichte f bezüglich μ, so folgt mit Satz 8.14 (i) $\mu(N) = 0 \implies f \cdot 1_N = 0$ μ-f.ü. $\implies \nu(N) = 0$.
Gilt dies, so sagt man, daß μ ν dominiert und schreibt $\mu >> \nu$. Der Satz von Radon-Nikodym besagt, daß die Eigenschaft bei σ-endlichen Maßen bereits die Existenz einer Dichte impliziert.

Beispiel 8.21 Der Fall $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}, \mathfrak{B}, l)$ ist von besonderem Interesse. Ist $f : \mathbb{R} \to \mathbb{R}$ eine Borel-meßbare, nicht-negative Funktion mit l-Integral 1, so wird durch $P : \mathfrak{B} \to [0, 1]$, $P(A) := \int\limits_A f(x)dx$ ein WMaß definiert, das WMaß mit (Lebesgue-) Dichte f.
Ist X eine Zufallsvariable (meßbare Abbildung) mit Verteilung (Bildmaß) P, so heißt f eine WDichte von X.

Satz 8.22 μ, ν Maße auf $(\Omega, \mathfrak{A})$, ν habe μ-Dichte f. Dann gilt für alle $(\mathfrak{A}, \mathfrak{B})$-meßbaren $g : \Omega \to \mathbb{R}$:
g ist genau dann ν-integrierbar, wenn $g \circ f$ μ-integrierbar ist, und dann: $\int g d\nu = \int g \circ f d\mu$.
Beweis: mit der "üblichen Maschinerie", Übungsaufgabe. □

8.6 Erwartungswerte

$(\Omega, \mathfrak{A}, P)$ WRaum, $X : \Omega \to \mathbb{R}$ ZV.

Definition 8.23
(i) Ist X P-integrierbar, dann heißt $EX := \int X dP$ der Erwartungswert von X. Gilt $E|X|^k \leq \infty$, so heißt EX^k das k-te Moment zu X (im Falle $E|X|^k = \infty$ sagen wir, daß das k-te Moment nicht existiert.)
(ii) Existiert zu X das zweite Moment, so heißt $var(X) := E(X - EX)^2$ die Varianz von X und $\sigma(X) := \sqrt{var(X)}$ die Standardabweichung.

Satz 8.24 (Wir setzen voraus, daß die beteiligten Erwartungswerte, etc. existieren.)
(i) $E(X + Y) = EX + EY$, $E(c \cdot X) = c \cdot EX$
(ii) $X \leq Y \implies EX \leq EY$
(iii) Ist X diskret verteilt, d.h. gibt es eine abzählbare Menge $A \subset \mathbb{R}$ mit $P(X \in A) = 1$, so gilt für alle Borel-meßbaren $g : \mathbb{R} \to \mathbb{R}$ $Eg(X) = \sum\limits_{x\in A} g(x) \cdot P(X = x)$.
(iv) Ist X absolut stetig verteilt, d.h. X hat eine Dichte f, so gilt für alle Borel-meßbaren $g : \mathbb{R} \to \mathbb{R}$ $Eg(X) = \int g(x)f(x)dx$.
Beweis: (i) und (ii): Satz 8.10
(iii) $P^X = \sum\limits_{x\in A} P(X = x)\delta_x$, Satz 8.12
(iv) Aus Satz 8.12, Satz 8.22 folgt $Eg(X) = \int\limits_\Omega g(X(\omega))P(d\omega) = \int\limits_\mathbb{R} g(x)P^X(dx) = \int\limits_\mathbb{R} g(x)f(x)dx$. □

Offensichtlich gilt weiterhin:
$$\begin{aligned} var(a + X) &= var(X) \qquad \forall a \in \mathbb{R} \\ var(aX) &= a^2 var(X) \qquad \forall a \in \mathbb{R} \\ var(X) &= EX^2 - (EX)^2 \end{aligned}$$

Beispiel 8.25 (siehe auch Beispiel 5.25)
$X \sim N(0, 1)$, d.h. X hat Dichte $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$.
Mit Satz 8.24(iv) folgt $EX = \int x \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx = 0$, $EX^2 = \int x^2 \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx = 1$, also $var(X) = 1$.

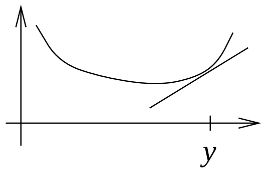
---
[3] mit $A := \bigcup_{i=1}^{\infty} A_i$
[4] Mit dem Satz von der monotonen Konvergenz für $(0 \leq)$ $f 1_{\bigcup_{i=1}^{n} A_i} \uparrow f 1_A$
[5] da A_i's disjunkt

Bekanntlich: $Y := \mu + \sigma X \sim N(\mu, \sigma^2)$, also $EY = E(\mu + \sigma X) = \mu + \sigma EX = \mu$,
$var(Y) = var(\mu + \sigma X) = var(\sigma X) = \sigma^2 var(X) = \sigma^2$.
Interpretation der Parameter einer Normalverteilung: $\mu \rightsquigarrow$ Erwartungswert, $\sigma^2 \rightsquigarrow$ Varianz.

8.7 Ungleichungen

Satz 8.26 $(\Omega, \mathfrak{A}, P)$ WRaum, $X : \Omega \to \mathbb{R}$ ZV.
(i) Markovsche Ungleichung
$P(|X| \geq a) \leq \frac{1}{a^k} E|X|^k$ für alle $a > 0$.
(ii) Ungleichung von Chebychev
Existiert zu X die Varianz, so gilt $P(|X - EX| \geq a) \leq \frac{1}{a^2} var(X)$.
Beweis:
(i) Definiere $Y : \Omega \to \mathbb{R}$ durch $Y(\omega) = \begin{cases} 0 & , \quad |X(\omega)| < a \\ a & , \quad |X(\omega)| \geq a \end{cases}$
Offensichtlich gilt dann $|Y| \leq |X|$, also $|Y|^k \leq |X|^k$ und es folgt $a^k P(|X| \geq a) = a^k P(|Y| = a) = E|Y|^k \leq E|X|^k$.
(ii) Wende Teil (i) auf $X - EX$, $k = 2$ an. □

Es sei $I \subset \mathbb{R}$ ein offenes Intervall und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion,
d.h. $\varphi(\alpha x + (1-\alpha)y) \leq \alpha\varphi(x) + (1-\alpha)\varphi(y)$.



Lemma 8.27 φ, I wie oben. Dann existiert zu jedem $y \in I$ ein $m \in \mathbb{R}$ mit $\varphi(x) \geq \varphi(y) + m(x - y)$ für alle $x \in I$
(anschaulich: durch jeden Punkt des Graphen geht eine Gerade (nicht notwendig eine Tangente, siehe $|\ldots|$), die unterhalb des Graphen liegt.

Satz 8.28 (Jensensche Ungleichung)
$I \subset \mathbb{R}$ sei ein offenes Intervall, $\varphi : I \to \mathbb{R}$ konvex, X ZV mit $E(X) < \infty$, $E|\varphi(X)| < \infty$, $P(X \in I) = 1$.
Dann gilt: $EX \in I$ und $\varphi(EX) \leq E\varphi(X)$. (Wichtige Spezialfälle: $|EX| \leq E|X|$, $(EX)^2 \leq EX^2$.)
Beweis: Ist $I = (-\infty, \infty)$, so gilt automatisch $EX \in I$.
Ist $X < a$ P-f.s., so gilt $EX \leq Ea = a$.
Wegen $a - X \geq 0$ würde $E(a - X) = 0$ auf $P(X = a) = 1$ führen, also $EX < a$ (analog: $X > a$ P-f.s. $\Longrightarrow$ $EX > a$.)
Also: $P(X \in I) = 1 \implies EX \in I$.
Nach Lemma 8.27 existiert ein $m \in \mathbb{R}$ mit $\varphi(X) \geq \varphi(EX) + m(X - EX)$. Nehme nun den Erwartungswert davon. □

$(\Omega, \mathfrak{A}, \mu)$ Maßraum, $\mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) := \{f : \Omega \to \mathbb{R} : \int |f|^p d\mu < \infty\}$, $||f||_p := (\int |f|^p d\mu)^{\frac{1}{p}}$.
Klar: $||f||_p \geq 0$, $||af||_p = |a| \cdot ||f||_p$.

Satz 8.29
(i) (Höldersche Ungleichung)
Hat man $f \in \mathfrak{L}^p$, $g \in \mathfrak{L}^q$, wobei $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, so folgt $f \cdot g \in \mathfrak{L}^1$ und $||f \cdot g||_1 \leq ||f||_p ||g||_q$.
(ii) (Minkowskische Ungleichung)
$f, g \in \mathfrak{L}^p$, $p \geq 1 \implies f + g \in \mathfrak{L}^p$, $||f + g||_p \leq ||f||_p + ||g||_p$.
Beweis: (i) Man kann f durch $|f|$ und g durch $|g|$ ersetzen, also $f, g \geq 0$ annehmen. Sei $c := \int f^p d\mu$.
Bei $c = 0$ hat man $f = 0$ μ-f.ü. und die Ungleichung ist trivial.
Sei also $c > 0$. Sei P das Maß mit μ-Dichte $\frac{1}{c} f^p$.
Sei $h(\omega) := \begin{cases} \frac{g(\omega)}{f(\omega)^{p-1}} & , \quad f(\omega) > 0 \\ 0 & , \quad \text{sonst.} \end{cases}$.
Dann $||f||_p ||g||_q \geq ||f||_p ||g 1_{\{f>0\}}||_q = c^{\frac{1}{p}} (\int h^q g^{q(p-1)} d\mu)^{\frac{1}{q}} = c^{\frac{1}{p}} (c \int h^q dP)^{\frac{1}{q}} \overset{\text{Jensen}}{\geq} c^{\frac{1}{p} + \frac{1}{q}} \int h dP = \int g f d\mu = ||fg||_1$.
(ii) Wegen $|f + g| \leq |f| + |g|$ gilt $||f + g||_p \leq ||\ |f| + |g|\ ||_p$, es reicht also wieder, die Ungleichung für $f, g \geq 0$ zu beweisen. Im Falle $p = 1$ hat man $||f + g||_1 = \int (f + g) d\mu = \int f d\mu + \int g d\mu = ||f||_1 + ||g||_1$.
Sei $p > 1$: $(f + g)^p \leq (2 \cdot \max\{f, g\})^p = 2^p \max\{f^p, g^p\} \leq 2^p (f^p + g^p)$. Dies zeigt, daß mit f und g auch $f + g$ in $\mathfrak{L}^p$ liegt. Mit $q := (1 - \frac{1}{p})^{-1}$ liefert Hölder: $||f + g||_p{}^p = \int (f + g)^p d\mu = \int f(f + g)^{p-1} d\mu + \int g(f + g)^{p-1} d\mu \leq (||f||_p + ||g||_p) ||(f + g)^{p-1}||_q$.
$||(f + g)^{p-1}||_q = (\int (f + g)^{(p-1)q} d\mu)^{\frac{1}{q}} = ||f + g||_p{}^{(p-1)}$.
Im Falle $||f + g||_p = 0$ ist nichts zu zeigen.
Im Falle $||f + g||_p > 0$ folgt die behauptete Ungleichung nach Division durch $||f + g||_p{}^{p-1}$ $(< \infty)$. □

Spezialfall von Hölder für $p = q = 2$. $|\int fg d\mu| \leq (\int f^2 d\mu)^{\frac{1}{2}} (\int g^2 d\mu)^{\frac{1}{2}}$. Dies ist die Cauchy-Schwarz-Ungleichung.

Spezialfall: $\Omega = \{1, \ldots, n\}$, $\mathfrak{A} = \mathbb{P}(\Omega)$, $\mu = \sum_{i=1}^{n} \delta_i$ ($\mu(A) = \#A$, "abzählendes Maß", "Zählmaß").

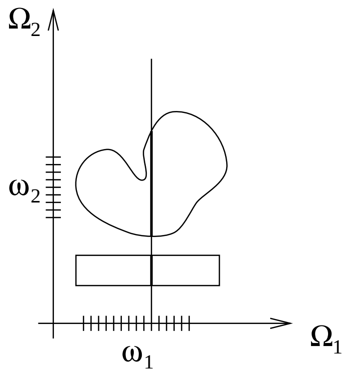
Dann: $\int a d\mu = \sum_{i=1}^{n} a_i$, $a = (a_1, \ldots a_n)$, also wird Cauchy-Schwarz zu $|\sum_{i=1}^{n} a_i b_i| \leq (\sum_{i=1}^{n} a_i^2)^{\frac{1}{2}} (\sum_{i=1}^{n} b_i^2)^{\frac{1}{2}}$.

Kapitel 9

Produktmaß und Unabhängigkeit

9.1 Das Produkt von zwei Maßräumen

Für $i = 1, 2$ sei $(\Omega_i, \mathfrak{A}_i, \mu_i)$ ein Maßraum. Sei $\Omega := \Omega_1 \times \Omega_2$.



Definition 9.1 Das <u>Produkt</u> $\mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2$ der σ-Algebren $\mathfrak{A}_1$ und $\mathfrak{A}_2$ über Ω ist die von der Rechteckmengen $A_1 \times A_2,\ A_i \in \mathfrak{A}_i$ erzeugte σ-Algebra.

Lemma 9.2 Für alle $A \in \mathfrak{A},\ \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2$ gilt: $A_{\omega_1} := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\} \in \mathfrak{A}_2,\ A_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\} \in \mathfrak{A}_1$. Man nennt A_{ω_1} den ω_1-Schnitt aus A.

Beweis: Sei $\omega_1 \in \Omega_1$. Wegen $\Omega_{\omega_1} = \Omega_2$ gilt $\Omega \in \mathfrak{A}' := \{A \in \mathfrak{A} : A_{\omega_1} \in \mathfrak{A}_2\}$.

$(\Omega \setminus A)_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \notin A\} = (\{\omega_2 : (\omega_1, \omega_2) \in A\})^c = (A_{\omega_1})^c$ liefert "Komplementstabilität" von $\mathfrak{A}'$.

$(\bigcup\limits_{n=1}^{\infty} A_n)_{\omega_1} = \bigcup\limits_{n=1}^{\infty} (A_n)_{\omega_1}$. Damit: abzählbare Vereinigungen führen nicht aus $\mathfrak{A}$' heraus.

Insgesamt: $\mathfrak{A}$' ist eine σ-Algebra.

$(A_1 \times A_2)_{\omega_1} = \begin{cases} A_2 & , \quad \omega_1 \in A_1 \\ \emptyset & , \quad \text{sonst} \end{cases}$.

Also enthält $\mathfrak{A}'$ ein Erzeugendensystem von $\mathfrak{A}$, und damit insgesamt: $\mathfrak{A}' \supset \mathfrak{A}$. □

Definition 9.3 Ein Maß μ auf dem meßbaren Raum $(\Omega, \mathfrak{A})$ heißt <u>σ-endlich</u>, wenn es eine Folge $(A_n)_{n\in\mathbb{N}} \subseteq \mathfrak{A}$ gibt mit $\mu(A_n) < \infty\ \forall n \in \mathbb{N},\ \bigcup\limits_{n=1}^{\infty} A_n = \Omega$.

Beispielsweise ist das Lebesgue-Maß l auf $(\mathbb{R}, \mathfrak{B})$ σ-endlich: $l([-n, n]) = 2n,\ \bigcup\limits_{n=1}^{\infty} [-n, n] = \mathbb{R}$.

Lemma 9.4 Die Maße μ_1 und μ_2 seien σ-endlich.
Dann gilt für alle $A \in \mathfrak{A}$: $\omega_1 \to \mu_2(A_{\omega_1})$ ist $(\mathfrak{A}_1, \overline{\mathfrak{B}})$-meßbar, $\omega_2 \to \mu_1(A_{\omega_2})$ ist $(\mathfrak{A}_2, \overline{\mathfrak{B}})$-meßbar.

Beweis: Da μ_2 σ-endlich ist, existiert eine Folge $(B_n)_{n\in\mathbb{N}} \subset \mathfrak{A}_2$ mit $B_n \uparrow \Omega_2,\ \mu_2(B_n) < \infty\ \forall n$.

Setze $f_A(\omega_1) := \mu_2(A_{\omega_1}),\ f_{A,n}(\omega_1) := \mu_2(A_{\omega_1} \cap B_n)$.

Sei $\mathfrak{D}$ die Menge aller $A \in \mathfrak{A}$ mit $(\mathfrak{A}_1, \mathfrak{B})$-meßbaren f_A. Man kann leicht zeigen: $\mathfrak{D}$ ist ein Dynkin-System, das den $\cap$-stabilen Erzeuger $\{A_1 \times A_2 : A_i \in \mathfrak{A}_i\}$ von $\mathfrak{A}_1 \otimes \mathfrak{A}_2$ enthält, also $\mathfrak{D} = \mathfrak{A}$.

Also: $f_{A,n}$ ist meßbar $\forall A \in \mathfrak{A},\ n \in \mathbb{N}$ und mit $f_A = \sup f_{A,n}$ folgt die Behauptung. □

Satz und Definition 9.5 Sind $\mu_1,\ \mu_2$ σ-endlich, so existiert genau ein Maß $\mu = \mu_1 \otimes \mu_2$ auf $\mathfrak{A}_1 \otimes \mathfrak{A}_2$, das <u>Produktmaß</u> von μ_1 und μ_2 mit der Eigenschaft $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \forall A_i \in \mathfrak{A}_i$.
Weiter gilt für alle $A \in \mathfrak{A},\ \mu(A) = \int \mu_2(A_{\omega_1})\mu_1(d\omega_1) = \int \mu_1(A_{\omega_2})\mu_2(d\omega_2)$.

Beweis: $f_A(\omega_1) := \mu_2(A_{\omega_1})$. Sind $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$ paarweise disjunkt, $A = \sum\limits_{n=1}^{\infty} A_n$, so gilt:

$$\int f_A d\mu_1 \overset{\mu_2 \text{ stetig}}{\underset{\text{von unten}}{=}} \int \lim_{n\to\infty} f_{\bigcup_{i=1}^n A_i} d\mu_1 \overset{\text{monotone}}{\underset{\text{Konvergenz}}{=}} \lim_{n\to\infty} \int f_{\bigcup_{i=1}^n A_i} d\mu_1 = \lim_{n\to\infty} \sum_{i=1}^{n} \int f_{A_i} d\mu_1 = \sum_{i=1}^{\infty} \int f_{A_i} d\mu_1.$$

Außerdem $\int f_\emptyset d\mu_1 = 0$. Also ist $\Pi : \mathfrak{A} \to [0, \infty],\ \Pi(A) := \int f_A d\mu_1$ ein Maß auf $\mathfrak{A}$. Für Π gilt:
$\Pi(A_1 \times A_2) = \int f_{A_1 \times A_2} d\mu_1 = \int \mu_2(A_2) 1_{A_1} d\mu_1 = \mu_2(A_2) \cdot \mu_1(A_1)$.

Analog wird durch $\Pi'(A) = \int \mu_1(A_{\omega_2})\mu_2(d\omega_2)$ ein Maß auf $\mathfrak{A}$ definiert, das auf den Rechteckmengen mit Π übereinstimmt. Das System der Rechteckmengen ist ein $\cap$-stabiler Erzeuger von $\mathfrak{A}$, Eindeutigkeit folgt mit Aufgabe 2 (hier geht die σ-Endlichkeit von $\mu_1,\ \mu_2$ ein.)

Mit μ_1 und μ_2 ist auch $\mu_1 \otimes \mu_2$ σ-endlich. Wie integriert man bzgl. $\mu_1 \otimes \mu_2$?

Ist $f : \Omega \to \mathbb{R}$ eine Abbildung, so schreiben wir $f_{\omega_1},\ f_{\omega_2}$ für die Abbildung

$f_{\omega_1} : \Omega_2 \to \mathbb{R},\ f_{\omega_1}(\omega_2) := f(\omega_1, \omega_2),\ f_{\omega_2} : \Omega_1 \to \mathbb{R},\ f_{\omega_2}(\omega_1) := f(\omega_1, \omega_2).$

Setzt man $f = 1_A$, so sieht man, das dies den Schnittbegriff bei Mengen verallgemeinert.

Lemma 9.6

Ist f $(\mathfrak{A}, \mathfrak{B})$-meßbar, so ist f_{ω_1} $(\mathfrak{A}_2, \mathfrak{B})$-meßbar für alle $\omega_1 \in \Omega_1$ und f_{ω_2} $(\mathfrak{A}_1, \mathfrak{B})$-meßbar für alle $\omega_2 \in \Omega_2$.

Beweis: $f_{\omega_1}^{-1}(B) = \{\omega_2 \in \Omega_2 : f_{\omega_1}(\omega_2) \in B\} = (\{\omega \in \Omega : f(\omega) \in B\})_{\omega_1} = (\underbrace{f^{-1}(B)}_{\in \mathfrak{A}})_{\omega_1} \overset{\text{Lemma 9.2}}{\in} \mathfrak{A}_2.$ □

Satz 9.7 (Fubini I) $\mu_1,\ \mu_2$ seien σ-endlich, $f : \Omega \to \mathbb{R}$ nicht-negativ und $(\mathfrak{A}, \mathfrak{B})$-meßbar.

Dann ist $\omega_1 \to \int f_{\omega_1} d\mu_2$ $(\mathfrak{A}_1, \overline{\mathfrak{B}})$-meßbar, $\omega_2 \to \int f_{\omega_2} d\mu_1$ $(\mathfrak{A}_2, \overline{\mathfrak{B}})$-meßbar, und es gilt:

$\int f d(\mu_1 \otimes \mu_2) = \int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \int (\int f_{\omega_1} d\mu_2) \mu_1(d\omega_1).$

Beweis: Im Falle $f = \sum\limits_{i=1}^{n} a_i 1_{A_i}$ erhält man $\int f_{\omega_2} d\mu_1 = \sum\limits_{i=1}^{n} a_i \int 1_{(A_i)_{\omega_2}} d\mu_1 = \sum\limits_{i=1}^{n} a_i \mu_1((A_i)_{\omega_2})$, also die Meßbarkeit von

$\omega_2 \to \int f_{\omega_2} d\mu_1$, und $\int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \sum\limits_{i=1}^{n} a_i \int \mu_1((A_i)_{\omega_2}) \mu_2(d\omega_2) = \sum\limits_{i=1}^{n} a_i \mu_1 \otimes \mu_2(A_i) = \int f d(\mu_1 \otimes \mu_2)$.

Ist $f \geq 0$ $(\mathfrak{A}, \mathfrak{B})$-meßbar, so existiert eine Folge $(f_n)_{n \in \mathbb{N}} \subset Prim_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f,\ \int f d\mu = \lim\limits_{n \to \infty} \int f_n d\mu$.

Wegen $(f_n)_{\omega_2} \uparrow f_{\omega_2}$ und $(f_n)_{\omega_2} \in Prim_+(\Omega, \mathfrak{A})$, folgt mit monotoner Konvergenz $g_n(\omega_2) := \int (f_n)_{\omega_2} d\mu_1 \uparrow \int f_{\omega_2} d\mu_1 \quad \forall \omega_2 \in \Omega_2$.

Nach dem bereits bewiesenen Teil gilt $\int g_n d\mu_2 = \int f_n d\mu$.

Wieder monotone Konvergenz liefert $\int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \lim\limits_{n \to \infty} \int g_n d\mu_2 = \lim\limits_{n \to \infty} \int f_n d\mu = \int f d\mu$.

Wiederholt man dies mit ω_1 anstelle von ω_2, so folgt der Rest der Behauptung. □

(Marginale) Erweiterung des Integralbegriffs:

Ist $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum, $A \in \mathfrak{A}$ mit $\mu(A^c) = 0,\ f : \Omega \to \mathbb{R}$, so nennen wir f $(\mathfrak{A}, \mathfrak{B})$-meßbar, μ-integrierbar, etc. ..., wenn dies auf $\tilde{f} : \Omega \to \mathbb{R},\ \tilde{f}(\omega) := \begin{cases} f(\omega) & , \quad \omega \in A \\ 0 & , \quad \text{sonst} \end{cases}$ zutrifft.

Wir schreiben dann $\int f d\mu$ anstelle von $\int \tilde{f} d\mu$. Hierdurch wird die Integration von Funktionen möglich, die nur fast überall definiert sind (Rechenregeln bleiben erhalten.)

Satz 9.8 (Fubini II)

$\mu_1,\ \mu_2$ σ-endlich, $\mu := \mu_1 \otimes \mu_2,\ f : \Omega \to \mathbb{R}$ sei μ-integrierbar. Dann sind μ_1-fast-alle f_{ω_1} μ_2-integrierbar, μ_2-fast-alle f_{ω_2} μ_1-integrierbar und es gilt: $\int f d\mu = \int (\int f_{\omega_2} d\mu_1) \mu_2(d\omega_2) = \int (\int f_{\omega_1} d\mu_2) \mu_1(d\omega_1)$.

Beweis: Reduktion auf Fubini I mit $f = f^+ - f^-$, etc. ... □

Man kann Fubini-Formel auch in der Form $\int f d\mu = \iint f(\omega_1, \omega_2) \mu_1(d\omega_1) \mu_2(d\omega_2) = \iint f(\omega_1, \omega_2) \mu_2(d\omega_2) \mu_1(d\omega_1)$ schreiben. Also: Integration bzgl. des Produktmaßes kann nacheinander über die Einzelkomponenten ausgeführt werden; Die Reihenfolge ist dabei egal.

9.2 Produkte von mehr als zwei Maßräumen

$(\Omega_i, \mathfrak{A}_i, \mu_i)$ σ-endliche Maßräume für $i = 1, \ldots, n$.

$(\ldots((\mathfrak{A}_1 \otimes \mathfrak{A}_2) \otimes \mathfrak{A}_3) \otimes \ldots) \otimes \mathfrak{A}_n$ ist eine σ-Algebra auf $(\ldots((\Omega_1 \times \Omega_2) \times \Omega_3) \times \ldots) \times \Omega_n$.

Identifiziert man dies mit $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$, so erhält man, daß die Produkt-σ-Algebra von den Rechteckmengen $A_1 \times A_2 \times \ldots \times A_n,\ A_i \in \mathfrak{A}_i$ für $i = 1. \ldots, n$ erzeugt wird und schreibt kurz: $\mathfrak{A} := \mathfrak{A}_1 \otimes \ldots \otimes \mathfrak{A}_n$ ($\otimes$ ist assoziativ.) Mit Induktion erhält man die Existenz (und Eindeutigkeit) eines Maßes μ mit $\mu(A_1 \times \ldots \times A_n) = \mu_1(A_1) \cdot \ldots \cdot \mu_n(A_n)$.

Schreibweise: $\mathfrak{A} = \bigotimes\limits_{i=1}^{n} \mathfrak{A}_i,\ \mu = \bigotimes\limits_{i=1}^{n} \mu_i$.

Der Satz von Fubini gilt in der entsprechenden Form.

Beispiel 9.9 $(\Omega_i, \mathfrak{A}_i, \mu_i) = (\mathbb{R}, \mathfrak{B}, l)$ für $i = 1, \ldots, n$.

$\mathop{\times}\limits_{i=1}^{n} \Omega_i = \mathbb{R}^n,\ \bigotimes\limits_{i=1}^{n} \mathfrak{B}\ (= \mathfrak{B}^{\otimes n}) =: \mathfrak{B}^n$ ist die σ-Algebra der n-dimensionalen Borel-Mengen und $l^n := l^{\otimes n}$ ist das n-dimensionale Lebesgue-Maß.

$\mathfrak{B}^n$ wird erzeugt von $I^n := \{(a, b] : a, b \in \mathbb{R}^n\},\ (a, b] := \{(x_1, \ldots, x_n) : a_i < x_i \leq b_i$ für $i = 1, \ldots, n\}$.

WMaße P auf $(\mathbb{R}^n, \mathfrak{B}^n)$ lassen sich durch die Verteilungsfunktion $F : \mathbb{R}^n \to [0, 1],\ F(x) := P((-\infty, x])$ charakterisieren.

Ist $(\Omega, \mathfrak{A}, P)$ ein WRaum und $X : \Omega \to \mathbb{R}^n$ $(\mathfrak{A}, \mathfrak{B}^n)$-meßbar, so heißt X (n-dimensionaler) Zufallsvektor, die Verteilungsfunktion zu $F_X(x) := P^X((-\infty, x]) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$ heißt auch Verteilungsfunktion zu X, usw.

Hat P^X eine Dichte f bzgl. l^n, so nennt man f eine Dichte zu X; X heißt dann absolutstetig verteilt.

Schwieriger: Produkte von unendlich vielen Maßräumen.

Einschränkung auf WRäume; Sei also $I \neq \emptyset$ und für jedes $i \in I$ sei $(\Omega_i, \mathfrak{A}_i, P_i)$ ein WRaum.

Für jedes $J \subset I$ wird durch $\Pi_J : \Omega \to \mathop{\times}\limits_{i \in J} \Omega_i,\ \Pi_J(\omega) := \omega|_J$ die Projektion auf die J-Koordinaten definiert.

Die Produkt-σ-Algebra $\mathfrak{A} := \bigotimes\limits_{i \in I} \mathfrak{A}_i$ auf Ω ist die von den Zylindermengen mit endlicher Basis $\Pi_J^{-1}(A_J),\ A_J \in \bigotimes\limits_{i \in J} \mathfrak{A}_i,\ J$

endlich, erzeugte σ-Algebra.
$\mathfrak{A}$ wird auch von den Mengen $\Pi_i^{-1}(A_i) := \Pi_{\{i\}}^{-1}(A_i)$, $A_i \in \mathfrak{A}_i$ erzeugt. Dieses System ist aber nicht $\cap$-stabil. Die Produkt-σ-Algebren werden von den Projektionen erzeugt.

Satz 9.10 Auf $(\Omega, \mathfrak{A})$ existiert genau ein WMaß $P = \bigotimes_{i\in I} P_i$ mit $P^{\Pi_J} = \bigotimes_{i\in J} P_i$ für alle endlichen $J \subset I$.

Beweis: Ähnlich wie bei der Existenz von l, siehe Bauer. $\square$

Münzwurf: $(\{0,1\}, \mathbb{P}(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1))$.
Der unendlich oft wiederholte Münzwurf: $(\{0,1\}, \mathbb{P}(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1))^{\mathbb{N}}$, $\Omega = \{0,1\}^{\mathbb{N}}$ Menge aller 0-1-Folgen.
Durch $\{0,1\}^{\mathbb{N}} \ni (\omega_i)_{i\in\mathbb{N}} \xrightarrow{T} \sum_{i=1}^{\infty} \omega_i \cdot 2^{-i} \in [0,1]$ wird eine $(\mathfrak{A}, \mathfrak{B}_{[0,1]})$-meßbare Abbildung mit der Eigenschaft $((\frac{1}{2}(\delta_0 + \delta_1))^{\otimes\mathbb{N}})^T = l|_{[0,1]}$ (siehe auch Bsp. 5.30) definiert.

9.3 Unabhängigkeit

Erinnerung (Def. 5.27):
- $\{\mathfrak{A}_i : i \in I\}$ heißt unabhängig, wenn $P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i)$, $A_i \in \mathfrak{A}_i$, $\forall J \subset I$, J endlich, gilt.
- Ist $(\Omega_i, \mathfrak{A}_i)$ ein meßbarer Raum $\forall i \in I$ und $X_i : \Omega \to \Omega_i$ $(\mathfrak{A}, \mathfrak{A}_i)$-meßbar, so heißt $\{X_i : i \in I\}$ unabhängig, wenn die σ-Algebren $\sigma(X_i) = \{X_i^{-1}(A_i) : A_i \in \mathfrak{A}_i\}$ unabhängig sind.

Klar: Eine Familie von σ-Algebren/Zufallsgrößen ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist. Beim Nachweis der Unabhängigkeit kann man sich auf $\cap$-stabile Erzeugendensysteme beschränken ($\rightsquigarrow$ Satz 5.28.)

Beispiel 9.11 ("∞-oft wiederholter Münzwurf")
Es sei $\Omega = \{0,1\}^{\mathbb{N}}$, $\mathfrak{A} = \mathbb{P}(\{0,1\})^{\otimes\mathbb{N}}$, $P = (\frac{1}{2}(\delta_0 + \delta_1))^{\otimes\mathbb{N}}$. Definiere $X_n : \Omega \to \mathbb{R}$ durch $X_n((\omega_i)_{i\in\mathbb{N}}) := \omega_n$ (Maßtheorie: Projektion auf die n-te Koordinate, Stochastik: Resultat des n-ten Wurfes). X_n ist nach Konstruktion meßbar.
Klar: $P(X_n = 0) = P(X_n = 1) = \frac{1}{2}$.
Ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{A} = \sigma(X_i)$ ist $\mathfrak{E}_i := \{A_i\}$ mit $A_i = \{X_i = 1\}$.
Ist $J \subset \mathbb{N}$ endlich, $J = \{i_1, \dots, i_n\}$, so gilt:

$$P(\bigcap_{i\in J} A_i) = P(X_{i_1} = 1, \dots X_{i_n} = 1) = P^{\Pi_J}(\{1\} \times \dots \times \{1\}) = (\tfrac{1}{2})^n = \prod_{j=1}^{n} P(X_{i_j} = 1).$$

Also: $\{X_n : n \in \mathbb{N}\}$ ist eine unabhängige Familie. Außerdem ist $X := (X_1, \dots, X_n)$ ein n-dimensionaler Zufallsvektor ($\rightsquigarrow$ Aufgabe 15). Die Verteilung von X heißt gemeinsame Verteilungsfunktion/Dichte von $X_1, \dots, X_n$.

Satz 9.12 (verallgemeinert Satz 4.18)
Es seien $X_1, \dots, X_n$ ZV auf $(\Omega, \mathfrak{A}, P)$.
(i) Dann gilt: $X_1, \dots, X_n$ sind unabhängig $\iff$ $P^{(X_1,\dots,X_n)} = P^{X_1} \otimes \dots \otimes P^{X_n}$.
(ii) Es sei F_i die Verteilungsfunktion zu X_i $(1 \le i \le n)$, F die gemeinsame Verteilungsfunktion. Dann gilt:

$$X_1, \dots, X_n \text{ unabhängig} \iff F(x_1, \dots, x_n) = \prod_{i=1}^{n} F_i(x_i) \qquad \forall x \in \mathbb{R}^n.$$

(iii) Es sei f_i eine Dichte von X_i; $f(x) := \prod_{i=1}^{n} f_i(x)$ für $x = (x_1, \dots, x_n)$.
Dann $X_1, \dots, X_n$ sind unabhängig $\iff$ f ist gemeinsame Dichte.

Beweis: $J := \{(-\infty, a] : a \in \mathbb{R}\}$ ist ein $\cap$-stabiles Erzeugendensystem zu $\mathfrak{B}$, also ist $\{X_i^{-1}((-\infty, a]) : a \in \mathbb{R}\}$ ein $\cap$-stabiles Erzeugendensystem zu $\sigma(X_i)$.
(i) und (ii) folgen aus Satz 5.28 und der Definition des Produktmaßes.
Teil (iii) folgt mit Fubini I aus Teil (ii). $\square$

Bekannt: Funktionen von unabhängigen Zufallsgrößen sind wieder unabhängig.

Satz 9.13 (Multiplikationssatz, verallgemeinert Satz 4.19)
Für unabhängige Zufallsvariablen X, Y gilt $E(XY) = (EX)(EY)$ (vorausgesetzt die Erwartungswerte existieren).

Beweis: $EXY = \int X(\omega)Y(\omega)dP =^1 \int xyP^{(X,Y)}(dx,dy) \overset{9.12(i)}{=} \int xyP^X \otimes P^Y(dx,dy) \overset{\text{Fubini II}}{=} \int(\int xyP^X(dx))P^Y(dy) = \int y(\int xP^X(dx))P^Y(dy) = EX \int yP^Y(dy) = (EX)(EY)$. $\square$

In Verallgemeinerung von Def. 4.21 nennen wir den Erwartungswert
$cov(X,Y) = E((X - EX)(Y - EY))$ $(= E(XY) - (EX)(EY))$ die Kovarianz der ZV X und Y.
Im Falle $cov(X,Y) = 0$ heißen X und Y unkorreliert.
Im Falle $var(X) > 0$, $var(Y) > 0$ heißt $\rho(x,y) := \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}}$ der Korrelationskoeffizient.

$$\text{Unabhängig} \underset{\not\Leftarrow}{\Rightarrow} \text{Unkorreliert.}$$

$$var(X_1 + \dots + X_n) = \sum_{i=1}^{n} var(X_i) + \sum_{\substack{i,j=1\\ i\neq j}}^{n} cov(X_i, X_j) \qquad \rightsquigarrow \text{Bienaymé.}$$

[1] Transformationsformel, Verlagerung der Integration auf den Bildbereich

Satz 9.14 (Das Borel-Cantelli-Lemma)
$(\Omega, \mathfrak{A}, P)$ WRaum, $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$ eine Folge von Ereignissen; $\limsup\limits_{n\to\infty} A_n = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$ (unendlich viele der A_k's treten ein, $\longrightarrow$ Bsp. 1.2(i)). Dann gilt:

(i) $\sum\limits_{n=1}^{\infty} P(A_n) < \infty \implies P(\limsup\limits_{n\to\infty} A_n) = 0$

(ii) $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$ und $\{A_n | n \in \mathbb{N}\}$ unabhängig $\implies P(\limsup\limits_{n\to\infty} A_n) = 1.$

Beweis: (i) Aufgabe 5c) zur Stochastik I.
(ii) Sei $p_n := P(A_n)$. Für alle $n \in \mathbb{N}$ gilt (mit $1 - p \le \exp(-p) \quad \forall p \in \mathbb{R}$):

$$0 \le P(\bigcap_{k=n}^{\infty} A_k^{\,c}) \overset{\text{Stetigkeit}}{\underset{\text{von unten}}{=}} \lim_{N\to\infty} P(\bigcap_{k=n}^{N} A_k^{\,c}) = \lim_{N\to\infty} \prod_{k=n}^{N} (1 - p_k) \le \lim_{N\to\infty} \exp(-\sum_{k=n}^{N} p_k) = 0.$$

also folgt: $0 \le P((\limsup\limits_{n\to\infty} A_n)^{\,c}) = P(\bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k^{\,c}) \le \sum\limits_{n=1}^{\infty} P(\bigcap\limits_{k=n}^{\infty} A_k^{\,c}) = 0.$ □

Kapitel 10

Gesetze der großen Zahlen

10.1 Konvergenz fast-sicher und Konvergenz in Wahrscheinlichkeit

Gesetz der großen Zahlen gibt es in "starker" und "schwacher" Form; sie beziehen sich auf unterschiedliche Konvergenzarten.

Definition 10.1 Es seien $X,\ X_1,\ X_2,\dots$ ZV auf $(\Omega, \mathfrak{A}, P)$.

(i) X_n konvergiert P-fast-sicher gegen X, Schreibweise: $X_n \xrightarrow{\text{f.s.}} X$, wenn gilt:
$P(\{\omega \in \Omega : \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

(ii) X_n konvergiert in Wahrscheinlichkeit gegen X, Schreibweise $X_n \xrightarrow{P} X$, wenn gilt:
$\lim\limits_{n\to\infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}) = 0$ für alle $\epsilon > 0$.

Satz 10.2 (Konvergenz fast-sicher impliziert Konvergenz in Wahrscheinlichkeit, die Umkehrung gilt nicht)

Beweis: $\xrightarrow{\text{f.s.}} \implies \xrightarrow{P}$

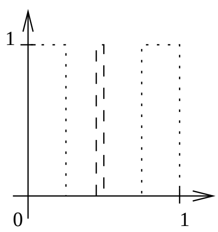
Sei $N := \{\omega \in \Omega : X_n(\omega) \not\to X(\omega)\}$, nach Voraussetzung: $P(N) = 0$.
Sei $\epsilon > 0$ und setze $A_n := \{\omega \in \Omega : \sup\limits_{m \geq n} |X_m(\omega) - X(\omega)| \geq \epsilon\}$.
Klar: $A_n \supseteq A_{n+1}$. Für reelle Zahlenfolgen $(a_n)_{n\in\mathbb{N}}$ gilt $\lim\limits_{n\to\infty} a_n = a \iff \lim\limits_{n\to\infty} \sup\limits_{m\geq n} |a_m - a| = 0$, also:

$$A_n \downarrow N_\epsilon \subseteq \{\omega \in \Omega : \limsup |X_n(\omega) - X(\omega)| \geq \epsilon\} \subseteq N \qquad (N_\epsilon := \bigcap_{n=1}^{\infty} A_n)$$

Also: $0 \leq P(|X_n - X| \geq \epsilon) \leq P(A_n) \longrightarrow P(N_\epsilon) \leq P(N) = 0$.

Gegenbeispiel zu $\xrightarrow{P} \implies \xrightarrow{\text{f.s.}}$: $(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}|_{[0,1)}, l|_{[0,1)})$.

Jedes $n \in \mathbb{N}$ läßt sich eindeutig in der Form $n = 2^m + k,\ m \in \mathbb{N}_0,\ k \in \{0, \dots, 2^m - 1\}$ schreiben. Sei $X_n := 1_{[k\cdot 2^{-m}, (k+1)\cdot 2^{-m})},\ X = 0$.
Dann gilt $\forall \epsilon > 0 :\ P(|X_n - X| > \epsilon) = P([k \cdot 2^{-m}, (k+1)\cdot 2^{-m})) = 2^{-m} \overset{n\to\infty}{\longrightarrow} 0$, also $X_n \xrightarrow{P} X$.
Andererseits gibt es zu jedem $x \in [0,1)$ und $m \in \mathbb{N}$ ein $k \in \{0, \dots, 2^m - 1\}$ mit $x \in [k \cdot 2^{-m}, (k+1)\cdot 2^{-m}),\ x \notin [j \cdot 2^{-m}, (j+1) \cdot 2^{-m})$ für $j \neq k$, d.h. die Folge $(X_n(x))_{n\in\mathbb{N}}$ nimmt ∞-oft den Wert 0 und ∞-oft den Wert 1 an.
Also: $P(\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\}) = 0$. □



10.2 Das schwache Gesetz der großen Zahlen

Sammelbegriff für Resultate der Form $\frac{1}{a_n}(\sum\limits_{i=1}^{n} X_i - b_n) \xrightarrow{P} 0$, mit geeigneten $(a_n)_{n\in\mathbb{N}},\ (b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$.
Ein einfaches Beispiel, das den diskreten Fall (Satz 4.28) verallgemeinert.

Satz 10.3 Es sei $E{X_n}^2 < \infty\ \forall n \in \mathbb{N}$, sowie $cov(X_i, X_j) = 0$ für $i \neq j$. Dann gilt:
$\lim\limits_{n\to\infty} \frac{1}{n^2} \sum\limits_{i=1}^{n} var(X_i) = 0 \quad \implies \quad \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - EX_i) \xrightarrow{P} 0.$

Beweis: Sei $Y_n := \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - EX_i)$.

Es gilt $EY_n = 0,\ var(Y_n) = \frac{1}{n^2} \sum\limits_{i=1}^{n} var(X_i) + \underbrace{\sum\limits_{i\neq j} cov(X_i, X_j)}_{=0}$

Mit der Ungleichung von Chebychev: $P(|Y_n| > \epsilon) \leq \frac{1}{\epsilon^2} var(Y_n) \overset{n\to\infty}{\longrightarrow} 0$. □

10.3 Das starke Gesetz der großen Zahlen

Sammelbegriff für Resultate der Form $\frac{1}{a_n}(\sum_{i=1}^n X_i - b_n) \xrightarrow{\text{f.s.}} 0$, wieder mit geeigneten $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$.

Die Partialsummen $S_n := \sum_{i=1}^n X_i$ werden durch $(b_n)_{n\in\mathbb{N}}$ in der 'Lage' und durch $(a_n)_{n\in\mathbb{N}}$ in der 'Skala' verändert. Wichtigstes Beispiel:

Satz 10.4 Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch verteilter ZV mit $E|X_1| < \infty$, so gilt:
$\overline{X_n}\ (= \frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n X_i) \xrightarrow{\text{f.s.}} EX_1$.

Beweis: Wir nehmen zunächst $X_n \geq 0\ \forall n \in \mathbb{N}$ an.
$Y_k := X_k \cdot 1_{\{X_k \leq k\}}$ ('bei k abgeschnitten')
$S_n^* := \sum_{k=1}^n Y_k$. Mit monotoner Konvergenz folgt: $EY_k = EX_k \cdot 1_{\{X_k\leq k\}} = EX_1 \cdot 1_{\{X_1 \leq k\}} \overset{k\to\infty}{\nearrow} EX_1$.

Aus der Analysis bekannt: $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ mit $a_n \to a \implies \frac{1}{n}\sum_{i=1}^n a_i \to a$.[1], d.h.

(1) $\lim_{n\to\infty} \frac{1}{n}ES_n^* = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n EY_k = EX_1$.

(2) Die Y-Variablen sind wieder unabhängig (aber i.A. nicht mehr identisch verteilt), also
$var(S_n^*) = \sum_{i=1}^n var(Y_i) \leq \sum_{i=1}^n EY_i{}^2 = \sum_{k=1}^n EX_k{}^2 \cdot 1_{\{X_k\leq k\}} \leq \sum_{k=1}^n EX_k{}^2 \cdot 1_{\{X_k \leq n\}} = nEX_1{}^2 \cdot 1_{[0,n]}(X_1)$.

Wähle $\alpha > 1$ und setze $m_n := \lfloor \alpha^n \rfloor$, $\Psi(x) := \sum_{n=1}^\infty \frac{1}{m_n}\cdot 1_{[0,m_n]}(x)$, $N(x) = \min\{n : m_n \geq x\}$.

Wegen $\frac{1}{m_n} = \frac{1}{\lfloor\alpha^n\rfloor} \leq \frac{2}{\alpha^n}$ und $\alpha^{N(x)} \geq \lfloor \alpha^{N(x)}\rfloor = m_{N(x)} \geq x$ folgt

(3) $\Psi(x) = \sum_{n=N(x)}^\infty \frac{1}{m_n} \leq 2\sum_{n=N(x)}^\infty \alpha^{-n} = 2\alpha^{-N(x)}\frac{1}{1-\alpha^{-1}} \leq \frac{k}{x}$ mit $k := \frac{2\alpha}{\alpha-1}$.

Mit Chebychev folgt nun für alle $\epsilon > 0$: $\sum_{n=1}^\infty P(\frac{1}{m_n}|S_{m_n}^* - ES_{m_n}^*| \geq \epsilon) \overset{(2)}{\leq} \sum_{n=1}^\infty \frac{1}{\epsilon^2 m_n} EX_1{}^2 \cdot 1_{[0,m_n]}(X_1)$

$= \frac{EX_1{}^2}{\epsilon^2}\sum_{n=1}^\infty \frac{1}{m_n}1_{[0,m_n]}(X_1) = \frac{1}{\epsilon^2}EX_1{}^2\Psi(X_1) \overset{(3)}{\leq} kEX_1 < \infty$.

Hieraus folgt wegen Aufgabe 18(ii) $\frac{1}{m_n}(S_{m_n}^* - ES_{m_n}^*) \longrightarrow 0\quad P$-f.s. und wegen (1) und Aufgabe 18(i)

(4) $\frac{1}{m_n}S_{m_n}^* \longrightarrow EX_1\quad P$-f.s.

Wie wollen "den $*$ loswerden."
$\sum_{n=1}^\infty \underbrace{P(X_n \neq Y_n)}_{P(X_n > n)} = \sum_{n=1}^\infty P(X_1 > n) \leq \int_0^\infty P(X_1 \geq x)dx = EX_1 < \infty$ ($\rightsquigarrow$ Aufgabe 14)

Mit Borel-Cantelli folgt, daß $N_0 := \{\omega \in \Omega : X_n(\omega) \neq Y_n(\omega)$ für unendlich viele $n\}$ eine P-Nullmenge ist. Für $\omega \in X_0{}^c$ existiert somit ein $k(\omega) < \infty$ mit $X_n(\omega) = Y_n(\omega)\ \forall n \geq k(\omega)$. Auf $N_0{}^c$ gilt also $\frac{1}{n}(S_n(\omega) - S_n{}^*(\omega)) = \frac{1}{n}(\sum_{k=1}^{k(\omega)} X_k(\omega) - Y_k(\omega)) \overset{n\to\infty}{\longrightarrow} 0$, d.h. $\frac{1}{n}(S_n - S_n{}^*) \longrightarrow 0$ P.-f.s. und mit (4) folgt

(5) $\frac{1}{m_n}S_{m_n} \longrightarrow EX_1$ P-f.s.

Als nächstes: Einschränkung auf Teilfolgen "loswerden".
$\frac{m_n}{m_{n+1}}\frac{S_{m_n}}{m_n} \leq \frac{S_k}{k} \leq \frac{m_{n+1}}{m_n}\frac{S_{m_{n+1}}}{m_{n+1}}$ für $m_n \leq k \leq m_{n+1}$ $\quad((S_k)_{k\in\mathbb{N}}$ hat Zuwächse $\geq 0)$

Mit $\frac{m_n}{m_{n+1}} \overset{n\to\infty}{\longrightarrow} \frac{1}{\alpha}$ folgt $\begin{array}{ccccc} \frac{1}{\alpha}EX_1 & \leq & \liminf_{k\to\infty}\frac{S_k}{k} & & \\ & & \limsup_{k\to\infty}\frac{S_k}{k} & \leq & \alpha EX_1 \end{array}$ P-f.s. $\forall \alpha > 1$ wegen (5).

Sei N_α die zugehörige Ausnahme-Nullmenge. Außerhalb der Nullmenge $\bigcup_{j=1}^\infty N_{1+\frac{1}{j}}$ gilt dann (mit $\alpha \to 1$)

$EX_1 \leq \liminf \frac{S_k(\omega)}{k} \leq \limsup \frac{S_k(\omega)}{k} \leq EX_1$, also

(6) $\overline{X_n} = \frac{1}{n}S_n \longrightarrow EX_1$ P-f.s.

Es bleibt die Einschränkung auf $X_i \geq 0$ zu beseitigen:
$\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i{}^+ - \frac{1}{n}\sum_{i=1}^n X_i{}^- \overset{(6)}{\longrightarrow} EX_1{}^+ - EX_1{}^- = EX_1$. □

Beispiel 10.5 (siehe auch Beispiel 9.11)
$(\Omega, \mathfrak{A}, P) = (\{0,1\}, \mathbb{P}(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1))^{\otimes\mathbb{N}}$ Modell für ∞-oft wiederholten Wurf einer fairen Münze.
X_n : Projektion auf n-te Koordinate (Resultat von Wurf Nr. n)
$(X_n)_{n\in\mathbb{N}}$ genügt den Voraussetzungen von Satz 10.4, also $\frac{1}{n}\#\{1 \leq i \leq n : X_i = 1\}$[2] $\longrightarrow EX_1 = \frac{1}{2}$ P-f.s.

[1] Césara-Mittelwerte haben den gleichen limes.
[2] $\frac{1}{n}S_n$, Anzahl der Kopfwürfe in den ersten n Versuchen.

In maßtheoretischer Sprechweise: Das Produktmaß P ist auf die Menge der 0-1-Folgen konzentriert, bei denen die relative Häufigkeit von Kopf gegen $\frac{1}{2}$ konvergiert.
In wtheoretischer Sprechweise: Die W. dafür, daß die relative Häufigkeit des Ereignisses Kopf gegen seine W. konvergiert, ist 1.
(Läßt sich auf $(1-p)\delta_0 + p\delta_1$ verallgemeinern, d.h. obige Aussagen gelten für Ereignisse mit beliebiger W. p.)

Beispiel 10.6 (siehe auch Aufgabe 28 zur Stochastik I)
Es sei $p \in \mathbb{N},\ p > 1$. Zu jedem $x \in [0,1)$ existiert eine eindeutige Entwicklung zur Basis p,

$$x = \sum_{k=1}^{\infty} a_k p^{-k},\ a_k \in \{0,\dots,p-1\},\ \#\{k : a_k \neq p-1\} = \infty.$$

Es sei B die Menge der zur Basis p normalen Zahlen, $B := \{x \in [0,1) : \lim_{n\to\infty} \frac{1}{n}\#\{1 \leq k \leq n : a_k = q\} = \frac{1}{p}$ für $q = 0,\dots,p-1\}$ (grob: alle möglichen Ziffern kommen gleichoft vor.)
$B \in \mathfrak{B}$ folgt mit Aufgabe 28 zur Stochastik I, also $l(B) =?$
$(\Omega,\mathfrak{A},P) = ([0,1), \mathfrak{B}|_{[0,1)}, l|_{[0,1)})$. Für jedes $n \in \mathbb{N},\ \omega \in \Omega$ sei $X_n(\omega)$ die n-te Ziffer der Entwicklung von ω. Dann ist $(X_n)_{n\in\mathbb{N}}$ eine iid Folge mit $P(X_n = q) = \frac{1}{p}$ für $q = 0,\dots,p-1$.
Sei $q \in \{0,\dots,p-1\},\ Y_n := 1_{\{q\}}(X_n)$. Dann ist $(Y_n)_{n\in\mathbb{N}}$ wieder eine iid Folge und es gilt:

$$P(B_q) = 1 \text{ mit } B_q = \{\omega \in \Omega : \lim_{n\to\infty} \underbrace{\frac{1}{n}\sum_{k=1}^{n} Y_k(\omega)}_{\frac{1}{n}S_n} = \underbrace{\frac{1}{p}}_{EY_1}\}$$

Wegen $\sum_{k=1}^{n} Y_k(\omega) = \#\{1 \leq k \leq n : a_k = q\}$ gilt $B = \bigcap_{q=0}^{p-1} B_q$, also $P(B) = 1$. Beachte: p war beliebig!
Daher: Satz von Borel: l-fast alle Zahlen sind normal in allen Basen $p > 1$.

Kapitel 11

Charakteristische Funktion

11.1 Grundlegendes

$(\Omega, \mathfrak{A}, P)$, $f : \Omega \to \mathbb{C}$, $f = g + ih$ (g, h $\mathbb{R}$-wertig)
$\int f d\mu := \int g d\mu + i \int h d\mu$, wenn g, h μ-integrierbar
Linearität, etc. gilt weiter, auch $Re(\int f d\mu) = \int Re(f) d\mu$.

Lemma 11.1 $|\int f d\mu| \leq \int |f| d\mu$.
Beweis: Annahme $\int f d\mu \neq 0$ (sonst ist nichts zu zeigen)
$Re(\overline{\int f d\mu} \cdot f) \leq |\overline{\int f d\mu} \cdot f| = |\int f d\mu| \cdot |f|$.
$|\int f d\mu|^2 = Re(\overline{\int f d\mu} \cdot \int f d\mu) = \int Re(\overline{\int f d\mu} \cdot f) d\mu \leq \int |\int f d\mu| \cdot |f| d\mu$
$\overset{|\int f d\mu| \neq 0}{\Longrightarrow} |\int f d\mu| \leq \int |f| d\mu$. □

Definition 11.2 (Charakteristische Funktion)
X ZV auf $(\Omega, \mathfrak{A}, P)$. Dann heißt $\varphi_X : \mathbb{R} \to \mathbb{C}$, $\varphi_X(\theta) = E \exp(i\theta X)$ die charakteristische Funhktion zu X.

$\varphi_X(\theta) = E \cos(\theta X) + iE \sin(\theta X)$. Dieser Erwartungswert existiert ($|\ldots| \leq 1$).
$E \exp(i\theta X) = \int\limits_{\Omega} \exp(i\theta X(\omega)) P(d\omega) \underset{\text{mationssatz}}{\overset{\text{Transfor-}}{=}} \int\limits_{\mathbb{R}} \exp(i\theta x) P^X(dx)$.

Also: Zufallsvariablen mit derselben Verteilung haben dieselbe charakteristische Funktion.

- Ist X diskret verteilt ($P(X \in A) = 1$ für eine abzählbare Menge $A \subset \mathbb{R}$), so gilt $\varphi_X(\theta) = \sum\limits_{x \in A} \exp(i\theta x) P(X = x)$.
- Ist X absolutstetig verteilt mit Dichte f, so gilt $\varphi_X(\theta) = \int \exp(i\theta x) f(x) dx$. ($\varphi$ ist Fourier-Transformierte zu P^X).

Beispiel 11.3

(i) $X \sim Bin(n,p)$: $\varphi_X(\theta) = \sum\limits_{k=0}^{n} \exp(i\theta k) \binom{n}{k} p^k (1-p)^{n-k} = \sum\limits_{k=0}^{n} \binom{n}{k} (p \exp(i\theta))^k (1-p)^{n-k} = (1 - p + p \exp(i\theta))^n$.

(ii) $X \sim unif(0,1)$: $\varphi_X(\theta) = \int\limits_0^1 \exp(i\theta x) dx = \begin{cases} \frac{1}{i\theta}(\exp(i\theta) - 1) & \text{für } \theta \neq 0 \\ 1 & \text{für } \theta = 0 \end{cases}$

Satz 11.4 Ist φ_X die charakteristische Funktion zu X, so gilt:
(i) $\varphi_X(0) = 1$
(ii) $|\varphi_X(\theta)| \leq 1 \quad \forall \theta \in \mathbb{R}$
(iii) φ_X ist gleichmäßig stetig auf $\mathbb{R}$
(iv) $\varphi_{(-X)} = \overline{\varphi_X}$
(v) $\varphi_{aX+b}(\theta) = \exp(i\theta b) \varphi_X(a\theta)$
Beweis: (i) $\varphi_X(0) = E \exp(0) = 1$
(ii) $|\varphi_X(\theta)| \underset{11.1}{\overset{\text{Lemma}}{\leq}} E|\exp(i\theta X)| = 1$.
(iii) $|\varphi_X(\theta + h) - \varphi_X(\theta)| = |E \exp(i(\theta + h)X) - \exp(i\theta X)| \leq E|\exp(i\theta X)(\exp(ihX) - 1)| \leq E|\exp(ihX) - 1|$
Sei nun $(h_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ mit $h_n \to 0$. Dann gilt für alle $\omega \in \Omega$: $\cos(h_n X(\omega)) \to 1$, $\sin(h_n X(\omega)) \to 0$.
Alle diese Funktionen sind betragsmäßig durch die P-integrierbare Funktion $\omega \mapsto 1$ beschränkt, also folgt mit majorisierter Konvergenz: $\sup\limits_{\theta \in \mathbb{R}} |\varphi_X(\theta + h_n) - \varphi_X(\theta)| \leq E|\cos(h_n X) - 1| + E|\sin(h_n X)| \to 0$.
(iv) $\varphi_{(-X)}(\theta) = E \exp(i\theta(-X)) = E \exp(i(-\theta)X) = E \cos(-\theta X) + iE \sin(-\theta X) = E \cos(\theta X) - iE \sin(\theta X) = \overline{\varphi_X(\theta)}$
(v) $\varphi_{aX+b}(\theta) = E \exp(i\theta(aX + b)) = \exp(i\theta b) E \exp(i\theta(aX)) = \exp(i\theta b) E \exp(i(\theta a)X) = \exp(i\theta b) \varphi_X(a\theta)$. □

11.2 Ableitungen und Momente

Lemma 11.5 Für alle $n \in \mathbb{N}_0$, $y \in \mathbb{R}$ gilt: $|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}| \le \min\{\frac{|y|^{n+1}}{(n+1)!}, \frac{2|y|^n}{n!}\}$.

Beweis: Es reicht, den Fall $y \ge 0$ zu betrachten ($|\overline{z}| = |z|$.) Zunächst:

(1) $e^{iy} = \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^{n+1}}{n!} \int_0^y (y-s)^n e^{is} ds$

Bei $n = 0$: $1 + i\int_0^y e^{is} ds = e^{iy}$. Partielle Integration ergibt:

(2) $\int_0^y (y-s)^n e^{is} ds = \frac{y^{n+1}}{n+1} + \frac{i}{n+1} \int_0^y (y-s)^{n+1} e^{is} ds$ liefert Schritt von n nach $n+1$.

$|\int_0^y (y-s)^n e^{is} ds| \le \int_0^y (y-s)^n ds = \frac{1}{n+1} y^{n+1}$.

Mit (1) folgt die Oberschranke $\frac{y^{n+1}}{(n+1)!}$.

In (2) erhält man mit $n-1$ anstelle von n: $\frac{i}{n}\int_0^y (y-s)^n e^{is} ds = \int_0^y (y-s)^{n-1} e^{is} ds - \frac{y_n}{n}$.

Also liefert (1) auch $e^{iy} = \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!} \int_0^y (y-s)^{n-1}(e^{is} - 1) ds$ und damit

$|e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!}| \le \frac{1}{(n-1)!} \int_0^y (y-s)^{n-1} \underbrace{|e^{is} - 1|}_{\le 2} ds \le \frac{2}{(n-1)!} \frac{y^n}{n}$. □

Satz 11.6 Im Falle $E|X|^k < \infty$ ist φ_X k-mal differenzierbar und es gilt $\varphi_X^{(k)}(\theta) = E(iX)^k \exp(i\theta X)$, insbesondere $\varphi_X^{(k)}(0) = i^k EX^k$. Außerdem ist $\varphi_X^{(k)}$ gleichmäßig stetig und beschränkt.

Beweis: $\frac{\varphi(\theta+h)-\varphi(\theta)}{h} = E\exp(i\theta X)(\frac{\exp(ihX)-1}{h})$

Lemma 11.5 ergibt bei $n = 1$: $|\exp(ihx) - 1 - ihx| \le \min\{\frac{h^2x^2}{2}, 2|hx|\}$, also

$|\frac{\exp(ihx)-1}{h}| \le 3|x|$, $\frac{\exp(ihX)-1}{h} \to iX$ (mit $h \to 0$.)

Gilt nun $E|X| \le \infty$, so läßt sich der Satz von der majorisierten Konvergenz anwenden und liefert

$\varphi'(\theta) = \lim_{h\to 0} \frac{\varphi(\theta+h)-\varphi(\theta)}{h} = E\exp(i\theta X)(iX)$.

Hat man bereits "$E|X|^k < \infty \implies \varphi^{(k)}(\theta) = E(iX)^k \exp(i\theta X)$" gezeigt, so folgt bei $E|X|^{k+1} < \infty$ mit denselben Argumenten: $\varphi^{(k+1)}(\theta) = \lim_{h\to 0} \frac{1}{h}(\varphi^{(k)}(\theta+h) - \varphi^{(k)}(\theta)) = \lim_{h\to 0} E(iX)^k \exp(i\theta X)(\frac{\exp(ihX)-1}{h}) = E(iX)^{k+1}\exp(i\theta X)$.

Rest wie im Beweis zu Satz 11.4 (iii). □

Satz 11.7 Für alle $\theta \in \mathbb{R}$ mit $\lim_{n\to\infty} \frac{|\theta|^n}{n!} E|X|^n = 0$ gilt $\varphi_X(\theta) = \sum_{k=0}^{\infty} \frac{(i\theta)^k}{k!} EX^k$.

Beweis: Lemma 11.5 liefert nach Integration ($y \rightsquigarrow \theta X$) : $|\varphi_X(\theta) - \sum_{k=0}^{n} \frac{(i\theta)^k}{k!} EX^k| \le 2\frac{|\theta|^n}{n!} E|X|^n$. □

Beispiel 11.8

$X \sim N(0,1)$. Alle Momente zu X existieren. Aus Symmetriegründen hat man $EX^{2k+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k+1} \exp(-\frac{x^2}{2}) dx = 0$.

Partielle Intergration liefert $EX^{2k} = (2k-1)EX^{2k-2}$.

$\frac{\theta^{2n}}{(2n)!} EX^{2n} = \theta^{2n} \frac{(2n-1)(2n-3)\cdot\ldots\cdot 3\cdot 1}{(2n)!} = \frac{\theta^{2n}}{2^n \cdot n!}$

$\varphi_X(\theta) = \sum_{k=0}^{\infty} \frac{(i\theta)^{2k}}{(2k)!}(2k-1)\cdot\ldots\cdot 5\cdot 3\cdot 1 = \sum_{k=0}^{\infty} \frac{(-\theta^2)^k}{2^k k!} = \exp(-\frac{\theta^2}{2})$.

$X \sim N(\mu, \sigma^2) \implies \varphi_X(\theta) = \exp(i\mu\theta - \frac{1}{2}\theta^2\sigma^2)$

11.3 Umkehrsätze

Satz 11.9 Es sei X ZV mit charakteristischer Funktion φ. Dann gilt für alle a, b mit $-\infty < a < b < \infty$:

$$\frac{1}{2}P(X=a) + P(a < X < b) + \frac{1}{2}P(X=b) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta} \varphi(\theta) d\theta.$$

Beweis: $S(y) = \int_0^y \frac{1}{x}\sin(x) dx$, $I(T) = \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta} \varphi(\theta) d\theta$

$$\Psi : \mathbb{R} \times [-T,T] \to \mathbb{C},\ \Psi(\theta, x) := \begin{cases} \frac{1}{i\theta}(\exp(-i\theta(a-x)) - \exp(-i\theta(b-x))) & , \theta \ne 0 \\ b-a & , \theta = 0 \end{cases}$$

Mit Lemma 11.5 (für $n = 1$) folgt, daß Ψ stetig ist und $|\Psi| \le b-a$ gilt.

Insbesondere: Ψ ist $P^X \otimes l|_{[-T,T]}$-integrierbar.

Also läßt sich Fubini II anwenden: $I(T) = \frac{1}{2\pi} \int_{-T}^{T} \frac{1}{i\theta}(\exp(-i\theta a) - \exp(-i\theta b))(\int \exp(i\theta x) P^X(dx)) d\theta =$

$\frac{1}{2\pi}\int(\int\limits_{-T}^{T}\frac{1}{i\theta}(\exp(-i\theta(a-x))-\exp(-i\theta(b-x)))d\theta)P^X(dx)$

Mit $\Psi_{a,b,T}(x)=2\int\limits_0^T\frac{1}{\theta}\sin((x-a)\theta)d\theta-2\int\limits_0^T\frac{1}{\theta}\sin((x-b)\theta)d\theta$ hat man $I(T)=\frac{1}{2\pi}\int\Psi_{a,b,T}(x)P^X(dx)$,

$\int\limits_0^T\frac{1}{\theta}\sin(c\theta)d\theta=\mathrm{sgn}(c)S(T|c|)$, $\qquad \mathrm{sgn}(c)=\begin{cases}1 & , \quad > \\ 0 & , \quad c=0 \\ -1 & , \quad <\end{cases}$, also

$\Psi_{a,b,T}(x)=2\mathrm{sgn}(x-a)S(T|x-a|)-2\mathrm{sgn}(x-b)S(T|x-b|)$ (Bekannt: $S(y)\to\frac{\pi}{2}$ mit $y\to\infty$)

$$\Psi_{a,b,T}(x)\overset{T\to\infty}{\longrightarrow}\Psi_{a,b}=\begin{cases}0 & , \quad x<a \\ \pi & , \quad x=a \\ 2\pi & , \quad a<x<b \\ \pi & , \quad x=b \\ 0 & , \quad x>b\end{cases}$$

Da S als Funktion auf $[0,\infty)$ stetig ist und mit $x\to\infty$ gegen einen endlichen Grenzwert strebt, gilt $\sup\limits_{x\geq 0}|S(x)|<\infty$.

Man hat also zu $(\Psi_{a,b,T})_{T>0}$ eine integrierbare Majorante, d.h. es folgt mit Lebesgue (majorisierte Konvergenz) $\lim\limits_{T\to\infty}I(T)=\frac{1}{2\pi}\int\Psi_{a,b}(x)P^X(dx)=\frac{1}{2\pi}(\pi\cdot P(X=a \text{ oder } X=b)+2\pi\cdot P(a<x<b))$. □

Korollar 11.10

Sind X und Y ZV mit derselben charakteristischen Funktion, so haben X und Y dieselbe Verteilung.

Beweis: Man sieht leicht, daß für jede abzählbare Menge D das System $\mathfrak{I}_D:=\{(a,b]: a\leq b,\ a,b\notin D\}$ ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{B}$ ist. Setzt man $D:=\{a\in\mathbb{R}: P(X=a)>0 \text{ oder } P(Y=a)>0\}$, so folgt mit Satz 11.9, daß P^X und P^Y auf $\mathfrak{I}_D$ übereinstimmen.

Sei $A(X):=\{x: P(X=x)>0\}$ die Menge der Atome von X.

Wegen $\#\{x\in\mathbb{R}: P(X=x)\geq\frac{1}{n}\}\leq n$, $A(X)=\bigcup\limits_{n=1}^{\infty}\{x: P(X=x)\geq\frac{1}{n}\}$ ist $A(X)$ und damit $D=A(X)\cup A(Y)$ abzählbar. □

Satz 11.11 Es sei X eine ZV mit charakteristischer Funktion φ.

Gilt $\int|\varphi(\theta)|d\theta<\infty$, so hat X eine stetige Dichte f, die gegeben wird durch $f(x)=\frac{1}{2\pi}\int\exp(-i\theta x)\varphi(\theta)d\theta \quad \forall x\in\mathbb{R}$.

Beweis: Lemma 11.5 liefert $\quad (*)\quad |\frac{1}{i\theta}(\exp(-i\theta a)-\exp(-i\theta b))|\leq|b-a|$.

Ist φ_X l-integrierbar, so ist $|b-a|\cdot|\varphi_X|$ eine integrierbare Majorante für den Grenzübergang in Satz 11.9, es gilt also $\frac{1}{2}P(X=a)+P(a<X<b)+\frac{1}{2}P(X=b)=\frac{1}{2\pi}\int\frac{1}{i\theta}(\exp(-i\theta a)-\exp(-i\theta b))\varphi(\theta)d\theta$.

Mit $(*)$ folgt $P(a<X<b)\leq\frac{1}{2\pi}|b-a|\int|\varphi(\theta)|d\theta$, insbesondere $P(X=x)=\lim\limits_{n\to\infty}P(x-\frac{1}{n}<X<x+\frac{1}{n})=0$, d.h. X hat keine Atome. Sei F die Verteilungsfunktion zu X.

Dann: $F(b)-F(a)=\frac{1}{2\pi}\int\frac{1}{i\theta}(\exp(-i\theta a)-\exp(-i\theta b))\varphi_X(\theta)d\theta \qquad \forall a<b$.

Wegen $(*)$ läßt sich wieder majorisierte Konvergenz anwenden:

$\lim\limits_{h\to 0}\frac{F(x+h)-F(x)}{h}=\frac{1}{2\pi}\int\exp(-i\theta x)\lim\limits_{h\to 0}(\frac{1-\exp(-i\theta h)}{i\theta h})\varphi(\theta)d\theta=\frac{1}{2\pi}\int\exp(-i\theta x)\varphi(\theta)d\theta$. □

11.4 Faltungen

$X,\ Y$ seien unabhängige ZV auf $(\Omega,\mathfrak{A},P)$, $Z:=X+Y$.

Nach Satz 9.12(i): Der Zufallsvektor $(X,Y):\Omega\to\mathbb{R}^2$ hat die Verteilung $P^X\otimes P^Y$, Z ist Funktion von (X,Y). P^Z ist das Bildmaß von $P^X\otimes P^Y$ unter h.

$$\begin{tikzcd} \Omega \arrow[r, "(X,Y)"] \arrow[dr, dashed, "Z"'] & \mathbb{R}^2 \arrow[d, "h(x,y):=x+y"] \\ & \mathbb{R} \end{tikzcd}$$

Definition 11.12 (Faltung)

Es seien $\mu_1,\ \mu_2$ WMaße auf $(\mathbb{R},\mathfrak{B})$. Dann heißt das Bildmaß von $\mu_1\otimes\mu_2$ unter der Abbildung $(x,y)\mapsto x+y$ die Faltung $\mu_1*\mu_2$ von μ_1 und μ_2.

Satz 11.13 Sind X und Y unabhängige ZV auf $(\Omega,\mathfrak{A},P)$, so gilt $P^{X+Y}=P^X*P^Y$.

Satz 11.14

(i) Für alle $A\in\mathfrak{B}$ gilt $\mu_1*\mu_2(A)=\int\mu_1(A-y)\mu_2(dy)=\int\mu_2(A-x)\mu_1(dx)$, wobei $A-x:=\{y-x: y\in A\}$, etc.

(ii) Sind $A_X,\ A_Y$ abzählbar mit $P(X\in A_X)=P(Y\in A_Y)=1$, so gilt $P(Z\in A_Z)=1$ mit
$A_Z=\{x+y: x\in A_X, y\in A_Y\}$, A_Z abzählbar, und
$P(Z=z)=\sum\limits_{x\in A_X}P(X=x)P(Y=z-x)=\sum\limits_{y\in A_Y}P(X=z-y)P(Y=y)$.

(iii) $X,\ Y$ seien unabhängige ZV. Sind $f_X,\ f_Y$ Dichten von X und Y, so ist auch $Z=X+Y$ absolutstetig verteilt, mit Dichte $f_Z(z)=\int f_X(z-y)f_Y(y)dy=\int f_X(x)f_Y(z-x)dx$.

Beweis: (i) Sei $B:=\{(x,y)\in\mathbb{R}^2: x+y\in A\}$ $\qquad B_y=\{x\in\mathbb{R}:(x,y)\in B\}=\{x\in\mathbb{R}: x+y\in A\}=A-y$

Dann gilt $\mu_1*\mu_2(A)=\mu_1\otimes\mu_2(B)\overset{\text{Satz}}{\underset{9.5}{=}}\int\mu_1(B_y)\mu_2(dy)=\int\mu_1(A-y)\mu_2(dy)$.

(ii) $P(Z=z) = \sum_{x\in A_X} P(Z=z, X=x) = \sum_{x\in A_X} P(Y=z-x, X=x) = \sum_{x\in A_X} P(Y=z-x)P(X=x).$[1]

(Diese Aussage zeigt, daß hier der in der Stochastik I eingeführte Faltungsbegriff ($\rightarrow$ Satz und Definition 4.25) verallgemeinert wird.)

(iii) Wir definieren $f : \mathbb{R} \rightarrow \mathbb{R}$ durch $f(z) := \int f_Y(z-x) f_X(x)dx$.

$$\int_{(-\infty,a]} f(z)dz = \int(\int 1_{(-\infty,a]}(z) f_Y(z-x) f_X(x)dx)dz \overset{\text{Fubini}}{\underset{1}{=}} \int(\underbrace{\int 1_{(-\infty,a]}(z) f_Y(z-x)dz}_{\text{Substitution } z'=z-x}) f_X(x)dx =$$

$$\int(\int 1_{(-\infty,a-x]}(z') f_Y(z')dz')P^X(dx) = \int P(\underbrace{Y \leq a-x}_{(-\infty,a]-x})P^X(dx) \overset{\text{Teil (i)}}{=} P^X * P^Y((-\infty,a]) \overset{\text{Satz}}{\underset{11.12}{=}} P(Z \leq a). \quad \square$$

Beispiel 11.15 X, Y unabhängig, $X \sim N(0,\sigma^2)$, $Y \sim N(0,\tau^2)$. Die Verteilung von $X+Y$ ist gesucht.
Satz 11.14 (iii) liefert:

$f_{X+Y}(z) = \int \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}x^2) \frac{1}{\sqrt{2\pi}\tau} \exp(-\frac{1}{2\tau^2}(z-x)^2)dx \overset{\text{quadratische}}{\underset{\text{Ergänzung}}{=}} \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \frac{1}{\sqrt{2\pi}} \int \exp(-\frac{1}{2}(u - z\frac{\sigma}{\tau\sqrt{\sigma^2+\tau^2}})^2 - \frac{z^2}{2(\sigma^2+\tau^2)})du = \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \exp(-\frac{1}{2(\sigma^2+\tau^2)}z^2)$, also folgt $Z \sim N(0,\sigma^2+\tau^2)$.

Im allgemeinen Fall, $X \sim N(\mu,\sigma^2)$, $Y \sim N(\lambda,\tau^2)$, liefert Anwendung auf $X-\mu$ und $Y-\lambda$:

X, Y unabhängig, $X \sim N(\mu,\sigma^2)$, $Y \sim N(\nu,\tau^2) \implies X+Y \sim N(\mu+\nu,\sigma^2+\tau^2)$.

Satz 11.16 Sind X, Y unabhängige ZV mit charakteristischen Funktionen φ_X und φ_Y, so gilt für die charakteristische Funktion φ_{X+Y} zu $X+Y$: $\varphi_{X+Y}(\theta) = \varphi_X(\theta) \cdot \varphi_Y(\theta) \quad \forall \theta \in \mathbb{R}$.

Beweis: Sei $\theta \in \mathbb{R}$. Mit X und Y sind auch $\exp(i\theta X)$ und $\exp(i\theta Y)$ unabhängig, also folgt:

$$\varphi_{X+Y}(\theta) = E\exp(i\theta(X+Y)) = E(\exp(i\theta X)\exp(i\theta Y)) \overset{\text{Multiplika-}}{\underset{\text{tionsregel}}{=}} (E\exp(i\theta X)) \cdot (E\exp(i\theta Y)) = \varphi_X(\theta) \cdot \varphi_Y(\theta). \quad \square$$

Beispiel 11.17 X, Y unabhängig, $X \sim N(\mu,\sigma^2)$, $Y \sim N(\nu,\tau^2)$.

$\varphi_X(\theta) = \exp(i\mu\theta - \frac{1}{2}\sigma^2\theta^2)$, $\varphi_Y(\theta) = \exp(i\nu\theta - \frac{1}{2}\tau^2\theta^2)$ (nach Beispiel 11.8)

Satz 11.16 liefert $\varphi_{X+Y}(\theta) = \exp(i(\mu+\nu)\theta - \frac{1}{2}(\sigma^2+\tau^2)\theta^2)$.

Dies ist die charakteristische Funktion (nach Beispiel 11.8) zu $N(\mu+\nu,\sigma^2+\tau^2)$, also mit Korollar 11.10

$X+Y \sim N(\mu+\nu,\sigma^2+\tau^2)$ (vergleiche mit Beispiel 11.15.)

Kapitel 12

Verteilungskonvergenz

12.1 Erinnerung, Zusammenhang zu anderen Konvergenzbegriffen

(Definition 6.2) Sind $P, P_1, P_2, \ldots$ WMaße auf $(\mathbb{R}, \mathfrak{B})$ mit Verteilungsfunktion $F, F_1, F_2, \ldots$, so konvergiert P_n schwach gegen P, $P_n \xrightarrow{W} P$, wenn gilt: $\lim\limits_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitspunkte x von F.

Sind $X, X_1, X_2, \ldots$ ZV auf (u.U. verschiedenen) WRäumen $(\Omega, \mathfrak{A}, P), (\Omega_1, \mathfrak{A}_1, P_1), \ldots$ so konvergiert X_n gegen X in Verteilung, $X_n \xrightarrow{D} X$, wenn ${P_n}^{X_n} \xrightarrow{W} P^X$, d.h. $\lim\limits_{n\to\infty} P_n(X_n \le x) = P(X \le x)$ für alle x mit $P(X = x) = 0$.

Satz 12.1 Sind $X, X_1, X_2, \ldots$ ZV auf demselben WRaum $(\Omega, \mathfrak{A}, P)$, so gilt:

$X_n \xrightarrow{\text{f.s.}} X \implies X_n \xrightarrow{D} X,\ X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X.$

Beweis: Da "$\xrightarrow{\text{f.s.}} \implies \xrightarrow{P}$" bekannt ist, reicht es, die zweite Aussage zu beweisen.

Es gelte also $X_n \xrightarrow{P} X, F, F_1, F_2, \ldots$ seien die zugehörigen Verteilungsfunktionen.

$P(X_n \le x) \le P(X \le x + \delta) + P(|X_n - X| > \delta)$ (denn $X_n \le x \implies X \le x + \delta$ oder $|X_n - X| > \delta$)

$P(X_n \le x) \ge P(X \le x - \delta) - P(|X_n - X| > \delta)$

Läßt man nun zunächst $n \to \infty$ gehen und dann $\delta \downarrow 0$, so folgt

$F(x-) \le \liminf\limits_{n\to\infty} F_n(x) \le \limsup\limits_{n\to\infty} F_n(x) \le F(x)$, also $F_n(x) \longrightarrow F(x)$ für alle Stetigkeitspunkte x von F

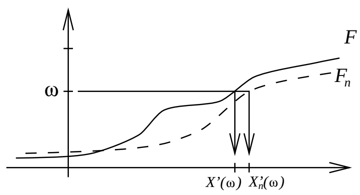
(in denen ja $F(x-) = F(x)$ gelten muß.) □

Schwache Konvergenz ist wirklich schwächer als $\xrightarrow{P}$:

$$X \sim N(0,1)\ X_n = \begin{cases} X & , \ n \text{ gerade} \\ -X & , \ n \text{ ungerade} \end{cases}$$

Satz 12.2 (Shorohod)

Es seien $X, X_1, X_2, \ldots$ ZV mit $X_n \xrightarrow{D} X$. Dann existieren ein WRaum $(\Omega, \mathfrak{A}, P)$ und hierauf ZV $X', X_1', X_2', \ldots$ mit $X' \stackrel{D}{=} X$,[1] $X_n' \stackrel{D}{=} X_n \quad \forall n \in \mathbb{N}$, $X_n' \xrightarrow{\text{f.s.}} X'$.



Beweis: Es seien wieder $F, F_1, F_2, \ldots$ die Verteilungsfunktionen zu $X, X_1, X_2, \ldots$.

Sei $(\Omega, \mathfrak{A}, P) := ([0,1), \mathfrak{B}|_{[0,1)}, l|_{[0,1)})$. $X' := F^{-1}$, $X_n' := {F_n}^{-1}$

(Definition 5.19) $F^{-1}(x) = \inf\{y : F(y) \ge x\}$ "Quantiltransformation"

Dann gilt: $X' \stackrel{D}{=} X$, $X_n' \stackrel{D}{=} X_n$ ($\rightsquigarrow$ Beweis zu Satz 5.21)

Es bleibt zu zeigen:

(1) $\lim X_n'(\omega) = X'(\omega)$ für P-fast alle $\omega \in \Omega$.

Sei $\omega \in (0,1)$. Zu gegebenem $\epsilon > 0$ wähle ein $x \in \mathbb{R}$ mit $X'(\omega) - \epsilon < x < X'(\omega)$ und $P(X = x) = 0$.

(X hat nur höchstens abzählbar viele Atome, also läßt sich in $(X'(\omega) - \epsilon, X'(\omega))$ ein solches x finden.)

Man hat $y \le F(x) \iff F^{-1}(y) \le x$ (Lemma 5.20), also $\omega \le F(x) \iff X'(\omega) \le x$.

Da F in x stetig ist, gilt $F_n(x) \longrightarrow F(x)$ mit $n \to \infty$, also existiert ein n_0 derart, daß $F_n(x) \le \omega$ für alle $n \ge n_0$ gilt, denn $F(x) < \omega$. Damit $X_n'(\omega) > x$ für alle $n \ge n_0$, denn: $\omega \le F_n(x) \iff X_n'(\omega) \le x$, also $\liminf\limits_{n\to\infty} X_n'(\omega) \ge x \ge X'(\omega) - \epsilon$.

Hieraus folgt mit $\epsilon \downarrow 0$ $\liminf\limits_{n\to\infty} X_n'(\omega) \ge X'(\omega)$.

Eine ähnliche Argumentation zeigt $\limsup\limits_{n\to\infty} X_n'(\omega) \le X'(\omega') \quad \forall \omega' > \omega$.

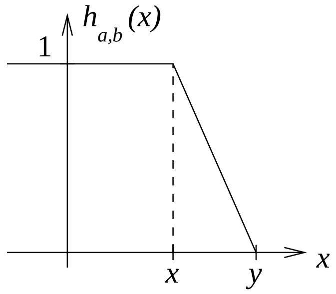
In den Punkten ω, in denen X' von rechts stetig ist, und das sind fast alle, kann man nun $\omega' \downarrow \omega$ ausführen und erhält insgesamt $X'(\omega) \le \liminf X_n'(\omega)$, $\limsup X_n'(\omega) \le X'(\omega)$. □

[1] $\mathfrak{L}(X') = \mathfrak{L}(X)$, X' hat dieselbe Verteilung wie X

Satz 12.3 Es sei $C_b(\mathbb{R})$ die Menge der stetigen und beschränkten Funktionen $h : \mathbb{R} \to \mathbb{R}$.
Dann gilt: $X_n \xrightarrow{D} X \iff Eh(X_n) \to Eh(X) \quad \forall h \in C_b(\mathbb{R})$.

Beweis: "$\Longrightarrow$": Nach Satz 12.2 existieren auf einem geeigneten WRaum $(\Omega, \mathfrak{A}, P)$ ZV $X', X'_1, X'_2, \ldots$ mit $X' \overset{D}{=} X$, $X'_n \overset{D}{=} X_n \quad \forall n$ und $X'_n \to X'$ fast sicher.
Da h stetig ist, gilt dann auch $h(X'_n) \to h(X')$ f.s..
Da h beschränkt ist, kann der Satz von der majorisierten Konvergenz angewendet werden und liefert
$Eh(X_n) = Eh(X'_n) \to Eh(X') = Eh(X)$

"$\Longleftarrow$": Für alle $a, b \in \mathbb{R}$ mit $a < b$ sei $h_{a,b}(x) = \begin{cases} 1 & , \quad x \le a \\ \frac{b-x}{b-a} & , \quad a < x < b \\ 0 & , \quad x \ge b \end{cases}$



Es seien F, F_k $(k \in \mathbb{N})$ die Verteilungsfunktionen zu X, X_k.
Dann gilt für alle $y > x$
$F_n(x) = E1_{(-\infty,x]}(X_n) \le Eh_{x,y}(X_n) \to Eh_{x,y}(X) \le E1_{(-\infty,y]}(X) = F(y)$.
Also: $\limsup\limits_{n\to\infty} F_n(x) \le F(y) \quad \forall y > x$, mit $y \downarrow x$: $\limsup\limits_{n\to\infty} F_n(x) \le F(x+) = F(x)$.
Analog erhält man für $y < x$: $F_n(x) \ge Eh_{y,x}(X_n) \to Eh_{y,x}(X) \ge F(y)$,
also mit $y \uparrow x$: $\liminf\limits_{n\to\infty} F_n(x) \ge F(x-)$.
In den Stetigkeitspunkten gilt: $F(x-) = F(x)$, also $F_n(x) \to F(x)$. □

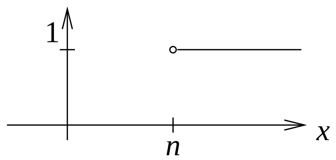
Satz 12.4 ("Continuous Mapping Theorem")
Es seien $X, X_1, X_2, \ldots$ ZV mit $X_n \xrightarrow{D} X$. Weiter sei $f : \mathbb{R} \to \mathbb{R}$ eine Borel-meßbare Abbildung mit der Eigenschaft $P(X \in \{x \in \mathbb{R} : f \text{ nicht stetig in } x\}) = 0$. Dann gilt $f(X_n) \xrightarrow{D} f(X)$.
Beweis: Übungsaufgabe. □

12.2 Straffheit und charakteristische Funktionen

Aus der Analysis:

Satz 12.5 (Helly)
Zu jeder Folge $(F_n)_{n\in\mathbb{N}}$ von Verteilungsfunktionen existiert eine Teilfolge $(F_{n_k})_{k\in\mathbb{N}}$ sowie eine schwach monoton steigende, rechtsstetige Funktion $G : \mathbb{R} \to [0,1]$ mit $\lim\limits_{k\to\infty} F_{n_k}(x) = G(x)$ in allen Stetigkeitspunkten x von G.
Grobe Beweisskizze: Für jedes $x \in \mathbb{R}$ hat $(F_n(x))_{n\in\mathbb{N}}$ als beschränkte Folge reeller Zahlen einen Häufungspunkt. Ist $(r_k)_{k\in\mathbb{N}}$ eine Abzählung von $\mathbb{Q}$, so wähle nacheinander Teil-Teil-Folgen $(F_{n_{k,j}})_{j\in\mathbb{N}}$ mit $F_{n_{k,j}}(r_k) \xrightarrow{j\to\infty} G_0(r_k)$ (Def. von G_0). Betrachte dann die Diagonalfolge $F_{n_{j,j}}(x) \to G_0(x) \quad \forall x \in \mathbb{Q}$. Setze $G(x) := \inf\{G_0(q) : q \in \mathbb{Q},\ q \ge x\}$.
Rest $\epsilon - \delta$.

Beachte: Das G aus Satz 12.5 ist nicht unbedingt eine Verteilungsfunktion.
Im Falle $F_n = 1_{(n,\infty)}$ hat man $G \equiv 0$ (Masse verschwindet im unendlichen.)



Definition 12.6 Eine Familie $\mathfrak{P}$ von WMaßen heißt <u>straff</u> (englisch: tight), wenn zu jedem $\epsilon > 0$ ein kompaktes Intervall $[a,b]$ existiert mit $P([a,b]) \ge 1 - \epsilon \quad \forall P \in \mathfrak{P}$.

Man sieht leicht: $\mathfrak{P}_1$ und $\mathfrak{P}_2$ straff $\Longrightarrow$ $\mathfrak{P}_1 \cup \mathfrak{P}_2$ straff.
$\mathfrak{P}$ straff, $\mathfrak{P}_0 \subset \mathfrak{P} \Longrightarrow \mathfrak{P}_0$ straff.
$\#\mathfrak{P} = 1 \Longrightarrow \mathfrak{P}$ straff.

Satz 12.7 Ist $\{P_n : n \in \mathbb{N}\}$ eine straffe Familie von WMaßen auf $(\mathbb{R}, \mathfrak{B})$, so existiert eine Teilfolge $(P_{n_k})_{k\in\mathbb{N}}$ und ein WMaß P mit $P_{n_k} \xrightarrow{W} P,\ k \to \infty$.

Beweis: Sei F_n die Verteilungsfunktion von P_n. Der Satz von Helly liefert $(n_k)_{k\in\mathbb{N}}$ und eine schwach monoton steigende, rechtsstetige Funktion G mit Werten in $[0,1]$ derart, daß $F_{n_k}(x) \to G(x) \quad \forall x$, G stetig in x.
Noch zu zeigen: $\lim\limits_{x\to-\infty} G(x) = 0$, $\lim\limits_{x\to\infty} G(x) = 1$ (Dann ist G Verteilunsgfunktion, und das zugehörige P tut's.)
Es sei $\epsilon > 0$. Da $\{P_n : n \in \mathbb{N}\}$ straff ist, existieren a, b mit $P_n([a+1, b]) \ge 1 - \epsilon \quad \forall n \in \mathbb{N}$.
Insbesondere gilt: $F_n(a) \le \epsilon \quad \forall n \in \mathbb{N}$. Da G höchstens abzählbar viele Unstetigkeitsstellen hat
($\rightsquigarrow$ Beweis zu Korollar 11.10), existiert $c < a$ mit G stetig in c.
Es folgt: $G(c) = \lim\limits_{k\to\infty} F_{n_k}(c) \le \epsilon$, also $G(x) \le \epsilon \quad \forall x \le c$.
Damit ist gezeigt: $\forall \epsilon > 0\ \exists c \in \mathbb{R}\ \forall x \le c : 0 \le G(x) \le \epsilon$.
Das ist die Aussage $\lim\limits_{x\to-\infty} G(x) = 0$ in $\epsilon - \delta$-Schreibweise.
Bew. zu $\lim\limits_{x\to\infty} G(x) = 1$ ganz analog. □

Satz 12.8 (Stetigkeitssatz für charakteristische Funktionen)
Es seien $X, X_1, \ldots$ ZV mit charakteristischen Funktionen $\varphi, \varphi_1, \ldots$. Dann gilt $X_n \xrightarrow{D} X \iff \varphi_n(\theta) \to \varphi(\theta) \quad \forall \theta \in \mathbb{R}$.
Beweis:
"$\Longrightarrow$": Sei $\theta \in \mathbb{R}$. Die Funktionen $x \to \cos\theta x,\ x \to \sin\theta x$ sind stetig und beschränkt, also folgt mit Satz 12.3:

$$\varphi_n(\theta) = E\exp(i\theta X_n) = E\cos(\theta X_n) + iE\sin(\theta X_n) \overset{n\to\infty}{\longrightarrow} E\cos(\theta X) + iE\sin(\theta X) = E\exp(i\theta X) = \varphi(\theta).$$

"$\Longleftarrow$": Wir zeigen zunächst: $\{P^{X_n} : n \in \mathbb{N}\}$ ist straff.

Fubini II (für $\mathbb{C}$-wertig Funktionen) liefert für alle $\delta > 0$: $\frac{1}{\delta}\int\limits_{-\delta}^{\delta} \underbrace{1}_{\varphi_n(0)} - \varphi_n(\theta)d\theta = \int(\frac{1}{\delta}\int\limits_{-\delta}^{\delta}(1-\exp(i\theta x))d\theta)P^{X_n}(dx) =$

$$2\int \underbrace{(1 - \frac{\sin(\delta x)}{\delta x})}_{\geq 0} P^{X_n}(dx) \geq 2 \int\limits_{|x|\geq\frac{2}{\delta}} \underbrace{(1 - \frac{1}{|\delta x|})}_{\geq\frac{1}{2} \text{ auf } |x|\geq\frac{2}{\delta}} P^{X_n}(dx) \geq P^{X_n}([-\tfrac{2}{\delta}, \tfrac{2}{\delta}]^c)$$

Sei nun $\epsilon > 0$. Da $\varphi(0) = 1$ und φ stetig in 0 (Satz 11.4), existiert ein $\delta > 0$ mit $|1 - \varphi(\theta)| < \frac{\epsilon}{4}$ für $-\delta \leq \theta \leq \delta$, d.h. $|\frac{1}{\delta}\int\limits_{-\delta}^{\delta}(1 - \varphi(\theta))d\theta| \leq \frac{1}{\delta}2\delta\frac{\epsilon}{4} = \frac{\epsilon}{2}$.

Da alle φ_n's betragsmäßig durch 1 beschränkt sind und über ein endliches Intervall integriert wird, kann majorisierte Konvergenz angewendet werden: $\int\limits_{-\delta}^{\delta}(1 - \varphi_n(\theta))d\theta \to \int\limits_{-\delta}^{\delta}(1 - \varphi(\theta))d\theta$, also existiert ein $n_0 \in \mathbb{N}$ mit $\frac{1}{\delta}\int\limits_{-\delta}^{\delta}(1 - \varphi_n(\theta))d\theta \leq \epsilon \quad \forall n \geq n_0$.

Die Abschätzung liefert also: $P(X_n \in [-\frac{2}{\delta}, \frac{2}{\delta}]) \geq 1 - \epsilon \quad \forall n \geq n_0$.
Zu zeigen: $\forall \epsilon > 0\ \exists a \in \mathbb{R}\ \forall n \in \mathbb{N} : P^{X_n}([-a, a]) \geq 1 - \epsilon$.
Zu jedem der (endlich vielen) $n = 1, \ldots, n_0 - 1$ existiert ein $a_n \in \mathbb{R}$ mit $P^{X_n}([-a_n, a_n]) \geq 1 - \epsilon$, denn $P^{X_n}([-m, m]) \overset{m\to\infty}{\longrightarrow} 1$.
Also folgt mit $a := \max\{a_1, \ldots, a_{n_0-1}, \frac{2}{\delta}\}$: $P^{X_n}([-a, a]) \geq 1 - \epsilon$ für alle $n \in \mathbb{N}$.
Angenommen, $X_n \xrightarrow{D} X$ gilt nicht. Dann existiert ein $x \in \mathbb{R}$ mit $P(X = x) = 0$ und $P(X_n \leq x) \not\to P(X \leq x)$.
Es gibt dann ein $\epsilon > 0$ und eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $(*)$: $|P(X_{n_k} \leq x) - P(X \leq x)| \geq \epsilon$ für alle $k \in \mathbb{N}$.
Da $\{P^{X_{n_k}} : k \in \mathbb{N}\}$ wieder straff ist, existiert nach Satz 12.7 eine Teil-Teilfolge $(X_{n_{k_j}})_{j\in\mathbb{N}}$ und ein WMaß P_0 mit $P^{X_{n_{k_j}}} \xrightarrow{W} P_0$. Aus der Voraussetzung $\varphi_n \to \varphi$ und dem bereits bewiesenen "$\Longrightarrow$"-Teil folgt $\varphi_0 = \varphi$ (wobei φ_0 die charakteristische Funktion zu P_0 bezeichnet.)
Der Eindeutigkeitssatz liefert $P_0 = P^X$, also $X_{n_{k_j}} \xrightarrow{D} X$, und damit $P(X_{n_{k_j}} \leq x) \to P(X \leq x)$.
Dies ist ein Widerspruch zu $(*)$. □

Nocheinmal: Verteilungskonvergenz bezieht sich auf Verteilungen. $Y_n \xrightarrow{D} Y,\ Y \sim N(0,1)$ steht für $P^{Y_n} \xrightarrow{W} N(0,1)$, nach Satz 12.8 und Beispiel 11.8 reicht hierfür der Nachweis von $E\exp(i\theta Y_n) \to \exp(-\frac{1}{2}\theta^2) \quad \forall \theta \in \mathbb{R}$.

Kapitel 13

Der Zentrale Grenzwertsatz (ZGWS)

13.1 Identisch verteilte Summanden

Es wird das aus §6 (Stochastik 1) bekannte Material (kumulative Form der Normalapproximation) verallgemeinert.

Lemma 13.1 Es seien $z_1, \ldots, z_n, w_1, \ldots, w_n \in \{z \in \mathbb{C} : |z| \leq 1\}$. Dann gilt: $|\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} w_k| \leq \sum_{k=1}^{n} |z_k - w_k|$.
Beweis: Ersetzt man in z-Produkt nacheinander die z-Werte durch w-Werte, so folgt mit der Dreiecksungleichung:
$|z_1 \cdot \ldots \cdot z_n - w_1 \cdot \ldots \cdot w_n| \leq |z_1 \cdot \ldots \cdot z_n - w_1 \cdot z_2 \cdot \ldots \cdot z_n| + |w_1 \cdot z_2 \cdot \ldots \cdot z_n - w_1 \cdot w_2 \cdot z_3 \cdot \ldots \cdot z_n| + \ldots +$
$|w_1 \cdot \ldots \cdot w_{n-1} \cdot z_n - w_1 \cdot \ldots \cdot w_n| \leq |z_1 - w_1| + |z_2 - w_2| + \ldots + |z_n - w_n|$. □

Satz 13.2
Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten ZV mit endlicher, positiver Varianz σ^2 und Erwartungswert μ. Dann gilt für $\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i \quad \sqrt{n}(\overline{X_n} - \mu) \xrightarrow{D} Z,\ Z \sim N(0, \sigma^2)$ (alternativ: $\mathfrak{L}(\sqrt{n}(\overline{X_n} - \mu)) \xrightarrow{W} N(0, \sigma^2)$.)
Beweis: O.B.d.A. $\mu = 0$ (sonst ersetze X_i durch $X_i - \mu$).
Aus $\mu = 0$ und $EX_n^2 = var(X_n) = \sigma^2$ folgt mit Satz 11.6 $\varphi(\theta) = 1 - \frac{1}{2}\theta^2\sigma^2 + \theta^2 r(\theta)$, wobei $\lim_{\theta \to 0} r(\theta) = 0$.
Sei φ_n die charakteristische Funktion zu $\sqrt{n}\overline{X_n} \quad (= \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}})$. Satz 11.4(i), Satz 11.16 $\Longrightarrow \varphi_n(\theta) = (\varphi(\frac{\theta}{\sqrt{n}}))^n$
Lemma 13.1 liefert: für jedes $\theta \in \mathbb{R}$ existiert ein $n_0 \in \mathbb{N}$, so daß $\forall n \geq n_0 : \ |\varphi_n(\theta) - (1 - \frac{1}{2n}\sigma^2\theta^2)^n| \leq \theta^2 r(\frac{\theta}{\sqrt{n}})$,
also $\lim_{n \to \infty} \varphi_n(\theta) = \lim_{n \to \infty} (1 - \frac{1}{2n}\sigma^2\theta^2)^n = \exp(-\frac{1}{2}\sigma^2\theta^2) \quad \forall \theta \in \mathbb{R}$.
Die rechte Seite ist die charakteristische Funktion zu $N(0, \sigma^2)$, also folgt die Behauptung mit dem Stetigkeitssatz.□

Bemerkung 13.3 In der Literatur wird der ZGWS häufig auch in der Form (vgl. Bemerkung 6.8)
$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} Z,\ Z \sim N(0,1)$ angegeben; es werden also die Partialsummen $S_n = \sum_{i=1}^{n} X_n$ anstelle der Mittelwerte betrachtet.
Wegen $\frac{1}{\sqrt{n}\sigma}(S_n - n\mu) = \frac{1}{\sigma}\sqrt{n}(\overline{X_n} - \mu)$ und der Eigenschaft "$Z_n \xrightarrow{D} Z,\ c \in \mathbb{R} \Longrightarrow cZ_n \xrightarrow{D} cZ$"[1] sind beide Formen äquivalent. Diese Grenzwertaussagen führen auf die folgenden, für die Praxis sehr wichtigen Approximationen:
$N(\mu, \frac{\sigma^2}{n})$ für $\mathfrak{L}(\overline{X_n})$ und $N(n\mu, n\sigma^2)$ für $\mathfrak{L}(S_n)$.

Zusammenfassung: Ist eine ZV S die Summe einer großen Anzahl unabhängiger, identisch verteilter ZV, so ist S in etwa normalverteilt, oder: Die Verteilung von S kann durch die Normalverteilung mit dem gleichen Erwartungswert und der gleicher Varianz wie S approximiert werden.
In Absatz 13.2 werden wir sehen, daß "identisch verteilt" abgeschwächt werden kann.

Beispiel 13.4 Es seien $X_1, X_2, \ldots$ die zufälligen Lebensdauern (in Stunden) eines bestimmten Typs Glühbirnen. Wir nehmen an, daß diese unabhängig und exponentialverteilt mit Parameter $\frac{1}{100}$ sind (d.h. die mittlere Lebensdauer einer solchen Glühbirne beträgt 100 Stunden.)
Wieviele Glühbirnen sollte man kaufen, wenn man eine Gesamtbrenndauer von 1000 Stunden benötigt? Antwort: unendlich viele.
Modifikation: Wieviele sollte man kaufen, wenn man mit W. 0.99 sicher sein will, daß sie insgesamt mindestens 1000 Stunden brennen?

Es gilt; $EX_i = 100,\ var(X_i) = (EX_i)^2 = 10000$ (dies erhält man mit einer ähnlichen Rechnung wie in Beispiel 8.25)
Gesucht ist ein (möglichst kleines) n mit $P(S_n \geq 1000) \geq 0.99$
Man hat $P(S_n \geq 1000) \geq 0.99 \iff P(\frac{S_n - ES_n}{\sqrt{var(S_n)}} \geq \frac{1000 - n \cdot 100}{\sqrt{n}\sqrt{10000}}) \geq 0.99$.

[1](Aufgabe 34)

Für eine $N(0,1)$-verteilte ZV Z gilt $P(Z \geq -2.326) \approx 0.99$, der ZGWS führt also auf die Näherung $\frac{10-n}{\sqrt{n}} \leq -2.326$, d.h. $n \geq 20.54$. Man kauft also 21 Glühbirnen.
In diesem Beispiel kann man den exakten Wert ausrechnen und erhält, daß 19 Glühbirnen gereicht hätten.

In Satz 13.2: $X_1, X_2, \dots$ unabhängig und identisch verteilt, $\mu = EX_i$, $\sigma := var(X_i) < \infty$, $\mathfrak{L}(\frac{\sqrt{n}}{\sigma}(\overline{X_n} - \mu)) \xrightarrow{W} N(0,\sigma^2)$.
Aus der Stochastik I bekannt: Satz von Moivre-Laplace (Ist Spezialfall hiervon mit $P(X_i = 1) = 1 - P(X_i = 0) = p$).

13.2 Der Satz von Lindeberg

Satz 13.5 (Lindeberg)

Für jedes $n \in \mathbb{N}$ seien $X_{n1}, \dots, X_{nr_n}$ unabhängige ZV auf einem WRaum $(\Omega_n, \mathfrak{A}_n, P_n)$ mit endlicher Varianz σ_{nk}^2 und Erwartungswert μ_{nk}, $1 \leq k \leq r_n$, $s_n^2 := \sum_{k=1}^{r_n} \sigma_{nk}^2 > 0$. Ist dann die Lindeberg-Bedingung

$$(*) \qquad \boxed{\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{nk}-\mu_{nk}|>\epsilon \cdot s_n} (X_{nk} - \mu_{nk})^2 dP_n = 0 \quad \forall \epsilon > 0}$$

erfüllt, so gilt $\frac{1}{s_n} \sum_{k=1}^{r_n} (X_{nk} - \mu_{nk}) \xrightarrow{D} Z$, $Z \sim N(0,1)$.

Bemerkung 13.6

(i) Ist $Y_1, Y_2, \dots$ eine iid Folge auf $(\Omega, \mathfrak{A}, P)$ mit endlicher Varianz $\sigma^2 > 0$, so genügt die Familie $\{X_{nk} : n \in \mathbb{N}\}$ (mit $r_n = n$), $X_{nk} := Y_k$, $(\Omega_n, \mathfrak{A}_n, P_n) = (\Omega, \mathfrak{A}, P)$ den Bedingungen von Satz 13.5:
$s_n^2 = n\sigma^2$
$\frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{nk}-\mu_{nk}|>\epsilon s_n} (X_{nk} - \mu_{nk})^2 dP_n = \frac{1}{n\sigma^2} n \int_{|Y_1-\mu_1|>\epsilon\sqrt{n}\sigma} (Y_1 - \mu_1)^2 dP \longrightarrow 0$ wegen majorisierter Konvergenz, denn $(Y_1 - \mu_1)^2$ ist wegen $var(Y_1) < \infty$ eine integrierbare Majorante.
(ii) Beim Beweis wird deutlich werden, daß durch die Lindeberg-Bedingung ein dominanter Einfluß einzelner X_{nk}'s auf die Summe $X_{n1} + \dots + X_{nr_n}$ ausgeschlossen wird, eine Bedingung dieser Art ist auch notwendig.

Beweis: Wir können $\mu_{nk} = 0$, $s_n = 1$ annehmen – ersetze sonst X_{nk} durch $X'_{nk} := \frac{1}{s_n}(X_{nk} - \mu_{nk})$.
Mit $n = 2$ liefert Lemma 11.5 $|\exp(i\theta x) - (1 + i\theta x - \frac{1}{2}\theta^2 x^2)| \leq \min\{|\theta x|^2, |\theta x|^3\}$.
Für die charakteristische Funktion φ_{nk} zu X_{nk} bedeutet dies $|\varphi_{nk}(\theta) - (1 - \frac{1}{2}\theta^2\sigma_{nk}^2)| \leq E\min\{|\theta X_{nk}|^2, |\theta X_{nk}|^3\}$.
Sei $\epsilon > 0$. Dann gilt:
$E\min\{\cdot\} \leq \int_{|X_{nk}|<\epsilon} |\theta X_{nk}|^3 dP_n + \int_{|X_{nk}|\geq\epsilon} |\theta X_{nk}|^2 dP_n \leq \epsilon\theta^3\sigma_{nk}^2 + \theta^2 \int_{|X_{nk}|\geq\epsilon} X_{nk}^2 dP_n$.
Summiere über k: $\sum \sigma_{nk}^2 = 1$, also

$$(1) \quad \lim_{n\to\infty} \sum_{k=1}^{r_n} |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\theta^2\sigma_{nk}^2)| \leq \epsilon\theta^3 + \theta^2 \underbrace{\overbrace{\sum_{k=1}^{r_n} \dots}^{\text{wie in } (*)}}_{\to 0 \text{ mit } n\to\infty} \xrightarrow{\epsilon\to 0} 0.$$

Wegen $s_n = 1$ ist $\prod_{k=1}^{r_n} \varphi_{nk}$ die charakteristische Funktion zu $\frac{1}{s_n}S_n$, $S_n := X_{n1} + \dots + X_{nr_n}$.
Wir behaupten nun:

$$(2) \quad \lim_{n\to\infty} |\prod_{k=1}^{r_n} \varphi_{nk}(\theta) - \prod_{k=1}^{r_n} (1 - \tfrac{1}{2}\theta^2\sigma_{nk}^2)| = 0 \qquad \forall \theta \in \mathbb{R}.$$

Für alle $\epsilon > 0$ gilt $\sigma_{nk}^2 \leq \epsilon^2 + \int_{|X_{nk}|>\epsilon} X_{nk}^2 dP_n$, also folgt mit der Lindeberg-Bedingung:

$$(0 \leq) \max\{\sigma_{nk}^2 : 1 \leq k \leq r_n\} \leq \epsilon^2 + \underbrace{\sum_{k=1}^{r_n} \int_{|X_{nk}|>\epsilon} X_{nk}^2 dP_n}_{\to 0 \text{ mit } n\to\infty}, \text{ also}$$

$$(3) \quad \lim_{n\to\infty} \max\{\sigma_{nk}^2 : 1 \leq k \leq r_n\} = 0$$

Für alle $\theta \in \mathbb{R}$ existiert also ein n_0 derart, daß für alle $n \geq n_0$
(4) $|1 - \frac{1}{2}\theta^2\sigma_{nk}^2| \leq 1$ für $k = 1, \dots, r_n$.
Also kann Lemma 13.1 verwendet werden

$$|\prod_{k=1}^{r_n} \varphi_{nk}(\theta) - \prod_{k=1}^{r_n} (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| \leq \sum_{k=1}^{r_n} |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| \quad \text{für alle } n \geq n_0,\ k = 1, \dots, r_n.$$

Also folgt (2) mit (1).
Es bleibt zu zeigen:

$$(5) \quad \lim_{n\to\infty} |\underbrace{\prod_{k=1}^{r_k} \exp(-\frac{1}{2}\theta^2\sigma_{nk}^2)}_{\exp(-\frac{1}{2}\theta^2) \text{ char. Fkt. zu } N(0,1)} - \prod_{k=1}^{r_n} (1 - \tfrac{1}{2}\theta^2\sigma_{nk}^2)| \overset{!}{=} 0 \qquad \forall \theta \in \mathbb{R}$$

Wegen Lemma 13.1 reicht es,

$$(6) \quad \lim_{n\to\infty} \sum_{k=1}^{r_k} |\exp(-\tfrac{1}{2}\theta^2\sigma_{nk}^2) - 1 + \tfrac{1}{2}\theta^2\sigma_{nk}^2| = 0 \quad \text{zu zeigen.}$$

Für reelle x mit $|x| \le \frac{1}{2}$ gilt: $|\exp(x) - 1 - x| \le \frac{1}{2} \sum_{j=2}^{\infty} |x|^j$ (geometrische Reihe)

Dies führt auf $\sum_{k=1}^{r_n} |\exp(-\frac{1}{2}\theta^2\sigma_{nk}^2) - 1 + \frac{1}{2}\theta^2\sigma_{nk}^2| \le \frac{1}{2}\theta^4 \sum_{k=1}^{r_k} \sigma_{nk}^4$

Wegen $\sum_{k=1}^{r_k} \sigma_{nk}^4 \le \max\{\sigma_{nk}^2 : 1 \le k \le r_n\} \sum_{k=1}^{r_n} \sigma_{nk}^2$ folgt hieraus (6) mit Hilfe von (3). □

Bemerkung 13.7 Der ZGWS hat eine lange "Verbesserungsgeschichte" hinter sich, die Lindeberg-Bedingung ist in gewisser Weise das Endresultat. Gelegentlich ist die Lyapunov-Bedingung

$$\boxed{\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{r_n} E|X_{nk} - \mu_{nk}|^{2+\delta} = 0}$$

für ein $\delta > 0$ einfacher zu verwenden. Daß diese die Lindeberg-Bedingung impliziert folgt aus
$\int_{|X_{nk}-\mu_{nk}|>\epsilon s_n} (X_{nk} - \mu_{nk})^2 dP_n \le \int_{|X_{nk}-\mu_{nk}|>\epsilon s_n} \frac{|X_{nk}-\mu_{nk}|^{2+\delta}}{\epsilon^\delta s_n^\delta} dP_n \le \frac{1}{\epsilon^\delta} \frac{1}{s_n^\delta} E|X_{nk} - \mu_{nk}|^{2+\delta}$.

Zusammenfassung: Die Summe vieler unabhängiger, individuell vernachlässigbarer ZV ist näherungsweise normalverteilt.

13.3 Anwendungen

13.3.1 Ein Sammelproblem

Einer Urne mit n Kugeln werden solange Kugeln zufällig und mit Zurücklegen entnommen, bis insgesamt r_n verschiedene Kugeln gezogen worden sind. S_n sei die Anzahl der hierfür notwendigen Züge.
Dies kann als Modell für die folgende Situation dienen: Jedem Paket Haferflocken liegt eines von n verschiedenen Bildchen bei. Wieviele Pakete muß man kaufen, bis man r_n verschiedene Bilder hat?
($\rightsquigarrow$ G. Polya: Eine Wahrscheinlichkeitsaufgabe zur Kundenwerbung, ZAMM 1930)
Angenommen $r_n = \lceil \rho n \rceil$ mit $0 < \rho < 1$ (d.h. mindestens $100 \cdot \rho\%$ der Bilder sollen gesammelt werden).
Bei n Kugeln werden die wiederholten Entnahmen von Kugeln beschrieben mit Hilfe des WRaums $(\Omega_n, \mathfrak{A}_n, P_n) = (\{1, \dots, n\}, \mathbb{P}(\{1, \dots, n\}), unif(\{1, \dots, n\}))^{\mathbb{N}}$ (siehe §9).
Sei X_{nk} die Anzahl der Züge, die nötig sind, um eine neue Kugel zu erhalten, wenn man bereits $k-1$ verschiedene Kugeln entnommen (und wieder zurückgelegt) hat. Dann ist X_{nk} geometrisch verteilt ($\rightsquigarrow$ Absatz 4.2.3) mit Parameter $\frac{n-(k-1)}{n}$ (klar: $X_1 = 1$). Außerdem ist X_{nk} unabhängig von $X_{n1}, \dots, X_{n,k-1}$, hieraus folgt, daß $X_{n1}, \dots, X_{nr_n}$ ($r_n := \lceil \rho n \rceil$) unabhängig sind.
Nach Konstruktion: $S_n = X_{n1} + X_{n2} + \dots + X_{nr_n}$.
Bekannt: Y geometrisch verteilt mit Parameter $p \implies EY = \frac{1}{p}$, $var(Y) = \frac{q}{p^2}$, $q := 1 - p$.
Auch: $EY^4 = \sum_{k=1}^{\infty} k^4 P(Y = k) = p \sum_{k=1}^{\infty} k^4 q^{k-1} \le p \sum_{k=1}^{\infty} (k+3)(k+2)(k+1)kq^{k-1} = p\left(\frac{d^4}{(dx)^4} \underbrace{\sum_{k=0}^{\infty} x^k}_{\frac{1}{1-x}}\right)\Bigg|_{x=q} = \frac{24}{p^4}$.
Für den Erwartungswert und die Varianz von S_n ergibt sich damit
$\mu_n = ES_n = \sum_{k=1}^{r_n} \frac{n}{n-k+1}$, $s_n^2 = var(S_n) = \sum_{k=1}^{r_n} \frac{1 - \frac{n-k+1}{n}}{(\frac{n-k+1}{n})^2} = n \sum_{k=1}^{r_n} \frac{k-1}{(n-k+1)^2}$.
Als nächstes: Lyapunov-Bedingung mit $\delta = 2$ erfüllt: Mit Y wie oben hat man
$E(Y - \frac{1}{p})^4 \le E(2\max\{Y, \frac{1}{p}\})^4 = 16E\max\{Y^4, (\frac{1}{p})^4\} \le 16E(Y^4 + \frac{1}{p^4}) \le \frac{400}{p^4}$.
Damit $\sum_{k=1}^{r_n} E(X_{nk} - EX_{nk})^4 \le 400 \sum_{k=1}^{r_n} \frac{n^4}{(n-k+1)^4} \le 400\lceil \rho n \rceil \frac{1}{(1-\rho)^4} = O(n)$.
Wegen $s_n^2 \ge n\frac{1}{n^2} \sum_{k=1}^{r_n} (k-1) \ge \frac{1}{2}\frac{1}{n}(\rho n - 1)(\rho n - 2) \sim cn$, $c > 0$, also $\lim_{n\to\infty} \frac{1}{s_n^{2+2}} \sum_{k=1}^{r_n} E(X_{nk} - EX_{nk})^{2+2} = 0$.
Also "gilt der ZGWS für S_n": $\frac{S_n - ES_n}{\sqrt{var(S_n)}} \xrightarrow{D} Z$, $Z \sim N(0,1)$.
Wie verhalten sich ES_n und $var(S_n)$ mit $n \to \infty$?
$\frac{1}{n}ES_n = \sum_{k=1}^{\lceil \rho n \rceil} \frac{1}{1-\frac{k-1}{n}} \frac{1}{n}$ Interpretation als Riemannsumme führt auf $\frac{1}{n}E(S_n) \to \int_0^\rho \frac{1}{1-x}dx + O(\frac{1}{n}) = -\log(1-\rho) + O(\frac{1}{n})$
und ähnlich erhält man $\lim_{n\to\infty} \frac{1}{n}var(S_n) = \int_o^\rho \frac{x}{(1-x)^2}dx = \frac{\rho}{1-\rho} + \log(1-\rho)$.
Mit $a(\rho) := -\log(1-\rho)$, $b(\rho) := \sqrt{\frac{\rho}{1-\rho} + \log(1-\rho)}$ folgt also $\frac{S_n - a(\rho)n}{b(\rho)\sqrt{n}} \xrightarrow{D} Z$, $Z \sim N(0,1)$.

Numerisches Beispiel: Wie groß muß Ihr Bekanntenkreis sein, damit mit W. ≥ 0.95 an 180 Tagen im Jahr Geburtstag gefeiert werden kann?
$n = 365$, $\rho = \frac{180}{365}$, bekannt: $P(Z \ge 1.645) \approx 0.05$, wenn $Z \sim N(0,1)$ und $S_n \ge k \iff \frac{S_n - a(\rho)n}{b(\rho)\sqrt{n}} \ge \underbrace{\frac{k - a(\rho)n}{b(\rho)\sqrt{n}}}_{\approx 1.645}$.
Wir erhalten den Mittelwert $k = 266$.

Das Verhalten von $a(\rho)$ und $b(\rho)$ mit $\rho \uparrow 1$ läßt vermuten, daß bei $\rho = 1$ (jede Kugel muß mindestens einmal erscheinen) ein anderes Verhalten vorliegt.

$var(S_n) = n\sum\limits_{k=1}^{n} \frac{k-1}{(n-k+1)^2} = n\sum\limits_{k=1}^{n} \frac{n-k}{k^2} = n^2\sum\limits_{k=1}^{n} \frac{1}{k^2} - n\sum\limits_{k=1}^{n} \frac{1}{k} = n^2\frac{\pi^2}{6} + o(n^2)$.

$var(X_{nn}) = \frac{1-\frac{1}{n}}{(n-k+1)^2} = n^2 + o(n^2)$, d.h. die einzelnen Summanden sind nicht mehr vernachlässigbar ($\max\limits_{1\le k\le r_n} \sigma_{nk}^2 \neq o(s_n^2)$), die Lindeberg-Bedingung ist nicht erfüllt.

(*) $\frac{S_n - n\log n}{n} \xrightarrow{D} Z,\ P(Z \le y) = \exp(-e^{-y})$ (Gumbel, Aufgabe 29)

Numerisches Beispiel: Soll der Bekanntenkreis so groß sein, daß mit W. 0.95 an jedem der 365 Tage im Jahr Geburtstag gefeiert werde kann, so erhält man die Bedingung: $k \ge 365\log 365 + 365\cdot 2.97 \approx 3237.51$

Heuristische Argumentation für die Gültigkeit von (*):
Es sei A_{ik} das Ereignis, das die i-te Kugel nach k Zügen noch nicht aufgetaucht ist. Dann gilt:
$P(A_{ik}) = (1-\frac{1}{n})^k$ für $i = 1,\dots,n$.
Wir unterstellen nun, daß die Ereignisse $A_{1k},\dots,A_{nk}$ zumindest näherungsweise unabhängig sind.
Dann $P(S_n \le k_n) = P(\bigcap\limits_{i=1}^{n} A_{ik_n}^c) \approx (1-(1-\frac{1}{n})^{k_n})^n$.
Für $k_n = n\log n + x\cdot n$ erhält man $P(\frac{S_n - n\log n}{n} \le x) = P(S_n \le k_n) \approx (1-\exp(\overbrace{k_n \underbrace{\log(1-\frac{1}{n})}_{\approx -\frac{1}{n}}}^{-\log n - x}))^n \approx (1-\frac{1}{n}e^{-x})^n$
$\to \exp(-e^{-x})$.

13.3.2 Rekorde

Es sei $X_1, X_2, \dots$ eine iid-Folge auf $(\Omega, \mathfrak{A}, P)$ mit stetiger Verteilungsfunktion F.
Wir setzen $R_n = \begin{cases} 1 & , \quad X_n > X_1, X_n > X_2, \dots, X_n > X_{n-1} \\ 0 & , \quad \text{sonst} \end{cases}$. R_n zeigt an, ob die n-te Versuchswiederholung einen Rekord liefert oder nicht. Mit Fubini folgt für $i \neq j$ $P(X_i = X_j) = P^{X_i} \otimes P^{X_j}(\{(x,x) : x \in \mathbb{R}\}) = \int \underbrace{\int 1_{\{x\}}(y) P^{X_i}(dy)}_{F(x)-F(x-)=0} P^{X_j} dx = 0$, also ist $A := \{\omega \in \Omega : \exists i,j,\ i \neq j,\ X_i(\omega) = X_j(\omega)\} = \bigcup\limits_{i=2}^{\infty}\bigcup\limits_{j=1}^{i-1}\{X_i = X_j\}$ eine P-Nullmenge. Die Wahrscheinlichkeit für das Auftreten "mehrfacher Werte" (engl.: ties) ist 0, kann also vernachlässigt werden.

Sei $\mathbb{P}_n$ die Menge der Permutationen von $\{1,\dots,n\}$. Sei $\Psi_n : \Omega \to \mathbb{P}_n$ die zufällige Permutation, die $X_1,\dots,X_n$ in aufsteigende Reihenfolge bringt:
$\Psi_n = \Pi \iff X_{\Pi(1)} < X_{\Pi(2)} < \dots < X_{\Pi(n)}$
$\Psi_n^{-1}(\{\Pi\})$ ist der Durchschnitt von Mengen $\{X_i < X_j\}$, also in $\mathfrak{A}$, d.h. Ψ_n ist eine Zufallsgröße (meßbare Abbildung), wenn man $\mathbb{P}_n$ mit der Potenzmenge als σ-Algebra ausstattet.
Für jedes feste $\Pi \in \mathbb{P}_n$ haben die Vektoren $(X_{\Pi(1)},\dots,X_{\Pi(n)})$ und $(X_1,\dots,X_n)$ dieselbe Verteilung, nämlich $(P^{X_1})^{\otimes n}$. Sei $B := \{(x_1,\dots,x_n) : x_1 < x_2 < \dots < x_n\}$. Dann gilt $P(\Psi_n = \Pi) = P((X_{\Pi(1)},\dots,X_{\Pi(n)}) \in B) = P((X_1,\dots,X_n) \in B) = P(\Psi_n = id)$, also:
(1) $P(\Psi_n = \Pi) = \frac{1}{n!}$ für alle $\Pi \in \mathbb{P}_n$.
Damit (2) $P(R_n = 1) = P(\Psi_n \in \{\Pi \in \mathbb{P}_n : \Pi(n) = n\}) = \frac{(n-1)!}{n!} = \frac{1}{n}$.

$\{R_{n+1} = 1\} \cap \{\Psi_n = \Pi\} = \{\Psi_{n+1} = \tilde{\Psi}\},\quad \tilde{\Psi}(i) = \begin{cases} \Pi(i) & , \quad 1 \le i \le n \\ n+1 & , \quad i = n+1 \end{cases}$

Es folgt mit (1) und (2) $P(\Psi_n = \Pi, R_{n+1} = 1) = P(\Psi_n = \Pi)P(R_{n+1} = 1)\quad \forall \Pi \in \mathbb{P}_n$, d.h. $\sigma(\Psi_n)$ und $\sigma(R_{n+1})$, also die erzeugten σ-Algebren, sind unabhängig. Wegen $\sigma(R_1,\dots,R_n) \subset \sigma(\Psi_n)$ sind dann auch $\sigma(R_1,\dots,R_n)$ und $\sigma(R_{n+1})$ unabhängig. Für beliebige $i_1 < i_2 < \dots < i_n$ bedeutet dies:
$P(R_{i_1} = j_1,\dots,R_{i_n} = j_n) = P(R_{i_1} = j_1,\dots,R_{i_{n-1}} = j_{n-1})P(R_{i_n} = j_n) = \dots = P(R_{i_1} = j_1)\cdot\ldots\cdot P(R_{i_n} = j_n)$, d.h. die ZV $R_1, R_2, \dots$ sind unabhängig.

Wieviele Rekorde gibt es unter den ersten n Versuchen? Sei $S_n := \sum\limits_{i=1}^{n} R_i$.

$ES_n = \sum\limits_{i=1}^{n} ER_i = \sum\limits_{i=1}^{n} P(R_i = 1) \overset{(2)}{=} \sum\limits_{i=1}^{n} \frac{1}{i} = \log n + \gamma + o(1)$ (γ ist Eulersche Konstante)

$var(S_n) \underset{\text{hängig}}{\overset{\text{unab-}}{=}} \sum\limits_{i=1}^{n} var(R_i) = \sum\limits_{i=1}^{n} \frac{1}{i}(1-\frac{1}{i}) = \sum\limits_{i=1}^{n} \frac{1}{i} - \sum\limits_{i=1}^{n} \frac{1}{i^2} = \log n + \gamma + o(1) - \frac{\pi^2}{6} + o(1)$

Chebychev liefert $P(|\frac{S_n}{ES_n} - 1| \ge \epsilon) \le \frac{1}{\epsilon^2}\frac{var(S_n)}{(ES_n)^2} \xrightarrow{n\to\infty} 0$, d.h. $\frac{S_n}{ES_n} \xrightarrow{P} 1$.
Der ZGWS erlaubt eine genauere Aussage:
Mit $s_n := (var(S_n))^{\frac{1}{2}} \sim (\log n)^{\frac{1}{2}}$ hat man $\frac{1}{s_n^3}\sum\limits_{i=1}^{n} E|R_i|^3 = \frac{1}{s_n^3}\sum\limits_{i=1}^{n} ER_i \xrightarrow{n\to\infty} 0$, d.h. Lyapunov mit $\delta = 1$ ist erfüllt,

also $\frac{S_n - ES_n}{\sqrt{var(S_n)}} \xrightarrow{D} Z,\ Z \sim N(0,1)$.

Verwendet man die Eigenschaften der Verteilungskonvergenz ($\rightsquigarrow$ Aufgabe 34), so folgt hieraus $\frac{S_n - \log n}{\sqrt{\log n}} \xrightarrow{D} Z$, $Z \sim N(0,1)$.

Für große n ist also beispielsweise die W. dafür, daß die Anzahl der in den ersten n Durchgängen erzielten Rekorde zwischen $\log n - 1.96\sqrt{\log n}$ und $\log n + 1.96\sqrt{\log n}$ liegt, ungefähr 0.95.

13.3.3 Primteiler

$\omega(n)$: Anzahl der Primfaktoren von n (ohne Vielfachheiten), $\omega : \mathbb{N} \to \mathbb{N}$.

Klar: $\omega(p) = 1 \quad \forall p \in \mathbb{P}$ =Menge der Primzahlen= $\{p_1, p_2, \dots\}$, $2 =: p_1 < p_2 < \dots$: $\omega(\prod_{i=1}^{k} p_i) = k$

Läßt sich etwas über das "mittlere Verhalten" von ω aussagen?

Für jedes $n \in \mathbb{N}$ ist $(\Omega_n, \mathfrak{A}_n, \mathfrak{P}_n) := (\{1, \dots, n\}, \mathfrak{P}(\{1, \dots, n\}), unif\{1, \dots, n\})$ ein WRaum.

Sei $r_n := \max\{k : p_k < n\}$ die Anzahl der Primzahlen $\leq n$, $X_{nk}(m) : \Omega_n \to \mathbb{R}$, $X_{nk} = \begin{cases} 1 & , \quad p_k | m \\ 0 & , \quad \text{sonst} \end{cases}$ $\quad (1 \leq k \leq r)$.

Mit diesen ZV gilt $\omega(m) = \sum_{k=1}^{r_n} X_{nk}(m)$, $m = 1, \dots, n$.

Fundamentaler Sachverhalt: $p_{i_1}, \dots, p_{i_j}$ teilen $m \iff (\prod_{l=1}^{j} p_{i_l}) | m$.

$$P(X_{i_1} = 1, \dots, X_{i_j} = 1) = \frac{1}{n} \left\lfloor \frac{n}{\prod_{l=1}^{j} p_{i_l}} \right\rfloor, \quad P_n(X_i = 1) = \frac{1}{n} \lfloor \frac{n}{p_i} \rfloor.$$

Für große n gilt also näherungsweise $P_n(X_{i_1} = 1, \dots, X_{i_j} = 1) \approx P(X_{i_1} = 1) \cdot \ldots \cdot P(X_{i_j} = 1)$, d.h. Unabhängigkeit.

$$E \sum_{k=1}^{r_n} X_{nk} \approx \sum_{p<n} \frac{1}{p} \sim \log\log n, \quad var(\sum_{k=1}^{r_n} X_{nk}) \approx \sum_{p \leq n} \frac{1}{p}(1 - \frac{1}{p}) \sim \log\log n$$ (aus der Zahlentheorie), also:

Satz 13.8 (Erdös–Kac)

$$\frac{1}{n} \#\{1 \leq m \leq n : \frac{\omega(m) - \log\log n}{\sqrt{\log\log n}} \leq x\} \overset{n \to \infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{y^2}{2}) dy$$

Weitere Details, Beweis, ... $\rightsquigarrow$ Billingsley, §30.

<u>Numerisches Beispiel:</u> $\prod_{i=1}^{1000} p_i \approx 0.6786 \cdot 10^{3393}$, $\log\log(\cdot) \approx 8.96$

$n = 10^{1.000.000}$: In etwa 99.9% der Zahlen $\{1, \dots, n\}$ haben nicht mehr als 27 Primfaktoren.

13.3.4 Maximum-Likelihood-Schätzer

Statistisches Problem: Gegeben sind Daten $x_1, \dots, x_n$ als Realisierungen von unabhängigen ZV $X_1, \dots, X_n$ mit derselben Verteilung, die die Lebesgue-Dichte $f(\cdot, \theta)$ hat. Hierbei ist θ ein unbekanntes Element des Parameterraumes $\Theta \subset \mathbb{R}$ (Stichprobe vom Umfang n). Der unbekannte Parameter soll geschätzt werden.

In Analogie zu §7.1 nennen wir $l : \Theta \to \mathbb{R}$, $l(\theta)$ $(= l(\theta | x_1, \dots, x_n)) := \prod_{i=1}^{n} f(x_i | \theta)$ die <u>Likeliehood-Funktion</u>,

$\log l$ die <u>log-Likelihood-Funktion</u>,

und eine Abbildung $\hat{\theta}_n : \mathbb{R}^n \to \Theta$ mit der Eigenschaft $l(\hat{\theta}(x_1, \dots, x_n) | x_1, \dots, x_n) = \sup_{\Theta} l(\theta(x_1, \dots, x_n))$ einen <u>Maximum-Likelihood-Schätzer</u> (ML-Schätzer).

Beispiel 13.9

(i) $\Theta = (0, \infty)$, $f(x|\theta) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$ (Exponentialverteilungen)

Man erhält $\log l(\theta) = n \log \theta - \theta \sum_{i=1}^{n} x_i$ und durch Ableiten und Nullsetzen etc.

$$\hat{\theta}_n \; (= \hat{\theta}_n(x_1, \dots, x_n) = \frac{n}{\sum_{i=1}^{n} x_i}) = \frac{1}{\overline{X}_n}.$$

(ii) Die Verteilung mit Verteilungsfunktion $F(x|\theta) = \frac{1}{1 + \exp(-(x - \theta))}$, $-\infty < x < \infty$, $\theta \in \Theta = \mathbb{R}$ bezeichnet man als <u>logistische Verteilung</u> mit Lageparameter θ (und Skalenparameter 1).

Eine zugehörige Dichte ist $f(x|\theta) = \frac{\exp(-(x-\theta))}{(1 + \exp(-(x-\theta)))^2}$.

Man erhält $\sum_{i=1}^{n} \log f(x_i|\theta) = -\sum_{i=1}^{n} (x_i - \theta) - 2 \sum_{i=1}^{n} \log(1 + \exp(-(x_i - \theta)))$, also $\frac{\partial}{\partial\theta} \sum_{i=1}^{n} \log f(x_i|\theta) = n - \sum_{i=1}^{n} \frac{2}{1 + \exp(x_i - \theta)}$.

Rechts steht eine in θ streng monoton fallende Funktion, die mit $\theta \to -\infty$ gegen n und mit $\theta \to \infty$ gegen $-n$ strebt, also genau eine Nullstelle hat: Diese ist der ML-Schätzer.

In manchen Fällen hat man einfache, explizite Formeln für ML-Schätzer und kann diese benutzen, um das asymptotische Verhalten von $\hat{\theta}_n$ herzuleiten (Aufgabe 40 für Beispiel 13.9(i)).

Skizze eines Argumentes für asymptotische Normalität in allgemeinen, aber regulären Fällen:

Sei $s_n(\theta) := \frac{\partial}{\partial\theta} \log l(\theta | X_1, \dots, X_n) = \sum_{i=1}^{n} Y_i$, $Y_i := \frac{\partial}{\partial\theta} \log f(X_i|\theta)$. Die Y_i sind iid.

Mittelwertsatz: $s_n(\theta) = \underbrace{s_n(\hat{\theta}_n)}_{=0, \text{ da ML-Schätzer}} + (\theta - \hat{\theta}_n) s_n'(\tilde{\theta}_n)$, $\tilde{\theta}_n$ zwischen θ und $\hat{\theta}_n$.

Also: $(*)\quad \sqrt{n}(\hat\theta_n-\theta)=\frac{-\frac{1}{\sqrt{n}}s_n(\theta)}{\frac{1}{n}s_n'(\tilde\theta_n)},\quad E_\theta Y_i \overset{Eg(x)=\int g(x)f(x)dx}{=} \int(\frac{\partial}{\partial\theta}f(x|\theta))\frac{1}{f(x|\theta)}f(x|\theta)dx \underset{\text{Regularität}}{\overset{\text{unter}}{=}} \frac{\partial}{\partial\theta}\int f(x|\theta)dx=0.$

Mit dem ZGWS folgt nun $\frac{1}{\sqrt{n}}\sum_{i=1}^n(Y_i-EY_i)\xrightarrow{D}Z,\ Z\sim N(0,I(\theta))$ mit $I(\theta):=var_\theta Y_1=E_\theta Y_1^2=E_\theta(\frac{\partial}{\partial\theta}\log f(X_1|\theta))^2$ ($\rightsquigarrow$ Fisher-Information, vgl. §7.2)

Was passiert im Nenner von $(*)$?

Ist $\hat\theta_n$ konsistent (d.h. $\hat\theta_n\to\theta$ in W.), so muß auch $\tilde\theta_n\to\theta$ in W. gelten. Mit einer gleichmäßigen Version der Gesetzes der großen Zahlen erhält man $\frac{1}{n}s_n'(\tilde\theta_n)=\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{(\partial\theta)^2}\log f(X_1|\theta)\big|_{\theta=\tilde\theta_n}\xrightarrow{P}E_\theta\frac{\partial^2}{(\partial\theta)^2}\log f(X_1|\theta)$.

$$E_\theta\frac{\partial^2}{(\partial\theta)^2}\log f(X_1|\theta)=\int\frac{\partial}{\partial\theta}\Big(\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\Big)f(x|\theta)dx=\int\Big(\frac{\frac{\partial^2}{(\partial\theta)^2}f(x|\theta)}{f(x|\theta)}-\frac{(\frac{\partial}{\partial\theta}f(x|\theta))^2}{f(x|\theta)^2}\Big)f(x|\theta)dx$$
$$=\underbrace{\int\frac{\partial^2}{(\partial\theta)^2}f(x|\theta)\,dx}_{=0}-\int\Big(\frac{\partial}{\partial\theta}\log f(x|\theta)\Big)^2f(x|\theta)dx=-E_\theta\Big(\frac{\partial}{\partial\theta}\log f(X_1|\theta)\Big)^2=-I(\theta).$$

Insgesamt folgt aus $(*)$ mit den Eigenschaften der Verteilungskonvergenz (Aufgaben 32 & 34)

$$\boxed{\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{D}Z,\ Z\sim N(0,\tfrac{1}{I(\theta)})}$$

Unter bestimmten Bedingungen ist $\frac{1}{I(\theta)}$ die kleinst mögliche asymptotische Varianz (vergleiche die Informationsgleichung aus Satz 7.8). ML-Schätzer sind dann asymptotisch optimal.

Beispiel 13.10 (Fortsetzung von Beispiel 13.9(ii))

Im Falle $f(x|\theta)=\frac{\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}$ erhält man $\frac{\partial^2}{(\partial\theta)^2}\log f(x|\theta)=\frac{-2\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}$ und damit

$$I(\theta)=-\int\Big(\frac{\partial^2}{(\partial\theta)^2}\log f(x|\theta)\Big)f(x|\theta)dx=\int_{-\infty}^{\infty}\frac{2\exp(-(x-\theta))^2}{(1+\exp(-(x-\theta)))^4}dx=\frac{1}{3}.$$

Obige Argumentation führt also auf

(1) $\quad\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{D}Z,\ Z\sim N(0,3).$

$E_\theta(X_k-\theta)=\int(x-\theta)\frac{\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}dx=0$ (Symmetrie)

Daher ist der Mittelwert $\overline{X_n}$ ein erwartungstreuer Schätzer für θ, dieser Schätzer ist auch konsistent (nach dem schwachen Gesetz der großen Zahlen).

$\int_{-\infty}^{\infty}x^2\frac{\exp(-x)}{(1+\exp(-x))^2}dx=\frac{\pi^2}{3}$, also $var_\theta(X_1)=\frac{\pi^2}{3}$: Der ZGWS (in der Form von Satz 13.2) liefert:

(2) $\quad\sqrt{n}(\overline{X_n}-\theta)\xrightarrow{D}Z,\ Z\sim N(0,\frac{\pi^2}{3}).$

Vergleicht man (1) und (2), so sieht man, daß der ML-Schätzer eine etwas kleinere asymptotische Varianz hat als $\overline{X_n}$.

Aussagen zur asymptotischen Normalität eines Schätzers lassen sich zur Konstruktion von asymptotischen Konfidenzintervallen verwenden (vergleiche §7.5). Hat man $\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{D}Z,\ Z\sim N(0,\sigma^2)$, so ist $[\hat\theta_n-\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}},\hat\theta_n+\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}]$ ein asymptotisch korrektes $100(1-\alpha)\%$ Konfidenzintervall für θ (hierbei: $P(Z\le u_\alpha)=\alpha$ bei $Z\sim N(0,1)$).

Kleinere asymptotische Varianz bedeutet: kürzere Intervalle.

Will man in der Situation von Beispiel 13.9 den schlechteren Wert von $\overline{X_n}$ durch größeres n ausgleichen, so wird man auf $\frac{n(\overline{X_n})}{n(\hat\theta_n)}=\frac{\frac{\pi^2}{3}}{3}\approx1.0966$ geführt, d.h. man braucht etwas 10% mehr Beobachtungen.

Kapitel 14

Zufallsvektoren

14.1 Allgemeines

Hier: Zufallsgrößen X mit Werten in $(\mathbb{R}^d, \mathfrak{B}^d)$.
Elemente von $\mathbb{R}^d$ sind Spaltenvektoren.

$< \cdot, \cdot >$ ist das Skalarprodukt: $< x, y >= \sum\limits_{k=1}^{d} x_k y_k,\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix}$.

Ist A^t die Transponierte einer $n \times m$-Matrix $A \in \mathbb{R}^{n\times m}$, und faßt man Elemente des $\mathbb{R}^d$ als $d \times 1$-Matrizen auf, so gilt $< x, y >= x^t y$.
Viele der Resultate der früheren Abschnitte lassen sich von $\mathbb{R}$ auf $\mathbb{R}^d$, $d > 1$, verallgemeinern (i.a. hier ohne Beweis.)
Ist X ein d-dimensinaler Zufallsvektor auf $(\Omega, \mathfrak{A}, P)$, so sind die Komponentenabbildungen $X_1, \dots, X_d$ "gewöhnliche" Zufallsvariablen (auch die Umkehrung hiervon gilt, siehe Aufgabe 15.)

Existieren zu diesen die Erwartungswerte, so nennen wir $EX := \begin{pmatrix} EX_1 \\ \vdots \\ EX_d \end{pmatrix}$ der Erwartungswertvektor zu X (die Linearität bleibt erhalten.)

Hat man $EX_i^2 < \infty$ für $i = 1, \dots, d$ (und damit auch $E|X_i X_j| < \infty$ für alle $i \neq j$ ($\rightsquigarrow$ Cauchy-Schwarz-Ungleichung), so nennt man

$$\boxed{cov(X) := \big(cov(X_i, X_j)\big)_{i,j=1}^{d}}$$

die Kovarianzmatrix zu X (Varianzmatrix wäre eigentlich besser.)
Auf der Diagonalen stehen die Varianzen der Komponenten.
Viele Rechenregeln lassen sich leicht übertragen: Ist z.B. $A : \mathbb{R}^d \to \mathbb{R}^k$ eine lineare Abbildung (also $k \times d$-Matrix), so gilt $E(AX) = AEX$ (Reihenfolge beachten, nicht $(EX)A$!!)
Dehnt man der Erwartungswertbegriff auf zufällige Matrizen aus (wieder komponentenweise), so gilt:

$$cov(X) = E(X - EX)(X - EX)^t.$$

Damit

$$\begin{aligned} cov(AX + b) &= E((AX + b) - (AEX + b))(AX + b - (AEX + b))^t) = \\ &= E((A(X - EX))(A(X - EX))^t) = AE((X - EX)(X - EX)^t)A^t = A \cdot cov(X) \cdot A^t. \end{aligned}$$

Ist X ein d-dimensionaler Zufallsvektor auf $(\Omega, \mathfrak{A}, P)$, so wird durch $\varphi_X : \mathbb{R}^d \to \mathbb{C}$, $\varphi_X(\theta) = E\exp(i < \theta, X >)\quad \forall \theta \in \mathbb{R}^d$ die charakteristische Funktion zu X definiert. Auch hier gilt $\varphi_X = \varphi_Y \implies \mathfrak{L}(X) = \mathfrak{L}(Y)$ (Eindeutigkeitssatz).

Gibt es "interessante" mehrdimensionale Verteilungen?

14.2 Mehrdimensionale Normalverteilungen

Ein d-dimensionaler Zufallsvektor X heißt (d-dimensional) normalverteilt, wenn alle Linearkombinationen von Komponenten von X, also alle Zufallsvariablen der Form $< a, X >$, $a \in \mathbb{R}^d$, im 1-dimensionalen Sinn normalverteilt sind. (Im Fall $Z \equiv c \in \mathbb{R}$ betrachten wir Z als $N(c, 0)$-verteilt.)
Man hat $X_i = (0, \dots, 0, 1, 0, \dots, 0)X$, also existiert der Erwartungswert μ und die Kovarianzmatrix Σ zu X nach den Rechenregeln aus §14.1.:

$Ea^tX = a^t\mu \qquad \forall a \in \mathbb{R}^d,\ var(a^tX) = a^t\Sigma a.$

Damit $\varphi_{a^tX}(\theta) = \exp(i\theta(a^t\mu) - \frac{1}{2}\theta^2(a^t\Sigma a))\quad \forall \theta \in \mathbb{R} \qquad$ ($\rightsquigarrow$ Beispiel 11.8)
Dies liefert $\varphi_X(\theta) = E\exp(i < \theta, X >) = E\exp(i \cdot 1 \cdot \theta^t X) = \varphi_{\theta^t X}(1) = \exp(i < \theta, \mu > -\frac{1}{2}\theta^t\Sigma\theta)$.

Erste Konsequenz: Auch im mehrdimensionalen Fall ist eine Normalverteilung durch die beiden Parameter μ und Σ eindeutig festgelegt. Weiter folgt hieraus (oder direkt aus der Definition), wie sich Normalverteilungen unter affinen

Transformationen verhalten:
$\varphi_{AX+b}(\theta) = E\exp(i < \theta, AX+b >)$
$= \exp(i < \theta, b >) E\exp(i < A^t\theta, X >)$
$= \exp(i < \theta, b >)\varphi_X(A^t\theta) \qquad = \exp(i < \theta, b + A\mu > -\frac{1}{2}\theta^t(A\Sigma A^t)\theta)$
Also: $X \sim N_d(\mu, \Sigma),\ A \in \mathbb{R}^{k\times d},\ b \in \mathbb{R}^k \implies Y := AX + b \sim N_k(A\mu + b, A\Sigma A^t)$.

Zu welchen $\mu \in \mathbb{R}^d,\ \Sigma \in \mathbb{R}^{d\times d}$ existiert eine solche Verteilung?
Wegen $a^t\Sigma a = var(a^tX) \geq 0 \quad \forall a \in \mathbb{R}^d$ muß Σ positiv semidefinit sein. Aus der Definition folgt, daß Σ symmetrisch sein muß. Zu jeder symmetrischen, positiv semidefiniten Matrix existiert eine "Wurzel", d.h. eine Matrix $A \in \mathbb{R}^{d\times d}$ mit $AA^t = \Sigma$.
Es seien nun $X_1, \ldots, X_d$ unabhängige, $N(0,1)$-verteilte Zufallsvariablen. Für $X = (X_1, \ldots, X_d)^t$ gilt

$$a^tX = \sum_{i=1}^d a_iX_i \sim N(0, \sum_{i=1}^d a_i^2) = N(0, a^tI_da).$$

Die Definition normalverteilter Zufallsvektoren zeigt also, daß X d-dimensional normalverteilt ist.
Klar: $X \sim N_d(0, I_d)$.
Für $Y := AX + \mu$ erhält man $Y \sim N_d(A\cdot 0 + \mu, AI_dA^t) = N_d(\mu, \Sigma)$.
Bei nicht-singulärem Σ ist auch A nicht-singulär und mit Hilfe einer geeigneten Formel für das Verhalten von Dichten unter affinen Transformationen folgt, daß dann $N_d(\mu,\Sigma)$ die (l^d-) Dichte

$$\boxed{(*) \qquad f(x|\mu,\Sigma) = (2\pi)^{-\frac{d}{2}}(det\Sigma)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu))} \quad \text{hat.}$$

Bei $det\Sigma = 0$ gibt es ein $a \in \mathbb{R}^d$ mit $a^t\Sigma a = 0$, also $var(a^tX) = 0$, d.h. X ist auf $H := \{x \in \mathbb{R}^d : a^tx = a^t\mu\}$ konzentriert. $l^d(H) = 0$, d.h. es existiert keine Dichte.

Konstruktion zu $N(\mu,\Sigma)$ begann mit $X_1, \ldots, X_d,\ X_i \sim N(0,1)$. Dann gilt $X = (X_1, \ldots, X_d)^t \sim N_d(0, I_d)$.

Hat man umgekehrt $X \sim N_d(0, I_d)$, dann gilt $f_X(x_1, \ldots, x_d) = (2\pi)^{-\frac{d}{2}}\exp(-\frac{1}{2}\overbrace{\sum_{i=1}^d x_i^2}^{x^tx})$, d.h. die gemeinsame Dichte der Komponenten $X_1, \ldots, X_d$ ist das Produkt der einzelnen Dichten $f_{X_i}(x_i) = (2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}x_i^2)$, d.h. $X_1, \ldots, X_d$ sind unabhängig und $N(0,1)$-verteilt.

Definition 14.1

(i) Das WMaß auf $(\mathbb{R}_+, \mathfrak{B}_+)$ mit Dichte $f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}},\ x > 0$ heißt Chiquadrat-Verteilung mit n Freiheitsgraden (kurz: χ^2_n.)

(ii) Die Studentsche t-Verteilung mit n Freiheitsgraden, kurz: t_n, ist die Verteilung auf $(\mathbb{R}, \mathfrak{B})$ mit Dichte
$$f(x) = \frac{1}{\sqrt{\pi n}}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}.$$

Klar: $\chi^2_n = \Gamma(\frac{n}{2}, \frac{1}{2})$ (siehe Aufgabe 26)

Lemma 14.2 (i) Sind X und Y unabhängige ZV mit Dichte f_X und f_Y, wobei $Y > 0$, so hat $Z := \frac{X}{Y}$ die Dichte

$$\boxed{f_Z(x) = \int_0^\infty yf_Y(y)f_X(yz)dy}.$$

(ii) Sind X und Y unabhängig mit $X \sim N(0,1),\ Y \sim \chi^2_n$, so gilt: $\frac{\sqrt{n}X}{\sqrt{Y}} \sim t_n$.
Beweis: Übungen.

Satz 14.3 Es sei $X_1, \ldots, X_n$ eine Stichprobe aus $N(\mu, \sigma^2)$.
$\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$ (Stichprobenmittelwert), $S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$ (Stichprobenvarianz), $T_n = \frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n}$. Dann gilt:

(i) $\overline{X_n}$ und S_n^2 sind unabhängig.
(ii) $\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma} \sim N(0,1)$.
(iii) $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$.
(iv) $T_n \sim t_{n-1}$.

Beweis: Sei $Y_i := \frac{X_i - \mu}{\sigma},\ 1 \leq i \leq n$. Dann ist $Y_1, \ldots, Y_n$ eine Stichprobe aus $N(0,1)$, und es gilt:

$$\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma} = \sqrt{n}\overline{Y_n},\ (n-1)\frac{S_n^2}{\sigma^2} = \sum_{i=1}^n (Y_i - \overline{Y_n})^2.$$

Sei nun A eine orthogonale $n \times n$-Matrix mit erster Zeile $(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$ (kann mit Gram-Schmidt konstruiert werden, für uns reicht hier die Existenz). Sei $Y := (Y_1, \ldots, Y_n)^t,\ Z := AY$. Dann gilt: $Y \sim N_n(0, I_n),\ Z \sim N_n(0, AI_nA^t) = N_n(0, I_n)$.
Also bilden die $Z_1, \ldots, Z_n$ eine Stichprobe aus $N(0,1)$. Nach Konstruktion gilt $Z_1 := \sqrt{n}\overline{Y_n}$, also folgt (ii).
Die Orthogonalität von A liefert außerdem $\sum_{i=1}^n Z_i^2 = Z^tZ = Y^tY = \sum_{i=1}^n Y_i^2$.

$$\sum_{i=1}^{n}(Y_i-\overline{Y_n})^2=\sum_{i=1}^{n}Y_i^2-2\overline{Y_n}\underbrace{\sum_{i=1}^{n}Y_i}_{n\overline{Y_n}}+n\overline{Y_n}^2=\sum_{i=1}^{n}Y_i{}^2-n(\overline{Y_n})^2=\sum_{i=1}^{n}Z_i^2-Z_1^2=\sum_{i=2}^{n}Z_i^2.$$

Folglich ist S_n^2 eine Funktion von $Z_2,\dots,Z_n$ und damit unabhängig von $\overline{X_n}$, das ja eine Funktion von Z_1 ist. Hiermit folgt (i).
Die Bezeichnung $\frac{(n-1)S_n^2}{\sigma^2}=\sum_{i=2}^{n}Z_i^2$ liefert auch (iii), den nach Aufgabe 33(a) zur Stochastik I ist Z_i^2 χ_1^2-verteilt, und mit der Faltungsgleichung der Gammafunktion (Aufgabe 26) folgt hieraus $\sum_{i=2}^{n}Z_i^2\sim\chi_{n-1}^2\quad(=\Gamma(\frac{n}{2},\frac{1}{2}))$.
Teil (iv) folgt nun mit Lemma 14.2(ii) aus den bereits bewiesenen Teilen (i)-(iii). □

Bemerkung 14.4 (Tests und Konfidenzintervalle für Erwartungswert bei Stichproben aus der Normalverteilung)
Es sei $t_{n;\alpha}$ das α-Quantil zur Verteilung t_n (siehe Tafelwerk, mit Computerprogramm).
Ist $X_1,\dots,X_n$ eine Stichprobe aus $N(\mu,\sigma^2)$ mit unbekanntem $\mu\in\mathbb{R}$ und $\sigma^2>0$, so folgt mit Satz 14.3(iv)

$$P_{\mu,\sigma^2}(t_{n-1,\frac{\alpha}{2}}\le\frac{\overline{X_n}-\mu}{\frac{S_n}{\sqrt{n}}}\le t_{n-1;1-\frac{\alpha}{2}})=1-\alpha$$

Dies gilt für alle $\mu\in\mathbb{R},\ \sigma^2>0$. Löst man die Ungleichung nach μ auf, so sieht man, daß

$$[\overline{X_n}-\tfrac{1}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}}S_n,\overline{X_n}-\tfrac{1}{\sqrt{n}}t_{n-1;\frac{\alpha}{2}}S_n]$$

ein $100(1-\alpha)\%$-Konfidenzintervall für μ ist.
Die Symmetrie der t-Verteilung impliziert $t_{n-1;\frac{\alpha}{2}}=-t_{n-1;1-\frac{\alpha}{2}}$ (damit: $\overline{X_n}\pm\frac{1}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}}S_n$). Für große n kann t_n durch $N(0,1)$ ersetzt werden (siehe Aufgabe 48).

Der in §7.4 besprochene Zusammenhang zwischen Tests und Konfidenzintervallen führt auf den von Anwendern sehr (zu) häufig benutzten t-Test zur Hypothese $H_0:\mu=\mu_0$ (μ_0 ist ein vorgegebener Wert):
Verwerfe, wenn $|\overline{X_n}-\mu_0|>\frac{1}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}}S_n$.
Die obige Rechnung zeigt, daß dies ein Test zum Niveau α ist, der bemerkenswerter Weise für alle $\sigma^2>0$ das Niveau α, die zugelassene maximale Wahrscheinlichkeit für einen Fehler 1. Art, "ausschöpft".
Es gibt hiervon auch 'einseitige' Versionen für Hypothesen der Form $H_0:\mu\le\mu_0,\ H_0:\mu\ge\mu_0$.

14.3 Grenzwertsätze

Gesetze der Großen Zahlen (stark und schwach)
Erweiterung von $\mathbb{R}$ auf $\mathbb{R}^d$ ist einigermaßen trivial (komponentenweise). Beim ZGWS muß zunächst "$\xrightarrow{D}$" definiert werden: $X_n\xrightarrow{D}X$ bedeutet $\lim_{n\to\infty}F_n(x)=F(x)\quad\forall$ Stetigkeitspunkte von F, wobei $F_n(n\in\mathbb{N}),\ F$ die Verteilungsfunktionen zu $X_n(n\in\mathbb{N}),\ X$.
Es gilt der Stetigkeitssatz für charakteristische Funktionen $X_n\xrightarrow{D}X\iff\varphi_{X_n}(\theta)\to\varphi_X(\theta)\quad\forall\theta\in\mathbb{R}^d$.
Damit:

Satz 14.5 (ein mehrdimensionaler ZGWS)
Es sei $(X_n)_{n\in\mathbb{N}}$ eine iid-Folge von d-dimensionalen Zufallsvariablen mit Erwartungswertvektor μ und Kovarianzmatrix Σ. Dann gilt:

$$\boxed{\sqrt{n}(\overline{X_n}-\mu)\to Z,\ Z\sim N_d(0,\Sigma).}$$

Beweis: Sei $\sigma\in\mathbb{R}^d$. Setze $Y_i:=\theta^tX_i,\ u\in\mathbb{N}$. Auf diese läßt sich Satz 13.2 anwenden:
$\sqrt{n}(\overline{Y_n}-\theta^t\mu)\longrightarrow Z_\theta,\ Z_\theta\sim N(0,\theta^t\Sigma\theta)$.
Der Stetigkeitssatz (für dim $=1$) liefert nun $E\exp(i(\sqrt{n}(\overline{Y_n}-\theta^t\mu)))\longrightarrow\exp(-\frac{1}{2}\theta^t\Sigma\theta)$.
Da θ beliebig war, und $\overline{Y_n}=\theta^t\overline{X_n}$ gilt, folgt hieraus $E\exp(i\theta^t(\sqrt{n}(\overline{X_n}-\mu)))\longrightarrow E\exp(i\theta^tZ)\quad\forall\theta$, also mit dem "mehrdimensionalen" Stetigkeitssatz die Behauptung. □

Beispiel 14.6
Gegeben sei ein stochastisches Experiment mit d möglichen Resultaten, die mit den Wahrscheinlichkeiten $p_1,\dots,p_d$ eintreten. X_n sei der d-dimensionale Zufallsvektor, dessen k-te Komponente 1 oder 0 ist, je nachdem, ob Resultat k in der n-ten Wiederholung eingetreten ist oder nicht. Dann beschreibt $S_n:=X_1+\ldots+X_n\ (=\overline{X_n})$ die Verteilung der Resultate nach n Durchgängen. $S_n\sim Multinom(n,p),\ p=(p_1,\dots,p_d)^d$ ($\rightsquigarrow$ Stochastik I, §6.2.5)
Aufgabe 45: $\mu=EX=p,\ \Sigma=diag(p)-pp^t$, in Komponenten: $(\Sigma)_{ij}=\begin{cases}p_i-p_i^2 & ,\quad i=j\\ -p_ip_j & ,\quad \text{sonst}\end{cases}$.

Satz 14.5 liefert $\quad(*)\quad\frac{1}{\sqrt{n}}(S_n-n\mu)\xrightarrow{D}Z,\ Z\sim N_d(0,\Sigma)$.

Die Testgröße des populären χ^2-Anpassungstests auf p lautet: $T(X_1,\dots,X_n):=\frac{1}{n}\sum_{i=1}^{n}\frac{(S_{ni}-np_i)^2}{p_i}$.

Konkretes Beispiel mit echten Daten:

189 Würfelwürfe haben das folgende Resultat ergeben:

i	1	2	3	4	5	6
Werte von S_{ni}	30	37	26	29	29	38

Ist der Würfel fair? Man erhält $T(X) = 3.73\ldots$

$Y_n := (\sqrt{n}\frac{\overline{X_{n1}}-p_1}{\sqrt{p_1}}, \ldots, \sqrt{n}\frac{\overline{X_{nd}}-p_d}{\sqrt{p_d}})^t = diag(\frac{1}{\sqrt{p_1}}, \ldots, \frac{1}{\sqrt{p_d}})\sqrt{n}(\overline{X_n} - p)$

Aus $(*)$ folgt: $Y_n \xrightarrow{D} diag(\frac{1}{\sqrt{p_1}}, \ldots, \frac{1}{\sqrt{p_d}})Z = \tilde{Z}$ (continuous mapping theorem, Eigenschaften von $\xrightarrow{D}$).

$\tilde{Z}$ ist wieder normalverteilt, mit Mittelwertvektor 0 und Kovarianzmatrix

$$\tilde{\Sigma} = diag(\frac{1}{\sqrt{p_1}}, \ldots, \frac{1}{\sqrt{p_d}})(diag(p) - pp^t)diag(\frac{1}{\sqrt{p_1}}, \ldots, \frac{1}{\sqrt{p_d}}) = I_d - \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix} (\sqrt{p_1}, \ldots, \sqrt{p_d}).$$

Der Vektor $\begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix}$ läßt sich zu einer orthogonalen Matrix A ergänzen (vergleiche Beweis zu Satz 14.3)

Es sei $\overline{Z} := A^t\tilde{Z}$. Dann ist $\overline{Z}$ wieder normalverteilt mit $E\overline{Z} = 0$ und Kovarianzmatrix $\overline{\Sigma} = A^t\tilde{\Sigma}A = \underbrace{A^tI_dA}_{I_d} - \underbrace{A^t\begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_d} \end{pmatrix}}_{(1,0,\ldots,0)^t}\underbrace{(\sqrt{p_1}, \ldots, \sqrt{p_d})A}_{(1,0,\ldots,0)} = diag(0, 1, \ldots, 1)$

Also: $\overline{Z_1} = 0$ f.s., $\overline{Z_2}, \ldots, \overline{Z_d}$ unabhängig und $N(0,1)$-verteilt.

Damit $\overline{Z}^t\overline{Z} = \sum_{i=2}^{d} \overline{Z_i}^2 \sim \chi^2_{d-1}$ (siehe Ende von Beweis zu Satz 14.3)

Insgesamt $T_{(n)} = Y_n^tY_n \xrightarrow{D} \tilde{Z}^t\tilde{Z} = \overline{Z}^t\overline{Z} \sim \chi^2_{n-1}$ (continuous mapping theorem)

Wir haben damit die asymptotische Verteilung der Testgröße; diese hängt nicht von p ab.

Also: Wähle bei Signifikanzniveau α als <u>kritischen Wert</u> c_α (d.h. Ablehnung bei $T > c_\alpha$) das $1-\alpha$-Quantil zu χ^2_{d-1} ($\rightsquigarrow$ Tafel, Computer)

Dieser Test hält bei großem n das Niveau α ein. Faustregel: $np_i \geq 5$ für $i = 1, \ldots, d$.

Im numerischen Beispiel: $\chi^2_{5;1-0.05} = 11.1\ldots$, somit keine Ablehnung.

Kapitel 15

Martingale

Wichtiges Hilfsmittel der modernen Stochastik und hat zahlreiche Anwendunngen (u.a. bei Glücksspielstrategien)

15.1 Bedingte Erwartungswerte

$(\Omega, \mathfrak{A}, P)$ WRaum, $\mathfrak{F}$ Unter-σ-Algebra von $\mathfrak{A}$.

Definition 15.1 Es sei X eine Zufallsvariable mit existierendem Erwartungswert (also $E|X| \leq \infty$). Dann heißt $Y : \Omega \to \mathbb{R}$ eine Version des bedingten Erwartungswertes von X unter F, wenn gilt:

$$\boxed{Y \text{ ist } \mathfrak{F}\text{-meßbar}, \int_F Y dP = \int_F X dP}$$

Extremfälle:

- $\mathfrak{F} = \{\emptyset, \Omega\}$. Alle $\mathfrak{F}$-meßbaren $Y : \Omega \to \mathbb{R}$ sind konstant. Wähle $Y \equiv EX$: $\int_F Y dP = EX \cdot P(F) = \int_F X dP$, da nur $F = \emptyset$, $F = \Omega$ in Frage kommen.
- $\mathfrak{F} = \mathfrak{A}$. Wähle $Y = X$.

Satz 15.2 Es sei X eine Zufallsvariable mit $E|X| < \infty$. Dann gilt:

(i) Es existiert eine Version der bedingten Erwartung von X unter $\mathfrak{F}$.

(ii) Sind Y_1 und Y_2 Versionen der bedingten Erwartung von X unter $\mathfrak{F}$, so gilt $P(Y_1 = Y_2) = 1$ (In Worten: bedingte Erwartungswerte sind fast sicher eindeutig).

(iii) Ist Y eine Version der bedingten Erwartung von X unter $\mathfrak{F}$, so exitiert zu Y der Erwartungswert, und es gilt $EY = EX$.

Beweis:

(i) Es sei P_0 die Restriktion von P auf $\mathfrak{F}$. Durch $\nu_+(F) = \int_F X^+ dP$ für alle $F \in \mathfrak{F}$ (wobei $X^+ := \max\{X, 0\}$) wird ein P_0-stetiges Maß auf $(\Omega, \mathfrak{F})$ definiert ($P_0(N) = 0 \implies \nu_+(N) = 0$), also existiert nach dem Satz von Radon-Nikodym eine Dichte, d.h. eine $\mathfrak{F}$-meßbare Abbildung $Y^+ : \Omega \to \mathbb{R}^+$ mit

$$\int_F X^+ dP = \nu_+(F) = \int_F Y^+ dP_0 = \int_F Y^+ dP \quad \forall F \in \mathfrak{F}.$$

Verfahre analog bei X^-, und setze $Y := Y^+ - Y^-$.

$$\underline{\int_F Y dP} = \int_F Y^+ dP - \int_F Y^- dP = \int_F X^+ dP - \int_F X^- dP = \underline{\int_F X dP}.$$

(ii) $F_1 := \{Y_1 < Y_2\}$, $F_2 := \{Y_1 > Y_2\}$, beide in $\mathfrak{F}$.

$$\int |Y_1 - Y_2| dP = \int_{F_1} (Y_2 - Y_1) dP + \int_{F_2} (Y_1 - Y_2) dP = \int_{F_1} Y_2 dP - \int_{F_1} Y_1 dP + \int_{F_2} Y_1 dP - \int_{F_2} Y_2 dP =$$

$$\int_{F_1} X dP - \int_{F_1} X dP + \int_{F_2} X dP - \int_{F_2} X dP = 0,$$ also $|Y_1 - Y_2| = 0$ P-f.s. und damit $Y_1 = Y_2$ P-f.s.

(iii) $\int Y^+ dP = \int_{\{Y>0\}} Y dP = \int_{\{Y>0\}} X dP \leq \int |X| dP < \infty$, analog: $\int Y^- dP < \infty$, also ist Y P-integrierbar.

Mit $\Omega \in \mathfrak{F}$ folgt schließlich $EY = \int_\Omega Y dP = \int_\Omega X dP = EX$. □

Aus Abschnitt 8.7: $\mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) = \{f : \Omega \to \mathbb{R} : \int |f|^p d\mu < \infty\}$, $||f||_p := (\int |f|^p d\mu)^{\frac{1}{p}}$.

Auf diesen Räumen wird durch $f \sim g \iff \mu(f \neq g) = 0$ eine Äquivalenzrelation definiert. Der Raum $L^p(\Omega, \mathfrak{A}, \mu)$ der Äquivalenzklassen $\tilde{f} := \{g \in \mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) : g \sim f\}$ wird mit $||\tilde{f}||_p := ||f||_p$ (hängt nicht von der Wahl von $f \in \tilde{f}$ ab) zu einem Banach-Raum.

Im Falle $p = 2$ hat man mit $< \tilde{f}, \tilde{g} > := \int f g d\mu$ sogar einen Hilbert-Raum[1] (auf $\mathfrak{L}^p$ ist $|| \cdot ||_p$ nur eine Seminorm, da

[1] Wo ist alles erlaubt? — Im Krieg, in der Liebe und im Hilbert-Raum!

$||f||_p = 0$ auch bei $f \neq 0$ gelten kann.)

Nach Satz 15.2 (ii) bilden die verschiedenen Versionen der bedingten Erwartungswerte von X bzgl. $\mathfrak{F}$ eine Äquivalenzklasse (d.h. $L^1(\Omega, \mathfrak{F}, P) \subset L^1(\Omega, \mathfrak{A}, P)$), wir schreiben hierfür $E[X|\mathfrak{F}]$ bzw. $E[X|Z]$ oder $E[X|Z_i, i \in I]$ im Falle $\mathfrak{F} = \sigma(Z)$ bzw. $\mathfrak{F} = \sigma(\{Z_i : i \in I\})$. Man kann den Übergang von X zu $E[X|\mathfrak{F}]$ als Projektion von $L^1(\Omega, \mathfrak{A}, P)$ auf $L^1(\Omega, \mathfrak{F}, P)$ auffassen. Im folgenden werden wir den Unterschied Funktion/Äquivalenzklasse weitgehend ignorieren.
Konkret: Aussagen zu bedingten Erwartungswerten gelten i.a. nur P-f.s., z.B. $E[E[X|\mathfrak{F}]|\mathfrak{F}] = E[X|\mathfrak{F}]$ P-f.s.

Satz 15.3

(i) $E[\alpha X + \beta Y|\mathfrak{F}] = \alpha E[X|\mathfrak{F}] + \beta E[Y|\mathfrak{F}]$ P-f.s. (Linearität)

(ii) $X \leq Y \implies E[X|\mathfrak{F}] \leq E[Y|\mathfrak{F}]$ P-f.s. (Monotonie)

(iii) $\varphi : \mathbb{R} \to \mathbb{R}$ konvex und gilt $E|\varphi(X)| < \infty$, so folgt $\varphi(E[X|\mathfrak{F}]) \leq E[\varphi(X)|\mathfrak{F}]$ P-f.s. (bedingte Jensensche Ungleichtung)

Beweis:

(i) Sei $F \in \mathfrak{F}$. Dann gilt:

$$\int_F (\alpha E[X|\mathfrak{F}] + \beta E[Y|\mathfrak{F}])dP = \alpha \int_F E[X|\mathfrak{F}]dP + \beta \int_F E[Y|\mathfrak{F}]dP = \alpha \int_F XdP + \beta \int_F YdP = \int_F (\alpha X + \beta Y)dP$$

Da mit $E[X|\mathfrak{F}]$, $E[Y|\mathfrak{F}]$ auch $\alpha E[X|\mathfrak{F}] + \beta E[Y|\mathfrak{F}]$ $\mathfrak{F}$-meßbar ist, leistet die Linearkombination das Verlangte.

(ii) Wegen der bereits bewiesenen Linearität, reicht es, den Spezialfall $X \geq 0 \implies E[X|\mathfrak{F}] \geq 0$ P-f.s. zu beweisen.
Sei $F := \{E[X|\mathfrak{F}] < 0\}$ $(\in \mathfrak{F})$.

$$0 \geq \int_F E[X|\mathfrak{F}]dP = \int_F XdP \geq 0.$$

Also: $\int_F E[X|\mathfrak{F}]dP = 0$, d.h. $P(E[X|\mathfrak{F}]1_F = 0) = 1$ und damit $P(E[X|\mathfrak{F}] < 0) = 0$.

(iii) Übungsaufgabe. □

Satz 15.4

(i) (bedingte Version des Satzes von der monotonen Konvergenz)
Sind $X, X_1, X_2, \ldots$ Zufallsvariablen mit $0 \leq X_n \uparrow X$, so gilt $E[X_n|\mathfrak{F}] \uparrow E[X|\mathfrak{F}]$ P-f.s..

(ii) (bedingte Version des Lemmas von Fatou)
Sind $X_1, X_2, \ldots$ nicht negative Zufallsvariablen, so gilt $E[\liminf\limits_{n\to\infty} X_n|\mathfrak{F}] \leq \liminf\limits_{n\to\infty} E[X_n|\mathfrak{F}]$ P-f.s..

(iii) (bedingte Version des Satzes von der majorisierten Konvergenz)
Sind $X, Y, X_1, X_2, \ldots$ Zufallsvariablen mit $X_n \to X$ P-f.s., $|X_n| \leq Y$ und $EY < \infty$, so gilt $E[X_n|\mathfrak{F}] \to E[X|\mathfrak{F}]$ P-f.s..

Beweis:

(i) Die Monotonie von $Y_n := E[X_n|\mathfrak{F}]$ folgt aus der Monotonie-Eigenschaft des bedingten Erwartungswerts (Satz 15.3 (ii)): Sei $Y := \sup\limits_{n\in\mathbb{N}} Y_n$, dann: $Y_n \uparrow Y$ (P-f.s.)
Nach dem "unbedingten" Satz von der monotonen Konvergenz, angewendet auf die Folgen $(1_F Y_n)_{n\in\mathbb{N}}$ und $(1_F X_n)_{n\in\mathbb{N}}$ gilt für jedes $F \in \mathfrak{F}$: $\int_F YdP = \sup_n \int_F Y_n dP = \sup_n \int_F X_n dP = \int_F XdP$.
Da Y auch $\mathfrak{F}$-meßbar ist, leistet es das Verlangte.

(ii),(iii) Übungsaufgaben □

Einige Eigenschaften, die kein Analogon bei gewöhnlichen Erwartungswerten haben (können):

Satz 15.5

(i) Ist $\mathfrak{G}$ eine Unter-σ-Algebra von $\mathfrak{F}$, so gilt $E[E[X|\mathfrak{F}]|\mathfrak{G}] = E[X|\mathfrak{G}]$

(ii) Ist Y $\mathfrak{F}$-meßbar, so gilt: $E[Y \cdot X|\mathfrak{F}] = Y \cdot E[X|\mathfrak{F}]$ P-fast sicher, vorausgesetzt Y ist beschränkt oder $E|Y|^p < \infty$, $E|X|^q < \infty$ für ein Paar (p, q) mit $\frac{1}{p} + \frac{1}{q} = 1$ (wird gebraucht für $E|X \cdot Y| < \infty$).

(iii) Ist $\mathfrak{G}$ eine von $\sigma(\{\sigma(X), \mathfrak{F}\})$ unabhängige σ-Algebra, so gilt: $E[X|\sigma(\mathfrak{F}, \mathfrak{G})] = E[X|\mathfrak{F}]$ P-fast sicher.
Inbesondere: Sind $\sigma(X)$ und $\mathfrak{G}$ unabhängig, so gilt $E[X|\mathfrak{G}] = EX$ P-f.s..

Beweis:

(i) $E[X|\mathfrak{G}]$ ist $\mathfrak{G}$-meßbar. Für alle $G \in \mathfrak{G}$ gilt $\int_G E[X|\mathfrak{G}]dP = \int_G XdP = \int_G E[X|\mathfrak{F}]dP$, da $G \in \mathfrak{F}$, also leistet $E[X|\mathfrak{G}]$ das von $E[E[X|\mathfrak{F}]|\mathfrak{G}]$ verlangte.

(ii) Ein Fall für die "übliche Maschinerie":
Sei $Y = 1_A$ mit $A \in \mathfrak{F}$. Dann gilt für alle $F \in \mathfrak{F}$:

$$\int_F Y \cdot E[X|\mathfrak{F}]dP = \int_{\underbrace{F \cap A}_{\in \mathfrak{F}}} E[X|\mathfrak{F}]dP = \int_{F\cap A} XdP = \int_F 1_A XdP = \int_F X \cdot YdP.$$

$Y \cdot E[X|\mathfrak{F}]$ ist $\mathfrak{F}$-meßbar, d.h. die Behauptung gilt für Indikatorfunktionen.

Linearität (Satz 15.3.(i)) liefert die Behauptung für primitive Y. Die bedingte Version des Satzes von der monotonen Konvergenz (Satz 15.4.(i)) liefert die Behauptung für $Y \geq 0$. Im letzten Schritt zerlege in Positiv- und Negativteil.

(iii) Wir können $X \geq 0$ annehmen (sonst X^+, X^- getrennt betrachten)
Betrachte $A \to \int_A XdP$, $A \to \int_A E[X|\mathfrak{F}]dP$. (Nach Definition von $E[X|\mathfrak{F}]$ stimmen diese Abbildungen auf $\mathfrak{F}$ überein.) Sei $F \in \mathfrak{F}$, $G \in \mathfrak{G}$. Da 1_G und $X1_F$ unabhängig sind, ebenso wie 1_G und $E[X1_F|\mathfrak{F}] = 1_F E[X|\mathfrak{G}]$ folgt:
$\int_{F\cap G} XdP = \int 1_G(X \cdot 1_F)dP \overset{\text{Multiplikations-}}{\underset{\text{satz}}{=}} \int 1_G dP \cdot \int X1_F dP = \int 1_G dP \cdot \int_F E[X|\mathfrak{F}]dP \overset{\text{Mult.-}}{\underset{\text{satz}}{=}} \int 1_G 1_F E[X|\mathfrak{F}]dP =$
$\int_{F\cap G} E[X|\mathfrak{F}]dP$, d.h. die beiden Maße stimmen auf einem $\cap$-stabilen Erzeuger von $\sigma(\mathfrak{F},\mathfrak{G})$ überein.
Da sie außerdem dieselben Gesamtmassen haben, sind sie gleich, d.h. $\int_A E[X|\mathfrak{F}]dP = \int_A XdP$ für alle $A \in \sigma(\mathfrak{F},\mathfrak{G})$.
Wegen der $\sigma(\mathfrak{F},\mathfrak{G})$-meßbarkeit von $E[X|\mathfrak{F}]$ (folgt aus $\mathfrak{F} \subset \sigma(\mathfrak{F},\mathfrak{G})$) hat man daher $E[X|\mathfrak{F}] = E[X|\sigma(\mathfrak{F},\mathfrak{G})]$ P-f.s..
Wähle $\mathfrak{F} = \{\emptyset, \Omega\}$. Dann $\sigma(\mathfrak{F},\mathfrak{G}) = \mathfrak{G}$, also $E[X|\mathfrak{G}] = E[X|\mathfrak{F}] = EX$ P-f.s.. □

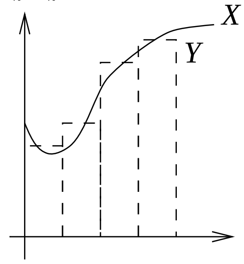
Satz 15.6 Gilt $EX^2 < \infty$, so minimiert $E[X|\mathfrak{F}]$ die Funktion $Y \to E(X-Y)^2$ auf der Menge der $\mathfrak{F}$-meßbaren Zufallsvariablen mit endlichem zweiten Moment (also: $E[\cdot|\mathfrak{F}]$ ist Orthogonalprojektion in $L^2(\Omega,\mathfrak{A},P)$ auf $L^2(\Omega,\mathfrak{F},P)$.)
Beweis: Für alle $\mathfrak{F}$-meßbaren Y mit $EY^2 < \infty$ gilt:
$E(X-Y)^2 = E((X - E[X|\mathfrak{F}]) + (E[X|\mathfrak{F}] - Y))^2 = E(X - E[X|\mathfrak{F}])^2 + E(E[X|\mathfrak{F}] - Y)^2$, denn
$E(X - E[X|\mathfrak{F}])(E[X|\mathfrak{F}] - Y) = E(E[(X - E[X|\mathfrak{F}])(E[X|\mathfrak{F}] - Y)|\mathfrak{F}]) = E((E[X|\mathfrak{F}] - Y)\underbrace{E[X - E[X|\mathfrak{F}]|\mathfrak{F}]}_{0}) = 0$
Der erste Term der Zerlegung hängt nicht von Y ab, der zweite ist ≥ 0 und $= 0$ für $Y = E[X|\mathfrak{F}]$. □

Stochastische Interpretation: $\mathfrak{F}$ steht für die "zugängliche Information", X soll 'vorhergesagt' werden. Wählt man die mittlere quadratische Abweichung als "Verlustfunktion", so besagt Satz 15.6, daß $E[X|\mathfrak{F}]$ optimal ist.

Beispiel 15.7

(i) $(\Omega,\mathfrak{A},P) = ([0,1), \mathfrak{B}_{[0,1)}, l|_{[0,1)})$, $\mathfrak{F} = \sigma(\{[\frac{k}{n}, \frac{k+1}{n}) : k = 0,1,\dots,n-1\})$ (n fest).
Jede $\mathfrak{F}$-meßbare Funktion $\varphi : \Omega \to \mathbb{R}$ muß auf den erzeugenden Intervallen $[\frac{k}{n}, \frac{k+1}{n})$, $k = 0,1,\dots,n-1$, konstant sein. Ist y_{nk} der Wert von $Y = E[X|\mathfrak{F}]$ auf $[\frac{k}{n}, \frac{k+1}{n})$, so muß $\int_{[\frac{k}{n},\frac{k+1}{n})} YdP = \frac{1}{n}y_{nk} = \int_{[\frac{k}{n},\frac{k+1}{n})} XdP$ gelten.

Also: $Y = n\sum_{k=0}^{n-1} (\int_{[\frac{k}{n},\frac{k+1}{n})} XdP) 1_{[\frac{k}{n},\frac{k+1}{n})}$



Es gilt $\mathfrak{F} = \sigma(Z)$, mit $Z(\omega) = \frac{\lfloor n\omega \rfloor}{n}$.
Z zeigt an, in welches Intervall das Ergebnis des Zufallsexperiments fällt.
$E[X|\mathfrak{F}] = \varphi(Z)$ mit $\varphi(\frac{k}{n}) = n \int_{[\frac{k}{n},\frac{k+1}{n})} XdP$.

(ii) Es sei $X_1, X_2, \dots$ eine iid-Folge von Zufallsvariablen mit $E|X_n| < \infty$ und $S_n := \sum_{k=1}^n X_k$. $E[X_1|S_n] =?$
(Vermutung: $\frac{1}{n}S_n$)
Für alle $B \in \mathfrak{B}$ gilt: $\int X_1 1_{\{S_n \in B\}}dP = \int\limits_{\{x\in\mathbb{R}^n : \Sigma x_i \in B\}} \dots \int x_1 P^{X_1}(dx_1)\dots P^{X_n}(dx_n) = \int\limits_{\{x\in\mathbb{R}^n : \Sigma x_i \in B\}} \dots \int x_j P^{X_1}(dx_1)\dots P^{X_n}(dx_n) =$
$\int X_j 1_{\{S_n \in B\}}dP$, denn $\bigotimes_{i=1}^n P^{X_i}$ ist invariant unter Koordinatenpermutation.
Hat also Y die Eigenschaft $\int_{\{S_n\in B\}} YdP = \int_{\{S_n\in B\}} X_1 dP$, so gilt sogar $\int_{\{S_n\in B\}} YdP = \int_{\{S_n\in B\}} X_j dP$ für $j = 1,\dots,n$, also:
$E[X_1|S_n] = E[X_2|S_n] = \dots = E[X_n|S_n]$.
Bilde Summen: $(nE[X_1|S_n] =) \sum_{i=1}^n E[X_i|S_n] = E[\sum_{i=1}^n X_i|S_n] = E[S_n|S_n] = S_n$.
Damit folgt: $\boxed{E[X_1|S_n] = \frac{1}{n}S_n}$
Betrachtet man $\mathfrak{F} = \sigma\{S_m : m \geq n\}$, so folgt mit Satz 15.5(iii): $E[X_1|S_n, S_{n+1},\dots] = \frac{1}{n}S_n$.

15.2 Martingale

Es seien $(\Omega,\mathfrak{A},P)$ ein WRaum und $(T,\leq)$ eine gerichtete Menge, d.h.
(i) "$\leq$" ist Halbordnung auf T.
(ii) $\forall s,t \in T\ \exists u \in T : s \leq u, t \leq u$.
Weiter sei $(\mathfrak{F}_t)_{t\in T}$ eine bzgl. "$\leq$" aufsteigende Familie von Unter-σ-Algebren von $\mathfrak{A}$, d.h. $s \leq t \implies \mathfrak{F}_s \leq \mathfrak{F}_t$ (Filtration)

Definition 15.8

Der stochastische Prozeß $(X_t)_{t\in T}$ (d.h. $\forall t \in T$ ist $X_t : \Omega \to \mathbb{R}$ eine ZV) heißt zu $(\mathfrak{F}_t)_{t\in T}$ adaptiert, wennX_t $\mathfrak{F}_t$-meßbar ist für alle t.
$(X_t,\mathfrak{F}_t)_{t\in T}$ heißt Martingal, wenn $E|X_t| < \infty$ $\forall t \in T$ und $s,t \in T, s \leq t \implies E[X_t|\mathfrak{F}_s] = X_s$ P-f.s..
Im Falle $X_s \leq E[X_t|\mathfrak{F}_s]$ für alle $s \leq t$ nennt man $(X_t,\mathfrak{F}_t)_{t\in T}$ ein Submartingal, bei $\geq$ Supermartingal.

Interpretation: T Zeit, $(X_t)_{t\in T}$ beschreibt die zufällige Entwicklung des "stochastischen Systems": X_t ist Position eines Partikels zur Zeit t, Aktienkurs o.ä.

Wenn $(X_t)_{t\in T}$ gegeben ist, so läßt sich eine zugehörige Filtration $(\mathfrak{F}_t)_{t\in T}$ konstruieren durch $\mathfrak{F}_t := \sigma(\{X_s : s \leq t\})$. Diese Familie nennt man die natürliche Filtration zu $(X_t)_{t\in T}$.
Klar: $(X_t)_{t\in T}$ ist zur zugehörigen Filtration adaptiert (X_t ist $\mathfrak{F}_t$-meßbar.)

Grob: Bei einem Martingal ist der gegenwärtige Wert der beste Vorhersagewert für zukünftige Werte (Satz 15.6). Bei einem Martingal bleiben wegen $EX_t = E(E[X_t|\mathfrak{F}_s]) = EX_s$ die Erwartungswerte konstant, bei einem Supermartingal nehmen sie im Laufe der Zeit ab, bei Submartingalen zu.

Ist $(X_t, \mathfrak{F}_t)_{t\in T}$ ein Supermartingal, so gilt:
(∗) $\int\limits_A X_s dP \geq \int\limits_A X_t dP$ für alle $s, t \in T$ mit $s \leq t$ und alle $A \in \mathfrak{F}_s$.
Ist umgekehrt $(X_t)_{t\in T}$ zu $(\mathfrak{F}_t)_{t\in T}$ adaptiert, und gilt $E|X_t| < \infty \quad \forall t \in T$, so impliziert (∗), daß $(X_t, \mathfrak{F}_t)_{t\in T}$ ein Supermatingal ist ($\rightsquigarrow$ Aufgabe 54.)

Die analoge Aussage für Submartingale und Martingale erhält man mit:
$(X_t, \mathfrak{F}_t)_{t\in T}$ Submartingal $\iff$ $(-X_t, \mathfrak{F}_t)_{t\in T}$ ist Supermatingal.
$(X_t, \mathfrak{F}_t)_{t\in T}$ Martingal $\iff$ $(X_t, \mathfrak{F}_t)_{t\in T}$ ist Super- und Submartingal.

Man sagt (abkürzend): "$(X_t)_{t\in T}$ Martingal", wenn $(X_t, \mathfrak{F}_t)_{t\in T}$ mit der natürlichen Filtration $(\mathfrak{F}_t)_{t\in T}$ ein Martingal ist.

Beispiel 15.9

(i) Es sei $T = \mathbb{N}$, $(X_n)_{n\in\mathbb{N}}$ eine iid-Folge von Zufallsvariablen mit Erwartungswert μ. $(S_n)_{n\in\mathbb{N}}$ sei die Folge der Partialsummen $S_n := \sum\limits_{k=1}^n X_k$. Weiter sei $\mathfrak{F}_n := \sigma(\{X_1, \dots, X_n\})$ $(= \sigma(\{S_1, \dots, S_n\}))$ (also: $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ ist die natürliche Filtration zu $(S_n)_{n\in\mathbb{N}}$).

$E[S_{n+1}|\mathfrak{F}_n] = E[S_n + X_{n+1}|\mathfrak{F}_n] = E[S_n|\mathfrak{F}_n] + E[X_{n+1}|\mathfrak{F}_n] = S_n + \overbrace{\mu}^{E(X_n)}$ (nach Satz 15.3(i), Satz 15.5(iii)) für alle $n \in \mathbb{N}$, also $E[S_{n+1}|\mathfrak{F}_n] \overset{\geq}{\underset{\leq}{=}} S_n$, wenn $\mu \overset{\geq}{\underset{\leq}{=}} 0$.

$E[S_{n+2}|\mathfrak{F}_n] = E[E[S_{n+2}|\mathfrak{F}_{n+1}]|\mathfrak{F}_n]$ ("Turm")

Hieraus folgt, daß $(S_n)_{n\in\mathbb{N}}$ ein Submartingal / Martingal / Supermartingal ist, wenn $\mu \geq 0$ / $\mu = 0$ / $\mu \leq 0$ gilt ($\rightsquigarrow$ Satz 15.5(i))

(ii) (In diesem Beispiel ist T keine Teilmenge von $\mathbb{R}$ …)
Es sei Q ein weiteres WMaß auf $(\Omega, \mathfrak{A})$, daß von P dominiert wird (d.h. $P(N) = 0 \implies Q(N) = 0$)
Sei T die Menge der endlichen meßbaren Zerlegungen (Partitionen) $\mathfrak{Z}$ von Ω, d.h. $\mathfrak{Z} = (A_1, \dots, A_n)$ mit $n \in \mathbb{N}$, $A_1, \dots, A_n \in \mathfrak{A}$, $A_i \cap A_j = \emptyset$ für $i \neq j$, $\bigcup\limits_{i=1}^n A_i = \Omega$. Sind $\mathfrak{Z}_1 = (A_1, \dots, A_n)$ und $\mathfrak{Z}_2 = (B_1, \dots, B_m)$ zwei solche Zerlegungen, so nennen wir $\mathfrak{Z}_2$ feiner als $\mathfrak{Z}_1$ und schreiben $\mathfrak{Z}_1 \leq \mathfrak{Z}_2$, wenn gilt: $\forall k \in \{1, \dots, n\} \exists I \subseteq \{1, \dots, m\}$: $A_k = \bigcup\limits_{i\in I} B_i$.
Zu zwei Partitionen gibt es stets eine gemeinsame feinere (beispielsweise das System aller Durchschnitte), also $(T, \leq)$ ist eine gerichtete Menge.
Für jedes $t = \mathfrak{Z} = (A_1, \dots, A_n) \in T$ setzen wir nun $\mathfrak{F}_t := \sigma(\{A_1, \dots, A_n\})$, $X_t = \sum\limits_{i=1}^n \frac{Q(A_i)}{P(A_i)} 1_{A_i}$
(=0, wenn $P(A_i) = 0$).
Offensichtlich ist dann $(X_t)_{t\in T}$ zur Filtration $(\mathfrak{F}_t)_{t\in T}$ adaptiert, und es gilt

$$E|X_t| = EX_t = \sum_{i=1}^n \frac{Q(A_i)}{P(A_i)} E1_{A_i} = \sum_{i=1}^n Q(A_i) = 1.$$

Es seien nun $s = (A_1, \dots, A_n)$ und $t = (B_1, \dots, B_m)$ zwei Zerlegungen mit $s \leq t$. Für $k = 1, \dots, n$ gilt:

$$\int_{A_k} X_t dP = \sum_{i=1}^m \frac{Q(B_i)}{P(B_i)} \underbrace{\int_{A_k} 1_{B_i} dP}_{P(A_k \cap B_i)} = \sum_{\{i: B_i \subseteq A_k\}} \frac{Q(B_i)}{P(B_i)} P(B_i) = Q(A_k).$$

Außerdem: $\int\limits_{A_k} X_s dP = \sum\limits_{i=1}^n \frac{Q(A_i)}{P(A_i)} \underbrace{\int_{A_k} 1_{A_i} dP}_{P(A_i \cap A_k)} = Q(A_k)$.

Mit Aussage (∗) folgt, daß $(X_t, \mathfrak{F}_t)_{t\in T}$ ein Martingal ist.

(iii) Es sei $(\mathfrak{F}_t)_{t\in T}$ eine Filtration, X eine Zufallsvariable mit endlichem Erwartungswert und $X_t := E[X|\mathfrak{F}_t]$.
Nach Konstruktion ist dann $(X_t)_{t\in T}$ zu $(\mathfrak{F}_t)_{t\in T}$ adaptiert und mit Satz 15.5(i) folgt für $s \leq t$ $E[X_t|\mathfrak{F}_s] = E[E[X|\mathfrak{F}_t]|\mathfrak{F}_s] = E[X|\mathfrak{F}_s] = X_s$ (wieder P-f.s.), also ist $(X_t, \mathfrak{F}_t)_{t\in T}$ eine Martingal.

Satz 15.10 (Submartingalungleichung von Doob)
Ist $(X_t, \mathfrak{F}_t)_{t=1,\dots N}$ ein Submartingal, so gilt für alle $c > 0$:

$$c \cdot P(\max_{1\leq k\leq n} X_k \geq c) \leq \int_{\{\max\limits_{1\leq k\leq n} X_k \geq c\}} X_n dP \leq EX_n^+$$

Beweis: $A := \{\max\limits_{1\le k\le n} X_k \ge c\}$, $A_i := \{X_j < c$ für $j = 1, \dots, i-1,\ X_i \ge c\}$.

Dann gilt: $A = A_1 + A_2 + \ldots + A_n$; $A_i \in \mathfrak{F}_i$, $X_i \ge c$ auf A_i.

Dann $\int\limits_{A_i} X_n dP \ge \int\limits_{A_i} X_i dP \ge c \cdot P(A_i)$.

Summiere nun über i : $\int\limits_{A} X_n dP \ge c \cdot P(A)$, der zweite Teil der Ungleichung folgt mit $X_n 1_A \le X_n^+$. □

Mit den Resultaten des ersten Paragraphen erhält man leicht eine Reihe von Eigenschaften der Klasse der Martingale (Supermartingale bzw. Submartingale). Aus der Linearität des bedingten Erwartungswertes folgt beispielsweise unmittelbar, daß mit $(X_t, \mathfrak{F}_t)_{t\in T}$ und $(Y_t, \mathfrak{F}_t)_{t\in T}$ auch $(\alpha X_t + \beta Y_t, \mathfrak{F}_t)_{t\in T}$ ein Supermartingal ist, wenn $\alpha \ge 0$, $\beta \ge 0$; die analoge Aussage für Martingale gilt für alle $\alpha, \beta \in \mathbb{R}$. Ist T_0 eine gerichtete Teilmenge von T, so ist mit $(X_t, \mathfrak{F}_t)_{t\in T}$ auch $(X_t, \mathfrak{F}_t)_{t\in T_0}$ ein Submartingal. Die bedingte Version der Jensenschen Ungleichung liefert unmittelbar:

Satz 15.11

Ist $(X_t, \mathfrak{F}_t)_{t\in T}$ ein Martingal und $\varphi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion mit $E|\varphi(X_t)| < \infty$ für alle $t \in T$, so ist $(\varphi(X_t), \mathfrak{F}_t)_{t\in T}$ ein Submartingal.

Insbesondere ist bei einem quadratisch integrierbaren Martingal $(X_t, \mathfrak{F}_t)_{t\in T}$, d.h. $E({X_t}^2) < \infty \quad \forall t \in T$, der Prozeß $({X_t}^2, \mathfrak{F}_t)_{t\in T}$ ein Submartingal. Weitere Eigenschaften dieser Art werden in den Übungsaufgaben behandelt.

15.3 Stoppzeiten und Transformationen

Wir betrachten in diesem Absatz Prozesse mit Zeitmenge $T = \mathbb{N}_0 = \{0, 1, 2, \dots\}$. Es ist $(\Omega, \mathfrak{A}, P)$ ein WRaum und $(\mathfrak{F}_t)_{t\in\mathbb{N}_0}$ eine Filtration. Martingale etc. sollen sich stets auf $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$ beziehen. Man sieht leicht, daß bei diesem T die (Super-, Sub-) Martingaleigenschaft bereits aus $E[X_{n+1}|\mathfrak{F}_n] \overset{\le}{\underset{\ge}{=}} X_n$ für alle $n \in \mathbb{N}_0$ folgt (Induktion, Satz 15.5(i)).

Im nachfolgenden ist gelegentlich eine Interpretation über Glücksspiele hilfreich. Hierbei ist X_n das Kapital eines Spielers nach n Durchgängen, wenn jede Runde eine Geldeinheit gesetzt wird (X_0 ist das Anfangskapital), $\mathfrak{F}_n := \sigma(\{X_0, \dots, X_n\})$ repräsentiert die nach n Durchgängen vorhandene Information. Weiter ist $X_{n+1} - X_n$ der Gewinn pro gesetzter Geldeinheit in der nächsten Runde, $E[X_{n+1} - X_n|\mathfrak{F}_n]$ ist der zu erwartende Gewinn in der nächsten Runde bei Kenntnis des gesamten bisherigen Spielverlaufs. Also wegen $E[X_{n+1} - X_n|\mathfrak{F}_n] = E[X_{n+1}|\mathfrak{F}_n] - X_n$ gilt:

$(X_n)_{n\in\mathbb{N}_0}$ ist $\left\{\begin{array}{l}\text{Supermartingal}\\ \text{Martingal}\\ \text{Submartingal}\end{array}\right\}$ bedeutet, das Spiel ist $\left\{\begin{array}{l}\text{nachteilig}\\ \text{fair}\\ \text{vorteilhaft}\end{array}\right\}$ (aus Sicht des Spielers).

Betrachten wir nun die strategischen Möglichkeiten: In der n-ten Runde werden c_n Geldeinheiten eingesetzt, wobei natürlich c_n $\mathfrak{F}_{n-1}$-meßbar sein muß. Das Kapital des Spielers nach n Durchgängen ist dann $X_0 + \sum\limits_{k=1}^{n} c_k(X_k - X_{k-1})$.

Definition 15.12 Ein reellwertiger stochastischer Prozeß $(c_n)_{n\in\mathbb{N}}$ heißt vorhersagbar bzgl. $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$, wenn c_n $\mathfrak{F}_{n-1}$-meßbar ist für alle $n \in \mathbb{N}$.

Satz 15.13 Es sei $(c_n)_{n\in\mathbb{N}_0}$ ein vorhersagbarer Prozeß, $(X_n)_{n\in\mathbb{N}_0}$ ein Prozeß mit $E|c_{n+1}(X_{n+1} - X_n)| < \infty$ für alle $n \in \mathbb{N}_0$, $Y_n := X_0 + \sum\limits_{k=1}^{n} c_k(X_k - X_{k-1})$. Dann gilt:

(i) Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal, so ist $(Y_n)_{n\in\mathbb{N}_0}$ ein Martingal.

(ii) Ist $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und ist $(c_n)_{n\in\mathbb{N}_0}$ nicht-negativ, so ist auch $(Y_n)_{n\in\mathbb{N}_0}$ ein Supermatingal.

Beweis: Da c_{n+1} $\mathfrak{F}_n$-meßbar ist, folgt mit Satz 15.5(ii) $E[Y_{n+1} - Y_n|\mathfrak{F}_n] = c_{n+1}E[X_{n+1} - X_n|\mathfrak{F}_n]$, woraus sich beide Behauptungen ergeben. □

Aus einem (un-)fairen Spiel wird also durch variieren der Einsätze kein vorteilhaftes Spiel (wenn man nicht über prophetische Gaben verfügt ...)
Kann man sein Glück dadurch verbessern, daß man im geeigneten Moment aufhört?

Definition 15.14

Eine Abbildung $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ heißt Stoppzeit (auch Optionszeit) bzgl. der Filtration $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$, wenn $\{\tau \le n\} \in \mathfrak{F}_n$ für alle $n \in \mathbb{N}_0$ gilt. Für eine Stoppzeit τ ist $\mathfrak{F}_\tau := \{A \subseteq \Omega : A \cap \{\tau \le n\} \in \mathfrak{F}_n \quad \forall n \in \mathbb{N}_0\}$ eine σ-Algebra (Übungsaufgabe), die σ-Algebra der τ-Vergangenheit.
(Interpretation: Die bis einschließlich zur (zufälligen) Zeit τ vorhandene Information.)

Als "optional stopping" bezeichnet man den Übergang von $(X_n)_{n\in\mathbb{N}_0}$ zu $(X_{\tau\wedge n})_{n\in\mathbb{N}_0}$[2], wobei $X_{\tau\wedge n} : \Omega \to \mathbb{R}$, $(X_{\tau\wedge n})(\omega) := X_{\tau(\omega)\wedge n}(\omega)$.
Ist $(\tau_n)_{n\in T_0}$, $T_0 \subseteq \mathbb{N}_0$, eine wachsende Folge von Stoppzeiten, so bezeichnet man den Übergang von $(X_n)_{n\in\mathbb{N}_0}$ zu $(Y_n)_{n\in T_0}$, $Y_n := X_{\tau_n}$ als "optional sampling". Mit τ ist auch $\tau_n := \tau \wedge n$ eine Stoppzeit ($\rightsquigarrow$ Übungsaufgabe), also ist optional stopping ein Spezialfall des optional sampling.

[2] $\wedge$ ist hierbei das Minimum ...

Bei der Glücksspiel-Situation könnte eine mögliche Strategie darin bestehen, bis zum Erreichen eines Mindestgewinnes c zu spielen und dann aufzuhören; dies entspricht dem optional stopping mit $\tau = \inf\{n \in \mathbb{N}_0 : X_n \geq X_0 + c\}$ (wir vereinbaren: $\inf \emptyset = \infty$).

Beispiel 15.15

(i) Konstante τ's sind Stoppzeiten, denn im Falle $\tau(\omega) = n_0 \quad \forall \omega \in \Omega$ gilt

$\{\tau \leq n\} = \Omega$ für $n \geq n_0$

$\{\tau \leq n\} = \emptyset$ für $n < n_0$, man erhält also jedesmal ein Element von $\mathfrak{F}_n$.

(ii) Ist $(Y_n)_{n\in\mathbb{N}_0}$ ein zu $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$ adaptierter (reellwertiger) Prozeß und A eine Borel-Menge, so ist die Eintrittszeit dieses Prozesses in A $\quad \tau_A(\omega) = \inf\{n \in \mathbb{N}_0 : Y_n(\omega) \in A\}$ (wieder mit $\inf \emptyset = \infty$) eine Stoppzeit, denn $\{\tau_A \leq n\} = \bigcup\limits_{i=0}^{n} \{Y_i \in A\} \in \mathfrak{F}_n$, wegen $\{Y_i \in A\} \in \mathfrak{F}_i$, $(\mathfrak{F}_n)_{n\in\mathbb{N}_0}$ isoton.

Die oben genannte 'Aufhörstrategie' basiert also tatsächlich auf einer Stoppzeit.

Wir wollen nun zeigen, daß unter bestimmten Voraussetzungen auch unter einer optional-sampling-Transformation die (Super-) Martingaleigenschaft erhalten bleibt. Hierzu benötigen wir einige einfache Hilfsaussagen:

Lemma 15.16

(i) Ist τ eine Stoppzeit, so ist $X_\tau^* : \Omega \to \mathbb{R}$, $X_\tau^*(\omega) := \begin{cases} X_{\tau(\omega)}(\omega) & , \quad \text{wenn } \tau(\omega) \leq \infty \\ 0 & , \quad \text{sonst} \end{cases}$ $\mathfrak{F}_\tau$-meßbar.

(ii) Sind τ_1, τ_2 Stoppzeiten mit $\tau_1 \leq \tau_2$, so gilt $\mathfrak{F}_{\tau_1} \subseteq \mathfrak{F}_{\tau_2}$.

Beweis: (i) Für jede Borelmenge $A \in \mathfrak{B}$ und jedes $n \in \mathbb{N}_0$ gilt: $\{X_\tau^* \in A\} \cap \{\tau \leq n\} = \bigcup\limits_{k=0}^{n} \{X_k \in A\} \cap \{\tau = k\}$.

Mit $\{\tau = k\} = \underbrace{\{\tau \leq k\}}_{\in \mathfrak{F}_k} \cap \underbrace{\{\tau \leq k-1\}^c}_{\in \mathfrak{F}_{k-1} \subseteq \mathfrak{F}_k}$ folgt $\{\tau = k\} \in \mathfrak{F}_k$, also insgesamt $\{X_\tau^* \in A\} \cap \{\tau \leq n\} \in \mathfrak{F}_n$.

(ii) Sei $A \in \mathfrak{F}_{\tau_1}$. Für alle $n \in \mathbb{N}_0$ gilt: $\{\tau_2 \leq n\} \subseteq \{\tau_1 \leq n\}$, also folgt

$A \cap \{\tau_2 \leq n\} = \underbrace{A \cap \{\tau_1 \leq n\}}_{\in \mathfrak{F}_n,\ \text{da } A \in \mathfrak{F}_{\tau_1}} \cap \underbrace{\{\tau_2 \leq n\}}_{\in \mathfrak{F}_n,\ \text{da } \tau_2 \text{ Stoppzeit}} \in \mathfrak{F}_n$, d.h. $A \in \mathfrak{F}_{\tau_2}$. □

Ohne weitere Voraussetzungen kann beim Übergang von (X_n) zu (Y_n) die Martingaleigenschaft verloren gehen.

Beispiel 15.17

Es sei $(Y_n)_{n\in\mathbb{N}}$ eine iid-Folge von Zufallsvariablen mit $P(Y_n = -1) = P(Y_n = 1) = \frac{1}{2}$, $X_n := \sum\limits_{i=1}^{n} 2^{i-1} Y_i$,

$\mathfrak{F}_n := \sigma(\{X_1, \dots, X_n\}) \quad (= \sigma(\{Y_1, \dots, Y_n\}))$, $\tau := \inf\{n \in \mathbb{N} : Y_n = 1\}$.

Dann ist $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein Martingal und τ eine Stoppzeit ($\rightsquigarrow$ Beispiel 15.15(ii))

Interpretation: Eine faire Münze wird wiederholt geworfen. Der Spieler setzt 2^{n-1}DM in der n-ten Runde und verliert seinen Einsatz, wenn 'Zahl' erscheint ($Y_n = -1$), bei 'Kopf' ($Y_n = 1$) gewinnt er 2^nDM. Beim ersten Eintreten von 'Kopf' wird das Spiel beendet; X_τ ist der Gesamtgewinn bei Spielende.

Es gilt: $P(\tau > k) = 2^{-k}$, insbesondere $P(\tau < \infty) = 1$, sowie

$$X_\tau = \sum_{k=1}^{\infty} X_k 1_{\{\tau = k\}} = \sum_{k=1}^{\infty} \underbrace{\left(-\sum_{i=1}^{k-1} 2^{i-1} + 2^{k-1} \right)}_{\equiv 1} 1_{\{\tau = k\}} \equiv 1 \quad (P\text{-f.s.})$$

Sei nun $T_0 = \{1, 2\}$, $\tilde{Y}_1 = X_1$, $\tilde{Y}_2 = X_\tau$.

Dann gilt: $E\tilde{Y}_1 = 0$, $E\tilde{Y}_2 = EX_\tau = 1$.

Also ist $(\tilde{Y}_n, \mathfrak{G}_n)_{n\in T_0}$ kein Martingal ($E|X_{\tau-1}| = \infty$.)

Lemma 15.18 Es seien $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}_0}$ ein Supermartingal und σ, τ Stoppzeiten mit $\sigma \leq \tau \leq \sigma + 1 \leq C$ für ein konstantes $C < \infty$. Dann gilt $E|X_\tau| < \infty$, sowie $E[X_\tau | X_\sigma] \leq X_\sigma$.

Beweis: Es gilt $E|X_\tau| = \sum\limits_{k=0}^{C} \int\limits_{\{\tau = k\}} |X_k| dP \leq \sum\limits_{k=0}^{C} E|X_k| < \infty$.

Sei nun $A \in \mathfrak{F}_\sigma$, $A_k := A \cap \{\sigma = k\} \cap \{\tau > \sigma\} = A \cap \{\sigma = k\} \cap \{\tau > k\} = A \cap \{\sigma = k\} \cap \{\tau = k+1\}$.

Dann gilt: $\int\limits_A (X_\sigma - X_\tau) dP = \sum\limits_{k=0}^{C-1} \int\limits_{A_k} (X_\sigma - X_\tau) dP = \sum\limits_{k=0}^{C-1} \int\limits_{A_k} (X_k - X_{k+1}) dP \geq 0$, wenn $A_k \in \mathfrak{F}_k$.

$A_k \in \mathfrak{F}_k$ folgt mit $A \cap \{\sigma = k\} = \underbrace{(A \cap \{\sigma \leq k\})}_{\in \mathfrak{F}_k,\ \text{da } \sigma \text{ Stoppzeit}} \cap \Big(\bigcup\limits_{i=1}^{k-1} \underbrace{(A \cap \{\sigma \leq i\})}_{\in \mathfrak{F}_i \subseteq \mathfrak{F}_k}\Big)^c$. □

Satz 15.19 (Optional Sampling Theorem)

Es seien $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}_0}$ ein Supermartingal (Submartingal, Martingal) und $(\tau_n)_{n\in\mathbb{N}_0}$ eine isotone Folge von beschränkten Stoppzeiten (d.h. $\forall n \in \mathbb{N}\ \exists C_n < \infty : P(\tau_n \leq C_n) = 1$); $Y_n := X_{\tau_n}$; $\mathfrak{G}_n := \mathfrak{F}_{\tau_n}$. Dann ist $(Y_n, \mathfrak{G}_n)_{n\in\mathbb{N}_0}$ wieder ein Supermartingal (Submartingal, Martingal).

Beweis: Es reicht, den Satz für Supermatingale zu beweisen.

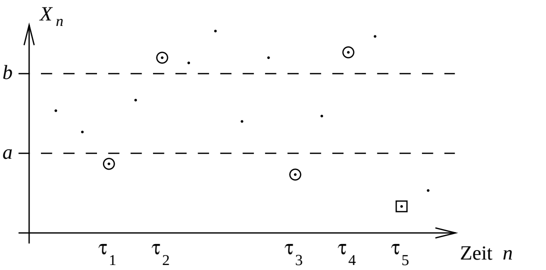
Nach Lemma 15.16(ii) liefert $\mathfrak{G}_n \subseteq \mathfrak{G}_{n+1}\ \forall n \in \mathbb{N}_0$; $(\mathfrak{G}_n)_{n\in\mathbb{N}_0}$ ist also eine Filtration.

Nach Lemma 15.16(i) ist $(Y_n)_{n\in\mathbb{N}_0}$ hierzu adaptiert. Es bleibt zu zeigen: $E[X_{\tau_{n+1}}|\mathfrak{F}_{\tau_n}] \le X_{\tau_n}$ P-f.s. für alle $n \in \mathbb{N}$.
Sei $n \in \mathbb{N}$. Setze $\sigma_k := \tau_{n+1} \wedge (\tau_n + k)$ für $k = 0, \dots, \lceil C_{n+1} \rceil =: m$.
Dann ist $\sigma_0, \dots, \sigma_m$ eine isotone Folge von Stoppzeiten mit $\sigma_0 = \tau_n$, $\sigma_m = \tau_{n+1}$ und $\sigma_{k+1} - \sigma_k \le 1$ für $k = 0, \dots, m-1$.
Lemma 15.18 liefert $E[X_{\sigma_{k+1}}|\mathfrak{F}_{\sigma_k}] \le X_{\sigma_k}$, also:
$E[X_{\tau_{n+1}}|\mathfrak{F}_{\tau_n}] = E[X_{\sigma_m}|\mathfrak{F}_{\sigma_0}] = E[E[X_{\sigma_m}|\mathfrak{F}_{\sigma_{m-1}}]|\mathfrak{F}_{\sigma_0}] \le E[X_{\sigma_{m-1}}|\mathfrak{F}_{\sigma_0}] \le \dots \le E[X_{\sigma_0}|\mathfrak{F}_{\sigma_0}] = X_{\sigma_0} = X_{\tau_n}$ P-f.s. □

15.4 Konvergenzsätze

15.4.1 Fast sichere Konvergenz

Zu einer Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen definieren wie die Größen $U_n[a,b]$, $-\infty < a < b < \infty$, $n \in \mathbb{N}$, wie folgt: $U_n[a,b]$ ist die Anzahl der aufsteigenden Überquerungen des Intervalls $[a,b]$ durch $X_1, X_2, \dots, X_n$, also das größte $k \in \mathbb{N}_0$ mit $\exists 1 \le i_1 < i_2 < i_3 < \dots < i_{2k} \le n$ mit $X_{i_{2j-1}} \le a$, $X_{i_{2j}} \ge b$ für $j = 1, \dots k$. ($U_n[a,b]$ ist eine Zufallsvariable.)



Lemma 15.20 Ist $(X_n)_{n\in\mathbb{N}}$ ein Supermartingal, so gilt: $EU_n[a,b] \le \frac{1}{b-a} E(X_n - a)^-$.

Beweis: Sei $p = \lfloor \frac{n}{2} \rfloor + 1$, $\tau_0 \equiv 0$, und für $k = 1, \dots, p$,

$$\tau_{2k-1} := \begin{cases} \min\{j \ge \tau_{2k-2}, j \le n : X_j \le a\} & , \text{ wenn } \{\dots\} \ne \emptyset \\ n & , \text{ sonst} \end{cases}$$
$$\tau_{2k} := \begin{cases} \min\{j \ge \tau_{2k-1}, j \le n : X_j \ge b\} & , \text{ wenn } \{\dots\} \ne \emptyset \\ n & , \text{ sonst} \end{cases}$$

Dies sind Stoppzeiten, klar: $1 \le \tau_1 \le \tau_2 \le \dots \le \tau_p = n$.
Die aufsteigenden Überquerungen gehören zu den Paaren $(\tau_1, \tau_2), \dots, (\tau_{2k_0-1}, \tau_{2k_0})$ mit $k_0 = U_n[a,b]$, d.h. $X_{\tau_{2k}} - X_{\tau_{2k-1}} \ge b - a$ für $k = 1, \dots, k_0$.
Die hierauf folgende Differenz $X_{\tau_{2k_0+2}} - X_{\tau_{2k_0+1}}$ ist nur dann nicht 0, wenn noch eine a-Unterschreitung (aber keine b-Überschreitung) mehr stattfindet. Sie ist dann von der Form $X_n - X_j$ mit einem $X_j \le a$, ist also $\ge X_n - a \ge \min\{0, X_n - a\} = -(X_n - a)^-$.
Es gilt $\quad (*) \quad \sum_{k=1}^{p} (X_{\tau_{2k}} - X_{\tau_{2k-1}}) \ge (b-a)U_n[a-b] - (X_n - a)^-$.
Das Optional Sampling Theorem (Satz 15.19) liefert $EX_{\tau_{2k}} \le EX_{\tau_{2k-1}}$, also ist der Erwartungswert auf der linken Seite von $(*)$ ≤ 0, d.h. $(b-a)U_n[a,b] - E(X_n - a)^- \le 0$.
(siehe Williams für einen auf Satz 15.13 beruhenden Beweis.) □

Ist $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ eine Filtration, so sei $\mathfrak{F}_\infty := \sigma(\bigcup_{n\in\mathbb{N}} \mathfrak{F}_n)$.

Satz 15.21 (Vorwärts-Konvergenzsatz von Doob)
Es sei $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein Supermartingal mit der Eigenschaft $\sup_{n\in\mathbb{N}} E|X_n| < \infty$
(Man nennt das Supermartingal dann $\mathfrak{L}^1$-beschränkt).
Dann existiert eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit $\lim_{n\to\infty} X_n = X_\infty$ P-fast sicher.

Beweis: Wie zeigen zunächst $P(N) = 0$ für $N := \{\omega \in \Omega : \liminf_{n\to\infty} X_n(\omega) < \limsup_{n\to\infty} X_n(\omega)\}$.

Mit $U_\infty[a,b] := \lim_{n\to\infty} U_n[a,b]$ gilt $N \subset \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \{\omega \in \Omega : U_\infty[a,b](\omega) = \infty\}$.

Lemma 15.20 liefert $(b-a)EU_n[a,b] \le E(X_n - a)^- \le |a| + E|X_n|$.

Nach Voraussetzung gibt es hierfür eine von n unabhängige Oberschranke, also folgt $EU_\infty[a,b] < \infty$ mit dem Satz von der der majorisierten Konvergenz – insbesondere $P(U_\infty[a,b] = \infty) = 0$.

Damit: N ist in einer abzählbaren Vereiniging von Nullmengen enthalten, also $P(N) = 0$.

Sei $\tilde{X}_\infty(\omega) := \begin{cases} \lim_{n\to\infty} X_n(\omega) & , \ \omega \in N^c \\ 0 & , \ \text{sonst} \end{cases}$ (bestimmte Divergenz ist zugelassen.)

Das Lemma von Fatou liefert $E|\tilde{X}_\infty| = E \liminf |X_n| \le \liminf E|X_n| \le \sup_n E|X_n| < \infty$, d.h. $P(\tilde{N}) = 0$ mit $\tilde{N} = \{\omega \in \Omega : |\tilde{X}_\infty| = \infty\}$.

$X_\infty := 1_{\tilde{N}^c} \cdot \tilde{X}_\infty$ leistet das Verlangte. □

Bemerkung 15.22 $\mathfrak{L}^1$-Beschränktheit folgt bereits aus $\sup_n EX_n^- < \infty$ (Übungsaufgabe 64)

15.4.2 Gleichgradige Integrierbarkeit

Hintergrund: Wann hat man (bei Martingalen) Konvergenz im Mittel (in $\mathfrak{L}^1$).

Definition 15.23 Eine Familie $\{X_i : i \in I\}$ von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$ heißt gleichgradig integrierbar, wenn gilt:

$\forall \epsilon > 0\ \exists C < \infty\ \forall i \in I : \int\limits_{|X_i|>C} |X_i| dP < \epsilon.$

Beispiel 15.24

(i) Ist X integrierbar (d.h. $E|X| < \infty$), so folgt mit dem Satz von der majorisierten Konvergenz $\lim\limits_{C\to\infty} \int\limits_{|X|>C} |X| dP = 0$, also ist die "Familie" $\{X\}$ gleichgradig integrierbar.
Offensichtlich ist die endliche Vereinigung gleichgradig integrierbarer Familien wieder gleichgradig integrierbar, insbesondere sind alle endlichen Familien von Zufallsvariablen mit endlichen Erwartungswerten gleichgradig integrierbar.

(ii) Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen, Y eine weitere Zufallsvariable mit $|X_n| < Y \quad \forall n,\ EY < \infty$ (Y ist integrierbare Majorante zur Folge $(X_n)_{n\in\mathbb{N}}$), so ist $\{X_n : n \in \mathbb{N}\}$ gleichgradig integrierbar:
$\int\limits_{|X_n|>C} |X_n| dP \le \int\limits_{Y>C} Y dP$ für alle $n \in \mathbb{N}$, die rechte Seite hängt nicht von n ab und geht mit $C \to \infty$ gegen 0 (vergleicht Teil (i)) (geht auch mit beliebigen Familien $\{X_i : i \in I\}$).

(iii) Ist $\{X_i : i \in I\}$ $\mathfrak{L}^p$-beschränkt für ein $p > 1$ (d.h. $\sup\limits_{i\in I} E|X_i|^p < \infty$), so folgt mit
$\int\limits_{|X_i|>K} |X_i| dP \le \int\limits_{|X_i|>K} K^{1-p}|X_i|^p dP \le K^{1-p} \sup\limits_{i\in I} E|X_i|^p$ die gleichgradig Integrierbarkeit der Familie (wegen $p > 1$ geht die rechte Seite, die nicht von i abhängt mit $K \to \infty$ gegen 0). Die analoge Aussage mit $p = 1$ gilt NICHT (Aufgabe 65).

Lemma 15.25 Es sei X ein Zufallsvariable mit $E|X| < \infty$. Dann gilt:

$\forall \epsilon > 0\ \exists \delta > 0\ \forall A \in \mathfrak{A} :\ P(A) < \delta \implies \int\limits_A |X| dP < \epsilon.$

Beweis: Gilt die Aussage nicht, so existiert ein $\epsilon_0 > 0$ und eine Folge $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$ mit

$P(A_n) \le 2^{-n},\ \int\limits_{A_n} |X| dP \ge \epsilon_0 \quad \forall n \in \mathbb{N}.$

Sei $A = \limsup\limits_{n\to\infty} A_n$. Das Borel-Cantelli-Lemma liefert $P(A) = 0$.

Andererseits folgt mit Fatou $\quad (0 =) \int\limits_A |X| dP = E|X| - \int\limits_{A^c} |X| dP = E|X| - \int \liminf(1_{A_n{}^c}|X|) dP$

$\ge E|X| - \liminf \int\limits_{A_n{}^c} |X| dP = \limsup \int\limits_{A_n} |X| dP \ge \epsilon, \quad \#.$ □

Satz 15.26

Ist X eine Zufallsvariable mit $E|X| < \infty$, so ist $\{E[X|\mathfrak{F}] : \mathfrak{F}$ Unter-σ-Algebra von $\mathfrak{A}\}$ gleichmäßig integrierbar.

Beweis: Sei $\epsilon > 0$. Nach Lemma 15.25 existiert ein $\delta > 0$ mit

$(*) \qquad P(A) < \delta \implies \int\limits_A |X| dP < \epsilon \quad \forall A \in \mathfrak{A}.$

Wähle nun, zu diesem δ ein C mit $\frac{1}{C}E|X| < \delta$. Sei nun $\mathfrak{F}$ irgendeine Unter-σ-Algebra von $\mathfrak{A}$ und Y eine Version von $E[X|\mathfrak{F}]$. Die Jensensche Ungleichung für bedingte Erwartungswerte liefert

$(**) \qquad |Y| \le E\big[|X|\big|\mathfrak{F}\big]$ P-fast sicher und damit $C \cdot P(|Y| \ge C) \le E|Y| \le E|X|$, also $P(|Y| \ge C) \le \frac{1}{C}E|Y| \overset{\text{s.o.}}{\le} \delta.$

Wegen $\{|Y| \ge C\} \in \mathfrak{F}$ folgt insgesamt $\int\limits_{|Y|\ge C} |Y| dP \overset{(**)}{\le} \int\limits_{|Y|\ge C} E\big[|X|\big|\mathfrak{F}\big] dP = \int\limits_{|Y|\ge C} |X| dP \overset{(*)}{\le} \epsilon.$ □

Das nun folgende Hauptresultat dieses Abschnitts verallgemeinert den Satz von der majorisierten Konvergenz (wegen Beispiel 15.24(ii))

Satz 15.27 Es seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit endlichem Erwartungswert. Dann sind äquivalent:

(i) $X_n \to X$ in $\mathfrak{L}^1$ (d.h. $E|X_n - X| \to 0$)

(ii) $X_n \to X$ in Wahrscheinlichkeit und $\{X_n : n \in \mathbb{N}\}$ gleichgradig integrierbar.

Beweis:

(i)$\implies$(ii): Wegen $(0 \le)\ \epsilon \cdot P(|X_n - X| > \epsilon) \le E|X_n - X|$ folgt aus der $\mathfrak{L}^1$-Konvergenz die Konvergenz in Wahrscheinlichkeit.
Zum Nachweise der gleichgradigen Integrierbarkeit sei $\epsilon > 0$.
Zu diesem ϵ existiert ein n_0 mit $E|X_n - X| \le \frac{\epsilon}{2} \quad \forall n \ge n_0$.
Lemma 15.25 (n_0-mal angewendet) liefert ein $\delta > 0$ derart, daß für alle $A \in \mathfrak{A}$ mit $P(A) < \delta$ gilt:

$(*) \qquad \int\limits_A |X_n| dP < \epsilon$ für $n = 1, \ldots, n_0 - 1,\ \int\limits_A |X| dP < \frac{\epsilon}{2}.$

Wegen $E|X_n| \le E|X_n - X| + E|X|$ ist die Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathfrak{L}^1$ beschränkt, es existiert also ein $C > 0$ mit $\frac{1}{C} \sup\limits_{n\in\mathbb{N}} E|X_n| < \delta$.
Insbesondere: $P(|X_n| \ge C) < \delta \quad$ für alle $n \in \mathbb{N} \quad$ ($\rightsquigarrow$ Markov-Ungleichung).
Mit $(*)$: $\int\limits_{|X_n|\ge C} |X_n| dP < \epsilon$ für $n = 1, \ldots, n_0 - 1,$

$$\int\limits_{|X_n|\geq C} |X_n|dP \leq \underbrace{\int\limits_{|X_n|\geq C} |X|dP}_{<\frac{\epsilon}{2}} + \underbrace{\int\limits_{|X_n|\geq C} |X_n - X|dP}_{\leq E|X_n-X|\leq\frac{\epsilon}{2}} < \epsilon \quad \text{für } n \geq n_0.$$

Da $\epsilon > 0$ beliebig war, folgt hieraus die behauptete gleichgradige Integrierbarkeit.

(ii)$\Longrightarrow$(i): Für alle $C > 0$ definiere $\varphi_C : \mathbb{R} \to \mathbb{R}$ durch $\varphi_C(x) := \begin{cases} C & , \quad x > C \\ x & , \quad |x| \leq C \\ -C & , \quad x < -C \end{cases}$

Zwischenbehauptung: $(*)$ $\quad \lim\limits_{n\to\infty} |\varphi_C(X_n) - \varphi_C(X)| = 0 \quad$ für alle $C > 0$.

Sei also $\epsilon > 0$. Man hat $|\varphi_C(x) - \varphi_C(y)| \leq |x - y|$, also

$E|\varphi_C(X_n) - \varphi_C(X)| = \int\limits_{|X_n-X|>\epsilon} |\varphi_C(X_n) - \varphi_C(X)|dP + \int\limits_{|X_n-X|\leq\epsilon} |\varphi_C(X_n) - \varphi_C(X)|dP \leq$

$2CP(|X_n - X| > \epsilon) + \epsilon \overset{n\to\infty}{\longrightarrow} \epsilon$, also folgt $(*)$.

Sei wieder $\epsilon > 0$. Der Satz von der majorisierten Konvergenz liefert $E|\varphi_C(X) - X| \to 0$ mit $C \to \infty$. Hiermit und mit der gleichgradigen Integrierbarkeit von $\{X_n : n \in \mathbb{N}\}$ erhält man ein $C > 0$ $\quad (C < \infty)$ mit:

$E|\varphi_C(X) - X| \leq \frac{\epsilon}{3}$.

$E|\varphi_C(X_n) - X_n| \quad (\leq \int\limits_{|X_n|\geq C} |X_n|dP) \quad < \frac{\epsilon}{3} \qquad \forall n \in \mathbb{N}$

und damit $E|X_n - X| \leq E|X_n - \varphi_C(X_n)| + E|\varphi_C(X_n) - \varphi_C(X)| + E|\varphi_C(X) - X| \leq \frac{2}{3}\epsilon + E|\varphi_C(X_n) - \varphi_C(X)|$.

Nach $(*)$ existiert ein n_0 mit: $E|\varphi_C(X_n) - \varphi_C(X)| < \frac{\epsilon}{3} \quad \forall n \geq n_0$. □

15.4.3 Gleichgradig integrierbare Martingale

Es sei wieder $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ eine Filtration und $\mathfrak{F}_\infty := \sigma(\bigcup\limits_{n=1}^{\infty} \mathfrak{F}_n)$.

Klar: $(\mathfrak{F}_n)_{n\in\mathbb{N}\cup\{\infty\}}$ ist wieder eine Filtration.

Satz 15.28 Es sei $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein gleichgradig integrierbares Supermartingal. Dann existiert eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit $X_n \to X_\infty$ $\quad P$-f.s. und in $\mathfrak{L}^1$ mit $n \to \infty$, sowie $X_n \geq E[X_\infty|\mathfrak{F}_n] \quad \forall n \in \mathbb{N}$. (auch $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}\cup\{\infty\}}$ ist also ein Supermartingal.)

Beweis:

Gleichgradig integrierbare Funktionen sind $\mathfrak{L}^1$-beschränkt ($\rightsquigarrow$ Aufgabe 65), also läßt sich Satz 15.21 anwenden und liefert ein X_∞ mit $X_n \to X_\infty \quad P$-f.s.

Hieraus folgt $X_n \to X_\infty$ in Wahrscheinlichkeit (Satz 10.2), also mit Satz 15.27 $E|X_n - X_\infty| \to 0$, d.h. die $\mathfrak{L}^1$-Konvergenz.

Sei nun $A \in \mathfrak{F}_n$. Die Supermaringaleigenschaft liefert $\int\limits_A X_n dP \geq \int\limits_A X_m dP \quad \forall m \geq n$.

Wegen $\quad |\int\limits_A X_m dP - \int\limits_A X_\infty dP| \leq E|X_m - X_\infty| \overset{m\to\infty}{\longrightarrow} 0 \quad$ folgt $\quad \int\limits_A X_n dP \geq \int\limits_A X_\infty dP \quad \forall A \in \mathfrak{F}_n$, d.h. $X_n \geq E[X_\infty|\mathfrak{F}_n]$ ($\rightsquigarrow$ Aufgabe 54) □

Sei nun T eine beliebige gerichtete Indexmenge.

Ist $(\mathfrak{F}_t)_{t\in T}$ eine Filtration und X eine Zufallsvariable mit $E|X| < \infty$, so ist $(X_t, \mathfrak{F}_t)_{t\in T}$ mit $X_t := E[X|\mathfrak{F}_t]$ ein gleichgradig integrierbares Martingal ($\rightsquigarrow$ Beispiel 15.9(iii), Satz 15.26)

Satz 15.29 Es sei $(X_t, \mathfrak{F}_t)_{t\in T}$ ein gleichgradig integrierbares Martingal, $\mathfrak{F}_\infty := \sigma(\bigcup\limits_{t\in T} \mathfrak{F}_t)$.

Dann existiert eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit $X_t = E[X_\infty|\mathfrak{F}_t] \quad (P$-f.s.) für alle $t \in T$.

Beweis: Wir zeigen zunächst:

(1) $\quad \forall\epsilon > 0\ \exists t(\epsilon) \in T\ \forall s, t \in T : s \geq t(\epsilon), t \geq t(\epsilon) \implies E|X_s - X_t| < \epsilon$ ($(X_t)_{t\in T}$ ist eine Cauchy-Folge "entlang T")

Angenommen, (1) ist falsch. Dann existiert $\epsilon_0 > 0$ mit

(2) $\quad \forall t \in T\ \exists s, u \in T, s \geq t, u \geq t$ mit $E|X_s - X_u| \geq 2\epsilon_0$.

Wir können dazu rekursiv eine isotone Folge $(t_n)_{n\in\mathbb{N}} \subset T$ konstruieren mit:

(3) $\quad E|X_{t_{n+1}} - X_{t_n}| \geq \epsilon_0 \quad \forall n \in \mathbb{N}$.

Sei hierzu $t_1 \in T$ beliebig. Sind $t_1, \ldots, t_n$ bereits konstruiert, so existieren nach (2) $s, u \geq t_n$ mit $E|X_s - X_u| \geq 2\epsilon_0$. Nach der Dreiecksungleichung muß dann $E|X_s - X_{t_n}| \geq \epsilon_0$ oder $E|X_u - X_{t_n}| \geq \epsilon_0$ gelten. Wähle entsprechend $t_{n+1} = s$ oder $t_{n+1} = u$.

Nun ist $(X_{t_n}, \mathfrak{F}_{t_n})_{n\in\mathbb{N}}$ ein gleichgradig integrierbares Martingal, konvergiert nach Satz 15.28 im Widerspruch zu (3) im Mittel. Damit ist (1) bewiesen.

Wähle $\epsilon = \frac{1}{n}$ in (1), die Folge $t(\frac{1}{n})$, $n \in \mathbb{N}$, kann als isoton vorausgesetzt werden (hierbei wird "T gerichtet" gebrauchtet.) Dann liefert (1)

(4) $\quad E|X_t - X_{t(\frac{1}{n})}| \leq \frac{1}{m} \quad$ für alle $m \in \mathbb{N},\ n \geq m,\ t \geq t(\frac{1}{m})$.

Abermalige Anwendung von Satz 15.28 liefert eine Zufallsvariable $X \in \mathfrak{L}^1$, gegen die $X_{t(\frac{1}{n})}$ mit $n \to \infty$ im Mittel konvergiert. Läßt man in (4) $n \to \infty$ gehen, so folgt:

(5) $E|X_t - X| \leq \frac{1}{m}$ für alle $m \in \mathbb{N}$, $t \geq t(\frac{1}{m})$ (X_t konvergiert gegen X "entlang T")
Da $\mathfrak{L}^1$-Konvergenz Konvergenz in Wahrscheinlichkeit impliziert, und da zu in W. konvergenten Folgen f.s. konvergente Teilfolgen existieren ($\rightsquigarrow$ Aufgabe 31), gibt es eine Folge $(m_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}$ mit $X_{t(\frac{1}{m_n})} \to X$ P-f.s. Setze nun $X_\infty(\omega) = \begin{cases} X(\omega) & , \text{ wenn } \lim\limits_{n\to\infty} X_{t(\frac{1}{m_n})} \text{ existiert} \\ 0 & , \text{ sonst} \end{cases}$
Dann ist X_∞ $\mathfrak{F}_\infty$-meßbar, da alle $X_{t(\frac{1}{m_n})}$ $\mathfrak{F}_\infty$-meßbar sind. Es bleibt zu zeigen, daß alle $t \in T$ X_t eine Version von $E[X_\infty|\mathfrak{F}_t]$ ist. Sei hierzu $t \in T$, $A \in \mathfrak{F}_t$. Die Martingaleigenschaft liefert
(6) $\int\limits_A X_t dP = \int\limits_A X_s dP$ für alle $s \geq t$.
Zu jedem $m \in \mathbb{N}$ existiert ein $s \geq t(\frac{1}{m})$, $s \geq t$, also folgt mit (6) und (5)
$|\int\limits_A X_t dP - \int\limits_A X_\infty dP| \overset{P(X=X_\infty)=1}{=} |\int\limits_A X_t dP - \int\limits_A X dP| \leq E|X_s - X| \leq \frac{1}{m}$ für alle $m \in \mathbb{N}$.
Mit $m \to \infty$ folgt $\int\limits_A X_t dP = \int\limits_A X_\infty dP$ $\forall A \in \mathfrak{F}_t$. □

15.4.4 Ein Rückwärtskonvergenzsatz

Jetzt: Indexmenge $T = -\mathbb{N}$, Grenzübergang $n \to -\infty$.

Satz 15.30 (Rückwärtskonvergenzsatz von Doob)
Es sei $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ ein Martingal, $\mathfrak{F}_{-\infty} := \bigcap\limits_{n=1}^{\infty} \mathfrak{F}_{-n}$. Dann gilt mit $n \to -\infty$:
$X_n \to X_{-\infty} := E[X_{-1}|\mathfrak{F}_{-\infty}]$ P-f.s und im Mittel.

Beweis: Es sei $U_{-n}[a,b]$ die Anzahl der aufsteigenden Überquerungen von $[a,b]$ durch $X_{-n}, \ldots, X_{-1}$, $U_{-\infty}[a,b] = \lim\limits_{n\to\infty} U_{-n}[a,b]$.
Lemma 15.20 liefert $EU_{-n}[a,b] \leq \frac{1}{b-a} E(X_{-1} - a)^-$ (unabhängig von n)
<u>Also:</u> monotone Konvergenz ergibt $EU_{-\infty}[a,b] < \infty$, und damit $P(\exists a, b \in \mathbb{Q} : U_{-\infty}[a,b] = \infty) = 0$.
Dies liefert die Existenz eines (möglicherweise $\pm\infty$-wertigen) f.s. Grenzwertes $\tilde{X}_{-\infty}$.
Mit dem Lemma von Fatou folgt $E|\tilde{X}_{-\infty}| = E \liminf\limits_{n\to\infty} |X_{-n}| \leq \liminf\limits_{n\to\infty} E|X_{-n}| \leq E|X_{-1}| < \infty$, da $(X_{-n}, \mathfrak{F}_{-n})_{n\in\mathbb{N}}$ ein Submartingal ist (Satz 15.11.)
Hieraus folgt $P(|X_{-\infty}| = \infty) = 0$; Abänderung auf einer Nullmenge liefert eine $\mathbb{R}$-wertige $\mathfrak{F}_{-\infty}$-meßbare Zufallsvariable $X_{-\infty}$ mit $X_{-n} \to X_{-\infty}$ f.s.; $\mathfrak{L}^1$-Konvergenz folgt mit Satz 15.26 und Satz 15.27.
Dies wiederum kann wie im Beweis zu Satz 15.28 verwendet werden, um zu zeigen, daß $X_{-\infty}$ eine Version von $E[X_{-1}|\mathfrak{F}_{-\infty}]$ ist. □

15.5 Anwendungen

15.5.1 Ein 0-1-Gesetz von Kolmogorov

Als 0-1-Gesetze bezeichnet man Aussagen, die für die Wahrscheinlichkeit bestimmter Ereignisse nur die "Extremfälle" 0 und 1 zulassen.

Satz 15.31 Es seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen. $\mathfrak{T}_n := \sigma\{X_m : m \geq n\}$, $\mathfrak{T} := \bigcap\limits_{n=1}^{\infty} \mathfrak{T}_n$
($\mathfrak{T}$ heißt auch "σ-Algebra der <u>terminalen Ereignisse</u>".) Dann gilt $P(A) = 0$ oder $P(A) = 1$ für alle $A \in \mathfrak{T}$.

Beweis: Sei $A \in \mathfrak{T}$. Setze $\mathfrak{F}_n := \sigma\{X_1, \ldots, X_n\}$, $Y := 1_A$, $Y_n := E[Y|\mathfrak{F}_n]$.
Wegen $|Y_n| < 1$ (Jensen) läßt sich Satz 15.28 anwenden: Es gibt eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable Y_∞ mit $Y_n \to Y_\infty$ P-f.s. und im Mittel, und es gilt $Y_n = E[Y_\infty|\mathfrak{F}_n]$ für alle $n \in \mathbb{N}$.
Hiermit $\int\limits_B Y_\infty dP = \int\limits_B Y_n dP = \int\limits_B Y dP$ für alle $B \in \mathfrak{F}$.
Man kann $Y_\infty \geq 0$ annehmen (Monotonie des bedingten Erwartungswertes)
Die Maße $B \to \int\limits_B Y_\infty dP$, $B \to \int\limits_B Y dP$ stimmen also auf dem $\cap$-stabilen Erzeuger $\bigcup\limits_{n=1}^{\infty} \mathfrak{F}_n$ von $\mathfrak{F}_\infty$ und damit auf $\mathfrak{F}_\infty$ überein. Wegen $\mathfrak{T}_n \subseteq \mathfrak{F}_\infty$ $\forall n$ gilt $\mathfrak{T} \subseteq \mathfrak{F}_\infty$, also ist Y $\mathfrak{F}_\infty$-meßbar.
Der Eindeutigkeitssatz für Dichten führt also auf $P(Y = Y_\infty) = 1$.
Andererseits ist $\mathfrak{T}$ als Teilmenge von $\mathfrak{T}_{n+1}$ von $\mathfrak{F}_n$ unabhängig, und damit $Y_n = E[1_A|\mathfrak{F}_n] = E1_A = P(A)$ P-f.s.
Insgesamt: $P(\{\omega \in \Omega : \lim\limits_{n\to\infty} P(A) = 1_A(\omega)\}) = 1$, d.h. $P(A) = 0$ oder $P(A) = 1$. □

Beispiel 15.32 Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen, $\overline{X_n} := \frac{1}{n} \sum\limits_{i=1}^{n} X_i$. Offensichtlich ändern sich die Häufungspunkte der Folge $(\overline{X_n})_{n\in\mathbb{N}}$ nicht, wenn man beliebig viele Anfangsglieder $X_1, \ldots, X_n$ verändert.
Dies bedeutet, daß $Y := \limsup\limits_{n\to\infty} \overline{X_n}$ $\mathfrak{T}_n$-meßbar ist für alle $n \in \mathbb{N}$, also sogar $\mathfrak{T}$-meßbar.
Nach Satz 15.31 gilt $P(Y \in A) \in \{0, 1\}$ $\forall A$, d.h. es existiert eine Konstante $C \in [-\infty, \infty]$ mit $P(\limsup \overline{X_n} = C) = 1$.

Analog: Auch der lim inf ist fast sicher konvergent.
Hieraus folgt insgesamt: Konvergiert $\overline{X_n}$ f.s. gegen eine Zufallsvariable $\overline{X_\infty}$, so ist $\overline{X_\infty}$ P-f.s. konstant.

15.5.2 Das starke Gesetz der großen Zahlen

Ein Martingalbeweis für Satz 10.4:

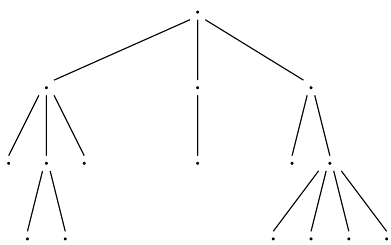
Satz 15.33 Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit (existierendem) Erwartungswert μ, so gilt $\overline{X_n} \to \mu$ P-f.s. $(\frac{1}{n}S_n \to \mu$ f.s.)
Beweis:
Es sei $\mathfrak{F}_{-n} := \sigma\{S_m : m \geq n\}$ $\forall n$, wobei $S_m := X_1 + \ldots + X_m$, $\mathfrak{F}_{-\infty} := \bigcap\limits_{n=1}^{\infty} \mathfrak{F}_{-n}$. Aus Beispiel 15.7(ii) ist $E[X_1|\mathfrak{F}_{-n}] = \overline{X_n}$, d.h. $(Y_n, \mathfrak{F}_n)_{n\in-\mathbb{N}}$ mit $Y_n := \overline{X_{-n}}$ ist ein Martingal (siehe auch Beispiel 15.9(iii)).
Satz 15.30 liefert die fast sichere Konvergenz gegen eine (f.s. endliche) $\mathfrak{F}_{-\infty}$-meßbare Zufallsvariable $X_{-\infty}$.
In der Notation von §15.5.1 gilt $\mathfrak{F}_{-\infty} = \mathfrak{T}$, also folgt mit Satz 15.31 und Beispiel 15.32 $Y_{-\infty} \equiv c$ P-f.s. mit einem $c \in \mathbb{R}$. Da außerdem Konvergenz in $\mathfrak{L}^1$ gilt, hat man $E\overline{X_n} \to EY_{-\infty}$, also $c = \mu$. □

15.5.3 Der Satz von Radon-Nikodym

Wir beschränken uns auf "normierte" Maße (WMaße). Der allgemeine Fall kann hieraus leicht hergeleitet werden.

Satz 15.34 Es sei $(\Omega, \mathfrak{A}, P)$ ein WRaum und Q ein weiteres WMaß auf $(\Omega, \mathfrak{A})$ mit $Q \ll P$ (Q wird von P dominiert, d.h. $P(A) = 0 \implies Q(A) = 0 \quad \forall A \in \mathfrak{A}$). Dann existiert eine $\mathfrak{A}$-meßbare Funktion $f : \Omega \to \mathbb{R}_+$ mit $Q(A) := \int\limits_A f dP \quad \forall A \in \mathfrak{A}$ (In Worten: Q hat eine Dichte bzgl. P.)
Beweis: Wir benötigen eine Hilfsaussage, die ganz ähnlich wie Lemma 15.25 bewiesen wird:
$(*) \qquad \forall \epsilon > 0\ \exists \delta > 0\ \forall A \in \mathfrak{A} : P(A) < \delta \implies Q(A) < \epsilon.$
Wir verwenden die Konstruktion aus Beispiel 15.9(ii):
T ist die Menge der (endlichen, meßbaren) Zerlegungen $t = \{A_1, \ldots, A_n\}$, $A_1, \ldots, A_n \in \mathfrak{A}$, $A_i \cap A_j = \emptyset$ für $i \neq j$, $\bigcup\limits_{i=1}^{n} A_i = \Omega$; $\mathfrak{F}_t := \sigma(\{A_1, \ldots, A_n\})$, $X_t = \sum\limits_{i=1}^{n} \frac{Q(A_i)}{P(A_i)} 1_{A_i}$ (Dichte von $Q|_{\mathfrak{F}_t}$ bzgl. $P|_{\mathfrak{F}_t}$.)
Bereits bekannt: $(X_t, \mathfrak{F}_t)_{t\in T}$ ist ein nicht-negatives Martingal. Jedes $B \in \mathfrak{F}_t$ ist endliche disjunkte Vereinigung von A_i-Mengen, insbesondere
$(**) \qquad \int\limits_B X_t dP = \sum\limits_{i:A_i\subseteq B} \frac{Q(A_i)}{P(A_i)} P(A_i \cap B) = \sum\limits_{A_i\subseteq B} Q(A_i) = Q(B).$
Speziell: $E_P X_t = 1$, $\int\limits_{X_t>C} X_t dP = Q(X_t > C) \quad \forall C > 0.$
Wir wollen nun zeigen, daß $\{X_t : t \in T\}$ gleichgradig integrierbar ist. Sei hierzu $\epsilon > 0$. Wähle δ gemäß $(*)$ und $C > \delta^{-1}$. Dann gilt $P(X_t > C) \leq \frac{1}{C} E_P X_t < \delta$, also $\int\limits_{X_t>C} X_t dP = Q(X_t > C) < \epsilon$ wegen $(*)$.
Nach Satz 15.29 existiert nun eine $\mathfrak{F}_\infty$-meßbare Zufallsvariable X_∞ mit
$(***) \qquad X_t = E[X_\infty|\mathfrak{F}_t] \quad \forall t \in T.$
Klar: $\mathfrak{F}_\infty \subseteq \mathfrak{A}$. Betrachte, zu gegebenem $A \in \mathfrak{A}$, die Zerlegung $t = \{A, A^c\}$. Dies zeigt $\mathfrak{A} \subset \mathfrak{F}_\infty$, also $\mathfrak{A} = \mathfrak{F}_\infty$; außerdem folgt mit $(**)$ und $(***)$ $Q(A) = \int\limits_A X_t dP = \int\limits_A X_\infty dP$, also leistet $f := X_\infty$ das Verlangte. □

15.5.4 Verzweigungsprozesse



$\{X_{nj} : n \in \mathbb{N}, j \in \mathbb{N}\}$ unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen, $p_k := P(X_{nj} = k)$ für $k \in \mathbb{N}_0$, $n, j \in \mathbb{N}$.
Interpretation: X_{nj} ist die Anzahl der Nachkommen vom Mitglied j der Generation n. Der Verzweigungsprozeß $(Z_n)_{n\in\mathbb{N}_0}$ wird rekursiv definiert durch
$Z_0 := 1$, $Z_{n+1} := \sum\limits_{k=1}^{Z_n} X_{n+1,k}$ für $n = 1, 2, \ldots$

Anwendungen:	überleben von Familiennahmen	(19. Jh.)
	Kernspaltung	(20. Jh.)

Mittlere Familiengröße $\mu := EX_{nj} = \sum\limits_{k=0}^{\infty} kp_k < \infty.$
Mit welcher Wahrscheinleichkeit stirbt ein solcher Prozeß aus?
Klar: $Z_n = 0 \implies Z_k = 0 \quad \forall k \geq n$
$A_n := \{Z_n = 0\}$, $q_n := P(A_n)$
$A_n \uparrow A := \{\exists n : Z_n = 0\}$ ("Prozeß stirbt aus")
$q_n \uparrow q := P(A)$ Frage: $q =?$
Der Fall $p_0 = 0$ liefert $q = 0$, und wird nun ausgeschlossen.

Wahrscheinlichkeitserzeugende Funktion: Y sei eine $\mathbb{N}_0$-wertige Zufallsvariable.
Dann heißt $g(t) := Et^Y = \sum\limits_{k=0}^{\infty} t^k P(Y = k)$ die wahrscheinlichkeitserzeugende Funktion zu Y

(bzw. zur Verteilung von Y.)

Man hat $g_{X+Y}(t) = g_X(t) g_Y(t)$, wenn X, Y unabhängig
$g'(1) = EX$ etc.
Ziel: "Darstellung" der Aussterbewahrscheinlichkeit q über die wahrscheinlichkeitserzeugende Funktion $g(t) = \sum t^k p_k$ der Nachkommenverteilung.

Lemma 15.35 Es sei g_n die wahrscheinlichkeitserzeugende Funktion zu Z_n.
Dann gilt $g_n = g \circ g \circ \ldots \circ g$ (n-fache Hintereinanderschaltung)

Beweis: $Z_1 = X_{11}$, die Behauptung gilt also für $n = 1$. Gilt sie für ein $n \in \mathbb{N}$, so folgt

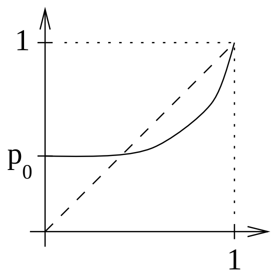
$$g_{n+1}(t) = Et^{Z_{n+1}} = \sum_{j=0}^{\infty} E \underbrace{t^{\sum_{k=1}^{j} X_{n+1,k}}} \underbrace{1_{\{Z_n = j\}}} = \sum_{j=0}^{\infty} g(t)^j P(Z_n = j) = g_n(g(t)).$$ □

Satz 15.36 Die Aussterbewahrscheinlickeit q ist die kleinste nicht-negative Lösung der Gleichung $g(t) = t$.

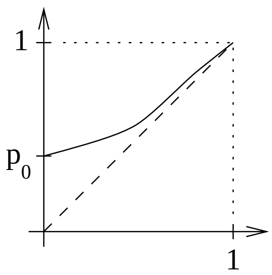
Beweis: Sei wieder $q_n := P(Z_n = 0)$. Mit Lemma 15.35 folgt $q_n = g_n(0) = g(g_{n-1}(0)) = g(q_{n-1})$.
Da g auf $[0,1]$ stetig ist, liefert der Grenzübergang $n \to \infty$ die Gleichung $q = g(q)$.
Ist andererseits r irgendeine nicht-negative Lösung von $g(t) = t$, so folgt, da g schwach monoton wachsend ist,
$q_1 = g(0) \leq g(r) = r$
$q_2 = g(q_1) \leq g(r) = r$, also $g_n \leq r \quad \forall n \in \mathbb{N}$ und damit $g \leq r$. □

Korollar 15.37 Es sei p_0 $(= P(X_{nk} = 0)) > 0$. Dann gilt: $q = 1 \iff \mu \leq 1$.

Beweis: $g''(t) = \sum\limits_{k=2}^{\infty} p_k k(k-1) t^{k-2} \geq 0$ für $0 \leq t \leq 1$. Also ist g auf $[0,1]$ konvex.



Fall $\mu > 1$



Fall $\mu \leq 1$

(Formaler Beweis: $\rightsquigarrow$ Krengel) □

Bei Modellen dieser Art ist die Suche nach zugehörigen Martingalen oft lohnend:
Setzt man $\mathfrak{F}_n := \sigma\{Z_1, \ldots, Z_n\}$, so erhält man

$$E[Z_{n+1} | \mathfrak{F}_n] = E[\sum_{j=1}^{Z_n} X_{n+1,j} | \mathfrak{F}_n] = E[\sum_{k=1}^{\infty} (\sum_{j=1}^{k} X_{n+1,j}) 1_{\{Z_n = k\}} | \mathfrak{F}_n] \overset{\text{monotone}}{\underset{\text{Konvergenz}}{=}} \sum_{k=1}^{\infty} E[(\sum_{j=1}^{k} X_{n+1,j}) 1_{\{Z_n = k\}} | \mathfrak{F}_n] \overset{15.5(\text{ii})}{\underset{15.5(\text{iii})}{=}}$$

$$\sum_{k=0}^{\infty} \underbrace{E(\sum_{j=1}^{k} X_{n+1,j})}_{k\mu} 1_{\{Z_n = k\}} = \sum_{k=0}^{\infty} k\mu 1_{\{Z_k = k\}} = \mu Z_n,$$ also ist $(Y_n, \mathfrak{F}_n)_{n \in \mathbb{N}}$ mit $\boxed{Y_n := \mu^{-n} Z_n}$ ein Martingal.

Mit $Z_n \geq 0$ folgt $E|Y_n| = EY_n = EY_1 = EX_{11} = \mu$, also ist das Martingal $\mathfrak{L}^1$-beschränkt und mit Satz 15.21 folgt:

Satz 15.38 Es gibt eine Zufallsvariable Y_∞ mit $\frac{1}{\mu^n} Z_n \to Z_\infty$ P-f.s.

Bei $\mu \leq 1$ hat man wegen $P(Z_n = 0) \to q = 1$ natürlich $Y_\infty \equiv 0$ P-f.s.
Dies liefert ein Beispiel für ein Martingal, das fast sicher, aber <u>nicht</u> im Mittel konvergiert: $EY_n \equiv \mu \not\to 0 = EY_\infty$.

Index