Stochastik

Inhaltsverzeichnis

1. Vorlesung
16.04.2018

Teil I

Grundlagen der Maßtheorethischen Wahrscheinlichkeitstheorie

1 0-1 Gesetze

Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.

1.1 Definition (liminf,limsup)

Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{F}$. Dann definiert man

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{w \in \Omega : w \in A_n \ \infty \text{ -oft}\}$$

„Unendlich viele der Ereignisse A_n treten ein “.

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{w \in \Omega : w \in A_n \text{für schließlich alle n}\}$$

„Bis auf endlich viele treten alle der Ereignisse A_n ein “.

1.2 Bemerkung

$\liminf\limits_{n\to\infty} A_n \subseteq \limsup\limits_{n\to\infty} A_n$

1.3 Lemma (Borel-Cantelli)

Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{F}$.

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\limsup_{n\to\infty} A_n) = 0$$

Ist $(A_n)_{n\in\mathbb{N}}$ zusätzlich Unabhängig, so gilt $\sum\limits_{n=1}^{\infty} P(A_n) = +\infty \Leftrightarrow P(\limsup\limits_{n\to\infty} A_n) = 1$

Beweis

Es gilt: $P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \leq \sum\limits_{i=1}^{\infty} P(A_i)$

(a) $\forall \varepsilon > 0 \ \exists N \in \mathbb{N}$ mit

$$P\left(\limsup_{n\to\infty} A_n\right) \leq P\left(\bigcup_{m=N}^{\infty} A_m\right) \leq \sum_{m=N}^{\infty} P(A_m) \leq \varepsilon$$

(b) „$\Rightarrow$":
Es gilt $\forall \alpha > 0: \ 1-\alpha \leq e^{-\alpha}$ und
$\left(\limsup_{n\to\infty} A_n\right)^c = \liminf_{n\to\infty} A_n^c = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c$
$P\left(\bigcap_{m=n}^{\infty} A_m^c\right) = \prod_{m=n}^{\infty} (1-P(A_m))$ wegen Unabhängigkeit und Monotonie.
Somit folgt
$P((\limsup_{n\to\infty} A_n)^c) \leq \sum_{n=1}^{\infty} P\left(\bigcap_{m=n}^{\infty} A_m^c\right) = \sum_{n=1}^{\infty} \prod_{m=n}^{\infty} (1-P(A_m)) \leq$
$\sum_{n=1}^{m} \prod_{m=n}^{\infty} e^{-P(A_m)} = \sum_{n=1}^{\infty} \underbrace{e^{\overbrace{-\sum_{m=n}^{\infty} P(A_m)}^{\infty}}}_{0} = 0$
„$\Leftarrow$": $\sum_{n=1}^{\infty} P(A_n) < \infty$ wäre ein Widerspruch zu a)

$\square$

$\limsup, \liminf$ sind Beispiele *Terminaler Ereignise*. 2. Vorlesung 17.04.2018

1.4 Definition (Terminale Ereignisse)

Seien $(\mathcal{F}_n)_{n\in\mathbb{N}} \subset F$ σ-Algebren und $T_n = \sigma\left(\bigcup_{m=n}^{\infty} \mathcal{F}_m\right)$.
Dann heißt $T_\infty := \bigcap_{n=1}^{\infty} \mathcal{F}_n$
terminale σ-Algebra (σ-Algebra der Terminalen Ereignisse).
Terminale Ereignisse hängen nicht von den ersten Endlich vielen $\mathcal{F}_n$ ab.

1.5 Beispiel

Sei $(A_n)_{n\in\mathbb{N}} \in F$. Setze $F_n = \{p, \Omega, A_n, A_n^c\}$.
Dann gilt

$$\bigcup_{m=n+i}^{\infty} A_m \in T_n \ \forall n \in \mathbb{N}, i \in \mathbb{N}_0$$

. Somit

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \in T_n \forall n \in \mathbb{N} \Rightarrow \limsup_{n\to\infty} A_n \in T_\infty$$

.
Da T_∞ σ-Algebra, gilt auch

$$\liminf_{n\to\infty} A_n = \left(\limsup_{n\to\infty} A_n^c\right)^c \in T_\infty$$

1.6 Definition (Unabhängige Mengensysteme)

Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $I \neq \emptyset$ eine Indexmenge. Eine Familie von Mengensystemen $(\varphi_n)_{n\in\mathbb{N}}$ mit $\varphi_i \subseteq \mathcal{F}$ heißt unabhängig, falls $\forall y \subseteq I, |y| < \infty, y \neq \emptyset$ gilt

$$P\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} P\left(A_j\right) \; \forall A_j \in \varphi_j$$

Beachte: Sei $(x_i)_{i\in I}$ eine Familie von Zufallsvariablen mit $x_i\left(\Omega, \mathcal{F}, P\right) \mapsto (\Omega_i, \mathcal{F}_i)$, dann sind diese Unabhängig genau dann, wenn die Familie von σ-Algebren $\left(x_i^{-1}\left(\mathcal{F}_i\right)\right)_{i\in I}$ unabhängig ist.

1.7 Satz (0-1 Gesetze von Kolmogorov)

Sei $(\mathcal{F}_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger σ-Algebren, $\mathcal{F}_n \subseteq \mathcal{F}$. Dann gilt

$$P\left(A\right) \in \{0,1\} \; \forall A \in T_\infty$$

Beweis:
Für $A \in T_\infty$ setze $\mathcal{D} = \{P \in \sigma\left(\bigcup_{n=1}^{\infty} F_n\right) :$ P ist unabhängig von $A\}$.
Es genügt zu zeigen: $T_\infty \subseteq \mathcal{D}$. Dann ist A unabhängig von A, und somit $P(A) = P(A \cap A) = P(A)^2$ und somit $P(A) \in \{0,1\}$.
Es gilt: T_{n+1} und $G_n := \sigma(\bigcup_{i=1}^{n} \mathcal{F}_i)$ sind unabhängig.
Da $A \in T_{n+1}$ folgt $G_n \subseteq \mathcal{D} \; \forall n \in \mathbb{N} \Rightarrow G_0 := \bigcup_{n=1}^{\infty} G_n \subseteq \mathcal{D}$. Wegen $G_n \subseteq G_{n+1}$ ist G_0 $\cap$-stabil. Mit dem Eindeutigkeitssatz für Maße sind folglich
$\vartheta : \sigma(G_0) \to \mathbb{R}^+, \quad G \mapsto P(G \cap A)$
$\mu : \sigma(G_0) \to \mathbb{R}^+, \quad G \mapsto P(G)P(A)$ identische Maße. $\Rightarrow \sigma(G_0) \subseteq \mathcal{D}$. Offensichtlich gilt $T_n \subseteq \sigma(G_0) \; \forall n \in \mathbb{N}$
$\Rightarrow T_\infty \subseteq \sigma(G_0) \subseteq \mathcal{D}$.

□

1.8 Korollar

Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Ereignisse. Dann gilt

$$P\left(\limsup_{n\to\infty} A_n\right) \in \{0,1\}$$

1.9 Korollar

Sei $X : (\Omega, \mathcal{F}) \to \left(\bar{\mathbb{R}}, B\left(\bar{\mathbb{R}}\right)\right)$ eine T_∞-Messbare Zufallsvariable. Dann gibt es eine Konstante $c \in \mathbb{R}$ mit $P\left(X = c\right) = 1$.

Beweis
$A_q := \{X \leq q\} \in T_\infty \ \forall q \in \mathbb{Q} \Rightarrow P(A_q) \in \{0,1\}$ und $P(A-q)$ monoton wachsend in q. Setze $c := \inf\{q \in \mathbb{Q} : P(A_q) = 1\}$. Dann folgt
$P(X < c) = P(\bigcup_{\substack{q<c \\ q\in\mathbb{Q}}} A_q) = 0$
$P(X > c) = P(\bigcup_{\substack{q>c \\ q\in\mathbb{Q}}} A_q) = 0$
$\Rightarrow P(X = c) = 1$

$\square$

1.10 Beispiel

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reeller Zufallsvariablen und $\alpha \searrow_{n\to\infty} 0$. Dann gilt:

$$P\left(\underbrace{\{\omega : \lim_{N\to\infty} \alpha_N \sum_{i=1}^{N} X_i(\omega) = 0\}}_{=:A}\right) \in \{0,1\}$$

Beweis
Sei $\mathcal{F}_n = X_n^{-1}(B(\mathbb{R}))$. Dann ist $(\mathcal{F}_n)_{n\in\mathbb{N}}$ eine unabhängige Folge von σ-Algebren.
$\alpha_N \sum\limits_{i=1}^{\infty} X_i = \underbrace{\alpha_N \sum_{i=1}^{n} X_i}_{\xrightarrow[N\to\infty]{} 0} + \underbrace{\alpha_N \sum_{i=n+1}^{\infty} X_i}_{\sigma_{n+1}-messbar} \quad \forall n \in \mathbb{N}$
$A = \bigcap\limits_{n=1}^{\infty} \underbrace{\{\alpha_N \sum_{i=n}^{N} X_i \xrightarrow[N\to\infty]{} 0\}}_{\in T_n} \in T_\infty$

$\square$

Inbesondere gilt:
$P(\{\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - E(X_i)) = 0\}) \in \{0,1\}$
für alle unabhängigen Folgen von Zufallsvariablen mit endlichem Erwartungswert, das heißt ein starkes Gesetz der Großen Zahlen gilt entweder fast sicher oder „fast sicher nicht“.

2 Konvergenzarten: fast sicher, stochastisch, in Lp

$(X_n)_{n\in\mathbb{N}}$ sei nun stets eine Folge von Zufallsvariablen und X eine Zufallsvariable mit $X-n, X : (\Omega, \mathcal{F}, P) \to \left(\mathbb{R}^d, B\left(\mathbb{R}^d\right)\right)$.
Dann gilt

$$\underbrace{\{\lim_{n\to\infty} X_n = X\}}_{\text{„Menge aller } \omega \text{, sodass } \lim \text{ existiert und gleich X ist"}} \qquad \in \mathcal{F}$$

2.1 Definition (Verschiedene Konvergenzarten (p,f.s.,Lp))

Wir sagen

1. X_n konvergiert p-fast sicher gegen X, falls

   $$P\left(\{\lim_{n\to\infty} X_n = X\}\right) = 1, \quad \text{in Zeichen: } X_n \xrightarrow{f.s.} X$$

   .

2. X_n konvergiert stochastisch in Wahrscheinlichkeit gegen X, falls

   $$P\left(\|X_n - X\| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0 \ \forall \varepsilon > 0 \text{ in Zeichen: } X_n \xrightarrow{p} X$$

   .

3. X_n konvergiert in $mcL^p\left(P\right)$ gegen X für $p \in (0, \infty)$, falls

   $$\|X_n\|, \|X\| \in \mathcal{L}^p \ \forall n \in \mathbb{N} \quad (E(\|X_n\|) < \infty, E(\|X\|) < \infty)$$
   $$E(\|X_n - X\|^p) \xrightarrow[n\to\infty]{} 0$$

   in Zeichen: $X_n \xrightarrow{\mathcal{L}^p} X$

4. X_n konvergiert in $\mathcal{L}^\infty\left(P\right)$ gegen X, falls

   $$\|X_n\|_\infty, \|X\|_\infty < \infty \ (d.h.\ X_n, X \in \mathcal{L}^\infty) \ \forall n \in \mathbb{N}$$
   $$E(\|X_n - X\|_\infty) \xrightarrow[n\to\infty]{} 0$$

   wobei für eine Zufallsvariable $Y \in \mathbb{R}^d$ gilt

   $$\|Y\|_\infty = \inf_{c\in\mathbb{R}} \{\|Y\| < c \ P - f.s\}$$

2.2 Lemma

3. Vorlesung 19.04.2018

Es gilt

(a) $\xrightarrow{f.s.} \Rightarrow \xrightarrow{p}$

(b) $\xrightarrow{\mathcal{L}^p} \Rightarrow \xrightarrow{p} \quad \forall p \in (0, \infty]$

Beweis Sei $\varepsilon > 0$ beliebig. Setze $A_n = \{\|X_n - X\| > \varepsilon\}$

(a) $\limsup\limits_{n\to\infty} P(\|X_n - X\| > \varepsilon) \overset{\text{Fatou}}{\leq} P\left(\limsup\limits_{n\to\infty} \{\|X_n - X\| > \varepsilon\}\right)$

$\qquad \| \qquad\qquad\qquad\qquad \|$

$E(1)_{\{\|X_n - X\| > \varepsilon\}} \qquad E\left(\limsup\limits_{n\to\infty} 1_{\{\|X_n - X\| > \varepsilon\}}\right)$

$= P(\|X_n - X\| > \varepsilon$ für ∞-viele n$) = 0$

$\left(\text{Beachte } \limsup\limits_{n\to\infty} 1_{A_n} = 1_{\limsup\limits_{n\to\infty} A_n}\right)$

(b) $p < \infty: \ P(A_n) = \int 1_{A_n} dP \leq \varepsilon^{-p} \int \|X_n - X\|^p \, dP \to 0 frn \to \infty$
$p = 0:$ offensichtlich

□

2.3 Bemerkung

(a) Die Umkehrungen in Lemma 2.2 sind im Allgemeinen falsch

(b) Bei allen Konvergenzarten aus der Definition 2.1. ist der Grenzwert (nur) fast sicher eindeutig: Gelte $X_n \overset{p}{\to} X, X_n \overset{p}{\to} Y$
$\Rightarrow P(\|X - Y\| > 2\varepsilon)$
$\leq P(\|X_n - X\| + \|X_n - Y\| > 2\varepsilon)$
$\leq P(\{\|X_n - X\| > \varepsilon\} \cup \{\|X_n - Y\| > \varepsilon\})$
$\leq P(\{\|X_n - X\| > \varepsilon\}) + P(\{\|X_n - Y\| > \varepsilon\}) \xrightarrow{n\to\infty} 0$
$\Rightarrow P(\|X - Y\| > 0) = 0$
$\Rightarrow X \overset{f.s.}{=\!=} Y$
In der Regel vernachlässigt man Unterschiede auf Nullmengen und nennt f.s. gleiche Zufallsvariablen einfach nur gleich. Formell sollte man (wie in Maßtheorie durchgeführt) bei diesen konvergenzarten Äquivalenzklassen fast sicher gleicher Zufallsvariablen betrachten, aber auch das vernachlässigt man fast immer in der Stochastik.

(c) Aus der Maßtheorie ist bekannt:

$(\mathcal{L}^\infty(P), \|\cdot\|_\infty)$ ist ein Banachraum

$(\mathcal{L}^p(P), \|\cdot\|_p)$ mit $\|X\|_p := (E(\|x\|^p))^{\frac{1}{p}}$ ist ein Banachraum für $p \in [1, \infty)$ und $\mathcal{L}^p(P)^* \xlongequal{\text{isometrisch isomorph}} \mathcal{L}^q(P)$ mit $\frac{1}{p} + \frac{1}{q} = 1$

Sei $Y \in \mathcal{L}^q$, dann ist $Y : \mathcal{L}^p \to \mathbb{R}, X \mapsto E(X^T Y))$

$(\mathcal{L}^2(P), \|\cdot\|_2)$ ist ein Hilbertraum mit Skalarprodukt $(X, Y) \mapsto E(X^T Y)$.

(d) Man kann zeigen:
$\mathcal{L}^p$ ist für $p \int (0,1)$ ein vollständiger metrischer Raum mit Metrik $d(X,Y) = E(\|X - Y\|^p)$

(e) Da $P(\Omega) = 1$ gilt $\mathcal{L}^q \subseteq \mathcal{L}^p \ \forall p, q \in \mathbb{R}^+$ mit q^p

(f) Fast sichere Konvergenz ist nicht metrisierbar.

2.4 Satz (Äquivalenz zu Konvergenz in p)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und X eine Zufallsvariable. Dann sind äquivalent:

(a) $X_n \xrightarrow{p} X \ , n \to \infty$

(b) Für jede Teilfolge $(n_k)_{k\in\mathbb{N}}$ von $\mathbb{N}$ $\exists$ eine Teilfolge dieser Teilfolge $(n_{k_l})_{l\in\mathbb{N}}$ mit $X_{n_{k_l}} \xrightarrow[l\to\infty]{f.s.} X$.

(c) $d(X_n, X) := E(1 \wedge \|X_n - X\|) \xrightarrow[n\to\infty]{} 0$

Beweis

a) $\Rightarrow$ *b)*
Sei $(n_k)_{k\in\mathbb{N}}$ beliebig.Für $j \in \mathbb{N}$ $\exists n_0(j)$, so dass $\forall n_k \geq n_0(j)$ gilt $P\left(\|X_{n_k} - X\| > \frac{1}{j}\right) \leq \frac{1}{j^2}$
$\Rightarrow \exists$ Teilfolge $\{n_{k_j}\}_{j\in\mathbb{N}}$ mit
$P\left(\left\|X_{n_{k_j}} - X\right\| > \frac{1}{j}\right) \leq \frac{1}{j^2}\ \forall j \in \mathbb{N}$
$\Rightarrow \sum_{j=1}^{\infty} P\left(\left\|X_{n_{k_j}}\right\| > \frac{1}{j}\right) \leq \sum_{j=1}^{\infty} \frac{1}{j^2} < \infty$
$\xRightarrow{Borel-Cantelli} P\left(\underbrace{\liminf_{j\to\infty}\left\{\left\|X_{n_{k_j}} - X\right\|\right\} \leq \frac{1}{j}}_{=A}\right) = 1$
$\Rightarrow X_{n_{k_j}}(\omega) \xrightarrow{j\to\infty} X(\omega)\, \forall \omega \in A$
$\Rightarrow X_{n_{k_j}} \xrightarrow{f.s.} X,\ j \to \infty$

b) $\Rightarrow$ *a)*
Für $\varepsilon > 0$ beliebig setze $a := \limsup_{n\to\infty} \underbrace{P(\|X_n - X\| > \varepsilon)}_{=:a_n}$
$\Rightarrow \exists$ Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $a_{n_k} \xrightarrow{n\to\infty} a$
Mit b) gibt es eine weitere Teilfolge $\left(n_{k_j}\right)_{j\in\mathbb{N}}$ mit $X_{n_{k_j}} \xrightarrow{f.s.} X,\ n \to \infty$, $\xRightarrow[\text{Lemma 2.2 a)}]{Beweis} a_{n_{k_j}} \xrightarrow{n\to\infty} 0$
$\Rightarrow a_{n_k} \to 0 \Rightarrow a = 0$

c) $\Rightarrow$ *a)*
analog zu Lemma 2.2b)

a) $\Rightarrow$ *c)*
folgt sofort aus dem kommenden Lemma 2.14

□

2.5 Bemerkung

Der Raum ($\mathcal{L}^0$) aller Zufallsvariablen mit Metrik $d(X,Y) = E(1 \wedge \|X-Y\|)$ ist ein vollständiger metrischer Raum und mit Satz 2.4 entspricht konvergenz in diesem der stochastischen Konvergenz von Zufallsvariablen.

2.6 Korollar (Stetigkeitssatz für Konvergenz in Wahrscheinlichkeit, continuous mapping theorem

(a) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge $\mathbb{R}^d$-wertiger Zufallsvariablen, X eine $\mathbb{R}^d$-wertige Zufallsvariable und $f : \mathbb{R}^d \to \mathbb{R}^m$ für $m \in \mathbb{N}$ stetig.
Dann gilt $X_n \xrightarrow{p} X \Rightarrow f(X_n) \xrightarrow{p} f(X)$

(b) Seien X, Y zwei Zufallsvariablen, $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ zwei Folgen von Zufallsvariablen mit $X_n \xrightarrow{p} X, Y_n \xrightarrow{p} Y$.
Dann gilt $X_n + Y_n \xrightarrow{p} X + Y,\ X_n * Y_n \xrightarrow{p} X * Y$

Beweis

(a) Die Aussage gilt offensichtlich für $\xrightarrow{f.s.}$, es genügt Satz 2.4 $a) \Leftrightarrow b)$ anzuwenden.

(b) $\left\| \begin{pmatrix} X_n \\ Y_n \end{pmatrix} - \begin{pmatrix} X \\ Y \end{pmatrix} \right\| \leq C\left(\|X_n - X\| + \|Y_n - Y\|\right)$ für C geeignet.

$$P\left(\left\| \begin{pmatrix} X_n \\ Y_n \end{pmatrix} - \begin{pmatrix} X \\ Y \end{pmatrix} \right\| > \varepsilon\right) \leq P\left(\left(\|X_n - X\| + \|Y_n - Y\|\right) > \frac{\varepsilon}{C}\right)$$
$$\leq P\left(\left(\|X_n - X\|\right) > \frac{\varepsilon}{2C}\right) + P\left(\left(\|Y_n - Y\|\right) > \frac{\varepsilon}{2C}\right)$$
$$\Rightarrow \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{p} \begin{pmatrix} X \\ Y \end{pmatrix}$$

Nun folgt b) direkt aus a)

$\square$

Wann $\xrightarrow{f.s}$ Konvergenz Konvergenz in $\mathcal{L}^p$ impliziert, folgt unmittelbar aus diversen Resultaten zur Integralkonvergenz, die in Maßtheorie gezeigt wurden. Wir fassen diese kurz aus stochastischer Sicht zusammen.

2.7 Satz (Beppo-Levy)

Seien $(X_n)_{n\in\mathbb{N}}$, X reellwertige Zufallsvariablen mit $X_n \nearrow_{f.s.} X$.
Dann gilt $E(X_n) \nearrow_{n\to\infty} E(X)$.
(Beachte: gegebenenfalls sind alle Erwartungswerte ∞).

2.8 Lemma (Fatou)

4. Vorlesung 23.04.2018

(a) Seien $(X_n)_{n\in\mathbb{N}}$ nichtnegative Zufallsvariablen, so gilt
$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} E(X_n)$

(b) Seien $(X_n)_{n\in\mathbb{N}}$ nichtnegative Zufallsvariablen. Existiert eine Zufallsvariable $X \in \mathcal{L}^1$ mit $X_n \leq X$ f.s. $\forall n \in \mathbb{N}$ so gilt
$E\left(\limsup_{n\to\infty} X_n\right) \geq \limsup_{n\to\infty} E\left(X_n\right)$

2.9 Satz (Dominierte Konvergenz)

Seien $(X_n)_{n\in\mathbb{N}}, X$ Zufallsvariablen mit $X_n \xrightarrow{f.s.} X$ und $\|X\| \leq Y$ f.s. für ein $Y \in \mathcal{L}^1$. Dann gilt
$X_n \xrightarrow{\mathcal{L}^1} X$ und $E(X_n) \xrightarrow{n\to\infty} E(X)$.

2.10 Satz (Vitali)

Seien $(X_n)_{n\in\mathbb{N}}, X$ Zufallsvariablen mit $X_n \xrightarrow{f.s.} X$ und $\limsup_{n\to\infty} \|X_n\|_p \leq \|X\|$ für ein $p \in (1,\infty)$. Dann gilt $X_n \xrightarrow{\mathcal{L}^p} X$.
Weitere wichtige Konvergenzsätze nutzen das Konzept der Gleichgradigen Integrierbarkeit (uniform integrability).

2.11 Definition (Gleichgradige Integrierbarkeit)

Sie $\mathcal{I}$ eine Indexmenge. Eine Familie $(X_i)_{i\in\mathcal{I}}$ von Zufallsvariablen heißt gleichgradig integrierbar, (UI), falls
$\forall \varepsilon > 0\ \exists k > 0:\ E\left(\|X_i\| 1_{\{\|X_i\|>k\}}\right) < \varepsilon\ \forall i \in \mathcal{I}$

2.12 Lemma (Einelementige Familien sind UI)

Sei $X \in \mathcal{L}^1$. Dann $\forall \varepsilon > 0\ \exists k > 0:\ E\left(\|X\| 1_{\{\|X\|>k\}}\right) < \varepsilon$, d.h. jede einelementige Familie integrierbarer Zufallsvariablen ist gleichgradig integrierbar. **Beweis**
Es gilt $\|X\| 1_{\{\|X\|>k\}} \xrightarrow[k\to\infty]{f.s.} 0$ und $\|X\| 1_{\{\|X\|>k\}} \leq \|X\| \in \mathcal{L}^1$.
Dominierte Konvergenz ergibt somit
$E\left(\|X\| 1_{\{\|X\|>k\}}\right) \xrightarrow{k\to\infty} 0$

□

2.13 Lemma (UI => Beschränkt)

Eine UI Familie von Zufallsvariablen ist beschränkt in $\mathcal{L}^1$.
Beweis
Sei K so,dass $E\left(\|X_i\| 1_{\{\|X_i\|>k\}}\right) < 1\ \forall i \in \mathcal{I}$.
$\|X_i\|_1 \leq E\left(\|X\| 1_{\{\|X\|\leq k\}}\right) + E\left(\|X\| 1_{\{\|X\|>k\}}\right) \leq k+1$.
Die Umkehrung gilt nicht.

2.14 Lemma

(a) Ist eine Familie von Zufallsvariablen beschränkt in $\mathcal{L}^p$ f+r $p > 1$, so ist sie UI.

(b) Sei $(X_i)_{i \in \mathcal{I}}$ eine Familie von Zufallsvariablen, $Y \int \mathcal{L}^1$ mit $\|X_i\| \leq Y$ f.s. $\forall i \in \mathcal{I}$. Dann ist $(X_i)_{i \in \mathcal{I}}$ UI.

Beweis

(a) $A := \sup_{i \in \mathcal{I}} E\left(\|X_i\|^p\right)$. Seien $v \geq k > 0$. Offensichtlich gilt $E\left(\|X_i\| 1_{\{\|X_i\| > k\}}\right) \leq k^{1-p} E\left(\|X_i\|^p 1_{\{\|X_i\| > k\}}\right) \leq k^{1-p} A$

(b) Offenstichtlich, da $E\left(\|X_i\| 1_{\{\|X_i\| > k\}}\right) \leq E\left(Y 1_{\{\|X_i\| > k\}}\right)$

2.15 Lemma (Beschränkte Konvergenz)

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen und X eine Zufallsvariable. Falls $X_n \xrightarrow[n \to \infty]{p} X$ und ein $k \geq 0$ existiert mit $\|X_n(\omega)\| \leq k$ $\forall n \in \mathbb{N}, \omega \in \Omega$, so gilt $X_n \xrightarrow[n \to \infty]{\mathcal{L}^1} X$ und insbesondere $E(X_n) \xrightarrow[n \to \infty]{E} (X)$

Beweis

Zunächst zeigen wir $P(\|X\| \leq k) = 1$.

Für $k \in \mathbb{N}$ gilt

$P\left(\|X\| > K + k^{-1}\right) \leq P\left(\|X - X_n\| > k^{-1}\right) \ \forall n \in \mathbb{N}$

$\Rightarrow P\left(\|X\| > k + k^{-1}\right) = 0$

$\Rightarrow P(\|X\| > k) = P\left(\bigcup_{k \in \mathbb{N}} \left\{\|X\| > k + k^{-1}\right\}\right) = 0.$

Sei nun $\varepsilon > 0$. Wähle n_0 so, dass

$$P\left(\|X_n - X\| > \frac{\varepsilon}{3}\right) < \frac{\varepsilon}{3k} \ \forall n \geq n_0$$

Dann gilt $\forall n \geq n_0$:

$E(\|X_n - X\|)$

$\leq E\left(\|X_n - X\| 1_{\{\|X_n - X\|\}} > \frac{\varepsilon}{3}\right) + \frac{\varepsilon}{3}$

$\leq 2kP\left(\|X_n - X\| > \frac{\varepsilon}{3}\right) + \frac{\varepsilon}{3} \leq \varepsilon \ \forall n \geq n_0$

2.16 Satz

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge in $\mathcal{L}^1$ und $X \in \mathcal{L}^1$.

Dann gilt $X_n \xrightarrow[n \to \infty]{\mathcal{L}^1} X$ genau dann, wenn

(i) $X_n \xrightarrow{p} X$

(ii) $(X_x)_{x\in\mathbb{N}}$ ist UI

Beweis

Nur für „$\Leftarrow$“ und reellwertige Zufallsvariablen!

Für $k \geq 0$ definiere $Y_k : \mathbb{R} \to [-k, k]$ durch

$Y_k(x) = k1_{\{x>k\}} + x1_{\{|x|\leq k\}} - k1_{\{x<-k\}}$.

Es gilt

$E\left(|Y_k(X_n) - X_n|\right) = E(|Y_k(X_n) - X_n|1_{\{|X_n|>k\}})$

$\leq kP(|X_n| > k) + E(|X_n|1_{\{|X_n|>k\}})$

$\leq 2E(|X_n|1_{\{|X_n|>k\}}) \xrightarrow[k\to\infty]{X_n\ UI} 0 \quad \forall n \in \mathbb{N}$

Wähle nun für $\varepsilon > 0$ beliebig k so, dass

$E(|Y_k(X_n) - X_n|) < \dfrac{\varepsilon}{3}\ \forall n,\ E(|Y_k(X) - X|) < \dfrac{\varepsilon}{3}$

Y_k ist stetig $\Rightarrow Y_k(X_n) \xrightarrow{p} Y_k(X)$

Mit Lemma 2.13 gibt es $n_0 \in \mathbb{N}$ so, dass $\forall n \geq n_0$

$E(|Y_k(X_n) - Y_k(X)|) < \dfrac{\varepsilon}{3}$

$\Rightarrow \forall n \geq n_0$ gilt $E(|X_n - X|) \leq E(|X_n - Y_k(X_n)|) + E(|Y_k(X_n) - Y_k(X)|) + E(|Y_k(X) - X|) < \varepsilon$

□

3 Konvergenz in Verteilung

$\mathcal{C}_b\left(\mathbb{R}^d\right) := \left\{f : \mathbb{R}^d \to \mathbb{R} \text{ stetig und beschränkt}\right\}$

3.1 Definition (Konvergenz in Verteilung)

(a) Seien $\mu_n,\ n \in \mathbb{N},\ \mu$ Wahrscheinlichkeitsmaße auf $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$.
Dann heißt μ_n schwach konvergent gegen μ, falls
$\int f d\mu_n \xrightarrow{n\to\infty} \int f d\mu \ \forall f \in \mathcal{C}_b\left(\mathbb{R}\right)$
in Zeichen: $\mu_n \xrightarrow{\omega} \mu$

(b) Sind $\left(X_n\right)_{n\in\mathbb{N}}, X :\ (\Omega, \mathcal{F}, P) \to \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ für $n \in \mathbb{N}$, dann sagt man X_n konvergiert in Verteilung gegen X, falls $\mathcal{L}(X_n) \xrightarrow{\omega} \mathcal{L}(X)$
in Zeichen: $X_n \xrightarrow{d} X$

3.2 Lemma

$X_n \xrightarrow{d} X \Leftrightarrow E(f(X_n)) \to E(f(x)) \ \forall f \in \mathcal{C}_b\left(\mathbb{R}^d\right)$

3.3 Bemerkung

(a) Schwache Konvergenz kann auf allgemeinen Räumen definiert werden und ist metrisierbar (Prohorov-Metrik)

(b) Die Konvergenz in Verteilung ist (auch) für Zufallsvariablen, die auf verschiedenen Grundräumen $(\Omega_n, \mathcal{F}_n, P_n)$ und $(\Omega, \mathcal{F}, P)$ definiert sind, wohldefiniert.

Im Folgenden betrachten wir der Einfachheit halber nur relle Zufallsvariablen.

3.4 Satz (Vergleich mit Konvegenz in Wahrscheinlichkeit)

Sei $\left(X_n\right)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen, X eine Zufallsvariable und $C \in \mathbb{R}$ eine Konstante. Dann gilt:

(a) $X_N \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

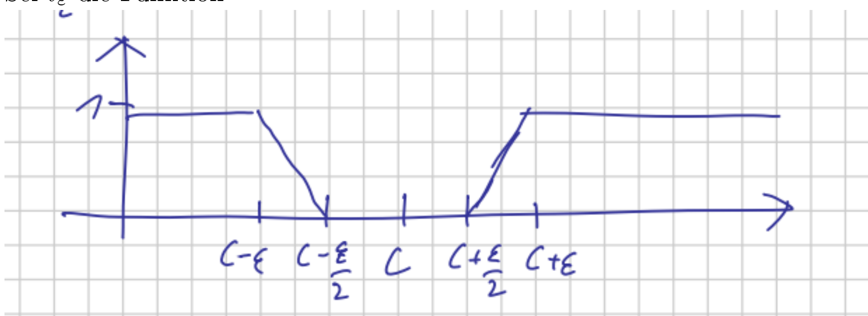
(b) $X_n \xrightarrow{d} C \Rightarrow X_n \xrightarrow{p} C$ (beachte: Konvergenz in Wahrscheinlichkeit gegen eine Konstante macht auch auf verschiedenen Grundräumen Sinn)

(c) $X_n \xrightarrow{d} X$ und $h : \mathbb{R} \to \mathbb{R}$ stetig $\Rightarrow h(X_n) \xrightarrow{d} h(X)$.

Beweis

(a) Korollar 2.6 + Lemma 2.15

(b) Sei t_ε die Funktion



für $\varepsilon > 0$.
Aus $X_n \xrightarrow{d} X$ folgt $P(|X_n - C| > \varepsilon) = \int 1_{\{|X_m - C| > \varepsilon\}} d\underbrace{P(X_n)}_{\mathcal{L}X_n} \leq$

$\int f_\varepsilon dP(X_n) = E(f_\varepsilon(X_n)) \xrightarrow[n\to\infty]{} E(f_\varepsilon(C)) = 0$

(c) klar, da $\forall f \in \mathcal{L}^p(\mathbb{R})$ gilt $f \circ h \in \mathcal{C}_b\left(\mathbb{R}^d\right)$

□

Die Umkehrung von Aussage a) ist falsch, ebenso ist es falsch aus $X_n \xrightarrow{d} X$ zu folgern, dass $P(X_n \in \mathcal{A}) \to P(X \in \mathcal{A})\ \forall \mathcal{A} \in \mathcal{B}(\mathbb{R})$

3.5 Definition (Quantilfunktion)

5.Vorlesung, 26.05.2018

Sei $F(x) = P(X \leq x)$ die Verteilungsfunktion einer reellen Zufallsvariable X.
Für $a \in (0,1)$ heißt $F^{\leftarrow} : (0,1) \to \mathbb{R}$
$u \mapsto F^{\leftarrow}(u) = \inf\left(t \in \mathbb{R} : F(t) \geq u\right)$
inverse Verteilungsfunktion/Quantilfunktion von F oder X.

3.6 Lemma (Eigenschaften der Quantilfunktion)

Sei F die Verteilungsfunktion einer Zufallsvariablen X. Dann gilt

(a) $F(F^{\leftarrow}(u)) \geq u\ \forall u \in (0,1)$

(b) Ist $F : \mathbb{R} \to (0,1)$ bijektiv, so ist $F^{\leftarrow} = F^{-1}$

(c) $F^{\leftarrow}$ ist (nicht zwangsweise streng) monoton steigend und linksstetig

(d) Für $x \in \mathbb{R}$ und $u \in (0,1)$ gilt $F(x) \geq u \Leftrightarrow x \geq F^{\leftarrow}(u)$

(e) $F(F^{\leftarrow}(u)) = u\ \forall u \in F(\mathbb{R}) \cap (0,1)$

Beweis

(a) F ist rechtsstetig $\Rightarrow$ inf in der Definition von $F^{\leftarrow}$ wird angenommen. Die Behauptung folgt direkt aus der Definition

(b) Folgt sofort aus $F(F^{-1}(u)) = u$ und Monotonie

(c) Übung

(d) $x \geq F^{\leftarrow}(u) \overset{F\nearrow}{\Longrightarrow} F(x) \geq F(F^{\leftarrow}(u)) \geq u$
$F(x) \geq u \Rightarrow x \geq \inf\{t : F(t) \geq u\} = F^{\leftarrow}(u)$

(e) „$\geq$“ folgt aus a)
Sei $t_0 \in \mathbb{R}$ mit $F(t_0) = u \in (0,1)$
$\Rightarrow F^{\leftarrow}(u) = \inf\{t : F(t) \geq u\} \leq t_0$
$\overset{F\nearrow}{\Longrightarrow} F(F^{\leftarrow}(u)) \leq F(t_0) = u.$

□

3.7 Satz (Verteilung durch Uniforme Verteilung berechnen)

Sei X eine reellwertige Zufallsvariable mit Verteilungsfunktion F und
$U \sim unif(0,1)$(Gleichverteilung auf $(0,1)$ =Lebesguemaß auf $(0,1)$)
Dann gilt

(a) $X \overset{\mathcal{D}}{=} F^{\leftarrow}(\mathcal{U})$, d.h. $\mathcal{L}(X) = \lambda|_{(0,1)} \circ (F^{\leftarrow})^{-1}$
Für $h \in \mathcal{L}(P(X))$ gilt $\int hdP(X) = \int\limits_0^1 h \circ F^{\leftarrow} d\lambda$

(b) Ist F stetig, so gilt
$F(X) = \mathcal{U}, d.h. \mathcal{L}(F(\lambda)) = \lambda|_{(0,1)}$
Ist $g \in \mathcal{L}^1(\lambda|_{(0,1)})$, so gilt
$\int\limits_0^1 gd\lambda = \int g \circ FdP(X)$

Beweis
Die Aussagen zu den Integralen folgen unmittelbar aus denen zu den Maßen

(a) $F(x) = P(\mathcal{U} \leq F(x)) = P(F^{\leftarrow}(\mathcal{U} \leq x) \; \forall x \in \mathbb{R}$

(b) F stetig $\xRightarrow{Zwischenwertsatz} (0,1) \subset F(\mathbb{R})$
Mit Lemma 3.6 e folgt $F(F^{\leftarrow}(u)) = u \; \forall u \in (0,1)$
$\mathcal{L}(F(X)) = \mathcal{L}(X) \circ F^{-1} = \lambda|_{(0,1)} \circ (F^{\leftarrow})^{-1} \circ F^{-1}$
$= \lambda|_{(0,1)} \circ (\underbrace{F \circ F^{\leftarrow}}_{=\text{Identität}})^{-1} = \lambda|_{(0,1)}$

□

Sobald man eine gleichverteilte Zufallsvariable realisieren/simulieren kann, kann man also eine beliebig verteilte simulieren.

3.8 Satz (Helly-Bray)

Seien $(X_n)_{n\in\mathbb{N}}$, $X : (\Omega, \mathcal{F}, P) \to \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ Zufallsvariablen mit Verteilungsfunktionen F_{X_n}, F_X. Setze
$S(F_X) = \{t \in \mathbb{R} : F_X(t) \text{ ist stetig in } t\}$.
Dann sind äquivalent

(a) $X_n \xrightarrow[n\to\infty]{d} X$

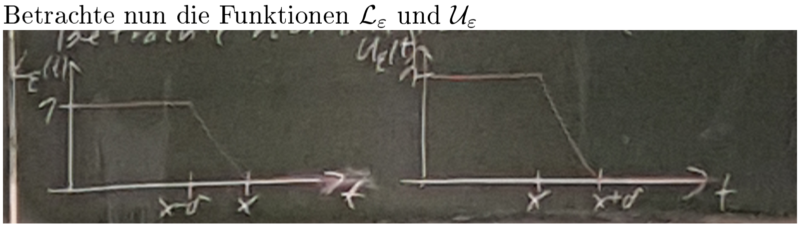
(b) $F_{X_n}(tr) \xrightarrow[n\to\infty]{} F_X(t)\ \forall t \in S(F_X)$

(c) $\exists D \subseteq \mathbb{R}$ dicht so, dass
$F_{X_n}(t) \xrightarrow[n\to\infty]{} F_X(t) \forall t \in D$

Beweis

- *b)* $\Rightarrow$ *c)*:
  Jede Verteilungsfunktion hat maximal abzählbar viele Sprungstellen $\Rightarrow$ $S(F_X)$ ist dicht in $\mathbb{R}$

- *c)* $\Rightarrow$ *b)*:
  Wähle $x_1, x_2 \in D$ mit $x_1 \leq x \leq x_2$ und $|F_X(x_{1/2}) - F_X(x)| < \varepsilon$
  $\Rightarrow \limsup\limits_{n\to\infty} F_{X_n}(X) \leq \lim\limits_{n\to\infty} F_{X_n}(x_2) = F_X(x_2) \leq F_X(x) + \varepsilon$
  $\liminf F_{X_n}(9x) \geq \lim\limits_{n\to\infty} F_{X_n}(x_1) = F_X(X_1) \geq F_X(x) - \varepsilon$
  Mit $\varepsilon \searrow 0$ folgt die Behauptung.

- *a)* $\Rightarrow$ *b)*:
  Sei $x \in S(F_X)$. Dann gibt es $\forall \varepsilon > 0$ ein $\delta > 0$ mit $|x - y| \leq \delta \Rightarrow |F_X(x) - F_X(y)| \leq \varepsilon$
  Betrachte nun die Funktionen $\mathcal{L}_\varepsilon$ und $\mathcal{U}_\varepsilon$
  

  Dann gilt
  $\limsup F_{X_n}(x) = \limsup \int 1_{(-\infty,x]} dP(X_n) \leq \limsup \int \mathcal{U}_\varepsilon dP(X_n) = \int \mathcal{U} dP(X) \leq F_X(x + \delta) \leq F_X(x) + \varepsilon$
  $\liminf F_{X_n}(x) \geq \liminf \int \mathcal{L}_\varepsilon dP(X_n) = \int \mathcal{L} dP(X) \geq F_X(x - \delta) \leq F_X(x) - \varepsilon$
  Mit $\varepsilon \searrow 0$ folgt *b)*

- *b)* $\Rightarrow$ *a)*:
  Der nächste Satz zeigt: wir können unter *b)* o.B.d.A. annehmen
  $X_n \xrightarrow[n\to\infty]{f.s.} X \Rightarrow X_n \xrightarrow{d} X$

$\square$

3.9 Satz (Skorohod)

Seien $(X_n)_{n\in\mathbb{N}}, X$ Zufallsvariablen mit Verteilungsfunktionen $F_{X-n}(t) \to F_X(t) \forall t \in S(F_X)$. Dann existiert ein Wahrscheinlichkeitsraum $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$ und Zufallsvariablen $\widehat{X}_n, \widehat{X} : (\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P}) \to \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ mit

(a) $\mathcal{L}(\widehat{X}_n) = \mathcal{L}(X_n), \ \forall n \in \mathbb{N}, \mathcal{L}(\widehat{X}) = \mathcal{L}(X)$

(b) $\widehat{X}_n \xrightarrow{n\to\infty} \widehat{X}$ $\widehat{P}$-f.s.

Wir zeigen zunächst:

3.10 Lemma

Seien $(F_n)_{n\in\mathbb{N}_0}$ Verteilungsfunktionen mit

$$F_n(t) \to F_0(t) \forall t \in s(F_0)$$

Dann gilt

$$F_n^{\leftarrow}(t) \to F_0^{\leftarrow}(t) \forall t \in S(F_0^{\leftarrow}) \cap (0,1)$$

Beweis

Sei $t \in S(F_0^{\leftarrow}), \ 0 < t < 1$. Zeige zunächst: 6.Vorlesung, 30.04.2018

$$\limsup_{n\to\infty} F_n^{\leftarrow}(t) \le F_0^{\leftarrow}(t) \tag{$*$}$$

Sei $t < t' < 1$ und $\varepsilon > 0$. Dann wähle $x \in S(F_0)$ mit $F_0^{\leftarrow}(t') < x < F_0^{\leftarrow}(t') + \varepsilon$
Es gilt $t < t' \le F_0(F_0^{\leftarrow}(t')) \le F_0(x)$.
Da $F_n(x) \to F_0(X)$, gibt es ein n_0 mit $F_n(x) > t \forall n \ge n_0$.
Mit 3.6]Lemma 3.6 d) $F_n^{\leftarrow}(t) \le x \forall n \ge n_0$
$\Rightarrow \limsup_{n\to\infty} F_n^{\leftarrow}(t) \le x < F_0^{\leftarrow}(t') + \varepsilon$.
Mit $t' \searrow t \in S(F_0^{\leftarrow})$ und $\varepsilon \searrow 0$ folgt $(*)$.
Analog zeigt man:
$\liminf_{n\to\infty} F_n^{\leftarrow}(t) \ge F_0^{\leftarrow}(t)$
$\Rightarrow$ Behauptung

$\square$

Beweis Satz 3.9

Wähle $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P}) = ((0,1), B((0,1), \lambda|_{(0,1)})$
Setze $\widehat{X}_n = F_{X_n}^{\leftarrow}, \ \widehat{X} = F_X^{\leftarrow}$. Dann gilt a) mit Satz 3.7 a
Zu b): Es Gilt $\widehat{P}(S(F_X^{\leftarrow})^c) = 0$, da F_X inf $\nearrow$
Somit folgt mit Lemma 3.10 aus $F_{X_n}(t) \to F_X(t) \forall t \in S(F_X)$, dass
$F_{X_n}^{\leftarrow}(t) = \widehat{X}_n(t) \to F_{X_0}^{\leftarrow}(t) = \widehat{X}_0(t)$ auf $S(F_X^{\leftarrow})$ also $\widehat{P}$-f.s.

$\square$

3.11 Proposition

Gilt $X_n \xrightarrow{d} X$, dann gilt auch

(a) $\liminf\limits_{n\to\infty} E(|X_n|) \geq E(|X|)$

(b) $E(X_n) \xrightarrow{n\to\infty} E(X)$, falls $(X_n)_{n\in\mathbb{N}}$ UI.

Beweis
Satz 3.9 + Lemma 2.8/Satz 2.16

□

3.12 Lemma (Slutzky)

Seien $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}, (Z_n)_{n\in\mathbb{N}}, X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ Zufallsvariablen, $d \in \mathbb{N}$ mit

- $X_n \xrightarrow{d} X$
- $\|X_n - Y_n\| \xrightarrow{p} 0$
- $Z_n \xrightarrow{p} c \in \mathbb{R}^d$

Dann gelten:

(a) $Y_n \xrightarrow{d} X$

(b) $\begin{pmatrix} X_n \\ Z_n \end{pmatrix} \to \begin{pmatrix} X \\ c \end{pmatrix}$

(c) $X_n + Z_n \xrightarrow{d} X + c$

(d) $X_n^T Z_n \xrightarrow{d} X^T c$

Beweis

(a) In der Übung wird gezeigt:
Es genügt zu zeigen $E(f(Y_n)) \to E(f(x)) \forall f \in \mathcal{C}_b(\mathbb{R})$ mit f gleichmäßig stetig
Sei also $f \in C(\mathbb{R}^d, \mathbb{R},\ \|f\|_\infty \leq M$ und $|f(x) - f(y)| \leq K\|x-y\|$ für $M, K > 0$. Dann gilt
$|E(f(X_n) - f(Y_n))| \leq E(|f(X_n) - f(Y_n)| * (1_{\{\|X_n - Y_n\| \leq \varepsilon\}} + 1_{\{\|X_n-Y_n\|>\varepsilon\}})$
$\leq K\varepsilon + 2MP(\|X_n - Y_n\| \geq \varepsilon) \xrightarrow[\varepsilon\searrow 0]{n\to\infty} 0$
$\Rightarrow E(f(Y_n)) \xrightarrow[n\to\infty]{} \lim\limits_{n\to\infty} E(f(X_n)) = E(f(X))$.

(b) Es gilt $\left\| \begin{pmatrix} X_n \\ Z_n \end{pmatrix} - \begin{pmatrix} X_n \\ C \end{pmatrix} \right\| \to 0$ und $\begin{pmatrix} X_n \\ C \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ C \end{pmatrix}$. Nun wende a) an.

(c) Folgt aus b) mit dem continuous mapping theorem

(d) Folgt aus b) mit dem continuous mapping theorem

□

3.13 Beispiel

Seien $(X_n)_{n\in\mathbb{N}} \sim N(0,1)$ iid. Dann folgt
$T_n = \frac{X_0}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2}}$ eine t_n-Verteilung.
Nach dem starken Gesetz der großen Zahlen gilt $\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \xrightarrow[f.s.]{n\to\infty} 1$
Mit Slutzky folgt $T_n \xrightarrow{d} \frac{X_0}{1} \sim N(0,1)$

3.14 Definition

(a) Eine Folge $(P_n)_{n\in\mathbb{N}}$ von Wahrscheinlichkeitsmaßen auf $\mathbb{R}$ heißt straff, falls $\forall\varepsilon > 0$ ein $k > 0$ existiert mit
$\sup\limits_{n\in\mathbb{N}} P_n([-K,K]^c) \leq \varepsilon$

(b) Eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen heißt straff, falls $(\mathcal{L}(X_n))_{n\in\mathbb{N}}$ straff ist.

3.15 Satz (Prohorov)

Jede straffe Folge von Zufallsvariablen besitzt eine in Verteilung konvergente Teilfolge von Zufallsvariablen
Beweis
Sei F_n die Verteilungsfunktion von X_n und $\mathcal{D} = \{x_m : m \in \mathbb{N}\}$ eine abzählbare dichte Teilmenge von $\mathbb{R}$.
Für $m \in \mathbb{N}$ betrachte die Folge $(F_n(x_m))_{n\in\mathbb{N}}$. Nach Bolzano-Weierstraß existiert eine Teilfolge $(n_{k,1})_{k\in\mathbb{N}}$ so, dass $(F_{n_{k,1}}(x_1)_{k\in\mathbb{N}}$ konvergiert. Wiederum existiert eine Teilfolge $(n_{k,2})_{k\in\mathbb{N}} \subset (n_{k,1})_{k\in\mathbb{N}}$ sodass $(F_{n_{k,2}}(x_2)_{k\in\mathbb{N}}$ konvergiert. Iterativ erhält man Teilfolgen $(n_{k,l})_{k\in\mathbb{N}} \quad \forall l \in \mathbb{N}$ mit $(F_{n_{k,l}}(x_i)_{k\in\mathbb{N}}$ konvergent $\forall i \leq l$.
Setze $n_k = n_{k,k}\ \forall k \in \mathbb{N}$.
Dann gilt $F_{nk}(x_i)$ konvergiert $\forall i \in \mathbb{N}$. Definiere nun $G : \mathcal{D} \to [0,1]$ durch $G(x_i) = \lim\limits_{n\to\infty} F_{n_k}(x_i)$ und setze G auf $\mathbb{R}$ fort durch $F(x) := inf\,\{G(y) : y \in \mathcal{D}, y > x\}$.
Gemäß Konstruktion ist F monoton wachsend und rechtsstetig.
F ist sogar eine Verteilungsfunktion, da wegen der Straffheit für jedes $\varepsilon > 0$ ein $k \in \mathbb{N}$ existiert mit $F(x) < \varepsilon \forall x < -k$ und $F(x) > 1 - \varepsilon \forall x > k$ und somit $\lim\limits_{x\to-\infty} F(x) = 1$ gilt.
Sei nun $x \in S(F)$. Dann gibt es zu jedem $\varepsilon > 0$ $y, z \in \mathcal{D}$ nut $y < x < z$ und $l_0 \in \mathbb{N}$ so, dass

$F(x) - \varepsilon \leq G(y) \leq F(x) \leq G(z) \leq F(x) + \varepsilon$ und
$F(x) - 2\varepsilon \leq F_{n_l}(y) \leq F_{n_l}(x) \leq F_{n_l}(z) \leq F_{n_l}(x) + 2\varepsilon \forall l \geq l_0$
$\Rightarrow F(x) - 2\varepsilon \leq \liminf\limits_{n\to\infty} F_{n_l}(x) \leq \limsup\limits_{n\to\infty} F_{n_l}(x) \leq F(x) + 2\varepsilon$
$\xRightarrow{\varepsilon \searrow 0} F_{n_l}(x) \xrightarrow{l\to\infty} F(x)$
Mit Helly-Bray folgt die Behauptung.

4 Charakteristische Funktionen

Wie üblich definieren wir für $f : \mathbb{R}^d \to \mathbb{C}$ messbar das Integral
$\int f d\mu = \int Re(f) d\mu + i \int Im(f) d\mu$ und (praktisch) alle Aussagen aus der Integrationstheorie bleiben mit den offensichtlichen Anpassungen gültig.

4.1 Definition

(a) Sei μ ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^d$. Dann heißt
$\widehat{\mu} : \mathbb{R}^d \to \mathbb{C}, t \mapsto \int e^{i\langle x,t\rangle} \mu(dx)$
die Fouriertransformierte von μ.

(b) Sei $x : (\Omega, \mathcal{F}, P) \to \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ eine Zufallsvariable. Dann heißt
$\varphi_X(t) = \widehat{P^X}(t) = E(e^{i\langle X,t\rangle})$
die charakteristische Funktion von X.

7.Vorlesung, 03.05.2018

Im Folgenden betrachten wir der Einfachheit halber fast immer nur $d = 1$. Praktisch alle Aussagen lassen sich unmittelbar auf beliebige Dimensionen Verallgemeinern. Da $|e^{i\langle x,t\rangle}| \leq 1 \forall x, t \in \mathbb{R}^d$ und $\mu(\mathbb{R}^d) = 1$ existiert die Fouriertransformierte stets.

4.2 Satz

Sei μ ein Wahrscheinlichkeitsmaß auf $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. Dann gelten:

(a) $\widehat{\mu}(0) = 1, |\widehat{\mu}(t)| \leq 1 \forall t \in \mathbb{R}^d$

(b) $\widehat{\mu}(\cdot)$ ist gleichmäßig stetig

(c) Für $a, b \in \mathbb{R}, X$ reellwertige Zufallsvariable
$\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$

(d) $\varphi_{-X} = \overline{\varphi_X}$

(e) (a) Sei ϑ ein weiteres Wahrscheinlichkeitsmaß auf $\mathbb{R}$:
$\widehat{\mu * \vartheta} = \widehat{\mu} * \widehat{\vartheta}$

(b) Seien $X_1, X_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen:
$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) Y_{X_2}(t)$

(f) $\widehat{\mu}$ ist positiv semidefinit, d.h. $\forall x_i \in \mathbb{R}, \lambda_i \in \mathbb{C}, n \in \mathbb{N}$ gilt
$\sum\limits_{j,l=1}^{n} \lambda_j \overline{\lambda_l} \widehat{\mu}(x_j - x_l) \geq 0.$

Beweis

(a) Trivial

(b) $\sup\limits_{t\in\mathbb{R}} |\varphi_X(t+h) - \varphi_X(t)|$
$= \sup\limits_{t\in\mathbb{R}} |E(e^{itX}(e^{ihx} - 1))|$
$\leq \sup\limits_{t\in\mathbb{R}} E(\underbrace{|e^{itx}|}_{=1}|e^{ihx} - 1|)$
$= E(|e^{ihx} - 1|) \xrightarrow[\substack{dom.\\ Konv.}]{h\to 0} 0$

(c) $\varphi_{aX+b}(t) = E(e^{i(aX+b)t}) = \underbrace{e^{iaXt}}_{\varphi_X(at)} e^{ibt}$

(d) $\overline{\varphi_X(t)} = \overline{E(e^{iXt})} = E(\overline{e^{iXt}}) = E(e^{-iXt}) = \varphi_{-X}(t)$

(e) $\varphi_{X_1+X_2}(t) = E(\underbrace{e^{it(X_1+X_2)}}_{|\cdot|\leq 1})$
$= \int_{\mathbb{R}^2} e^{it(X_1+X_2)} d(P^{X_1} \otimes P^{X_2})(X_1, X_2)$
$= \int\limits_{\mathbb{R}} \int\limits_{\mathbb{R}} e^{itX_1} e^{itX_2} dP^{X_1}(X_1) dP^{X_2}(X_2)$
$= \int\limits_{\mathbb{R}} e^{itX_1} dP^{X_1}(X_1) \int\limits_{\mathbb{R}} e^{itX_2} dP^{X_2}(X_2)$
$= \varphi_{X_1}(t)\varphi_{X_2}(t)$

(f) Es gilt
$\sum\limits_{j,l=1}^{n} \lambda_j \overline{\lambda_l} \widehat{\mu}(x_j - x_l) = \int\limits_{\mathbb{R}} \sum\limits_{j,l=1}^{n} \lambda_j \overline{\lambda_l} e^{i(x_j - x_l)z} \mu(dz)$
$= \int\limits_{\mathbb{R}} \sum\limits_{j,l=1}^{n} \lambda_j e^{ix_j z} \overline{\lambda_l e^{ix_l z}} \mu(dz)$
$= \int\limits_{\mathbb{R}} |\sum\limits_{j=1}^{n} x_j e^{ix_j z}|^2 \mu(dz) > 0.$

$\square$

4.3 Bemerkung

(a) Der Satz von Bochner besagt: Ist $\varphi : \mathbb{R} \to \mathbb{C}$ positiv semidefinit, $\varphi(0) = 1$ und φ stetig, dann gibt es ein Wahrscheinlichkeitsmaß μ mit $\varphi = \widehat{\mu}$.

(b) Gilt $X \stackrel{d}{=} -X$ (d.h. X ist symmetrisch um Null verteilt), so folgt $Im(E(e^{iuX})) = E(sin(uX)) = 0$

4.4 Beispiel

1. Sei $X \sim$ Bernoulli(p):
$\varphi_X(t) = e^{it}P(X = 1) + e^0 P(X = 0) = pe^{it} + (1 - p)$

2. Sei $X \sim$ Binomial (n,p)
$\varphi_X(t) \stackrel{4.2c}{=\!=} (pe^{it} + (1 - p))^n$

3. Sei $X \sim N(\mu, \sigma)$. Dann ist
$\varphi_X(t) = exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$

Beweis
Mit Satz 4.2 c) genügt es, die Aussage für $\mu = 0, \sigma^2 = 1$ zu zeigen. Sei $Y \sim N(0,1)$, dann ist Y symmetrisch um 0 verteilt. $\varphi_Y(t) = Re\varphi_Y(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{\mathbb{R}} cos(tx)e^{-\frac{x^2}{2}} dx$

$$\frac{d}{dt}\varphi_Y(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{\mathbb{R}} sin(tx) \left(-x^{-\frac{x^2}{2}} \right) dx$$

$$\overset{\substack{part.\\ int.}}{=\!=\!=} \frac{1}{\sqrt{2\pi}} \left(sin(tx)e^{-\frac{x^2}{2}} |_{-\infty}^{\infty} - t \int\limits_{\mathbb{R}} cos(tx)e^{-\frac{x^2}{2}} dx \right) = t\varphi_Y(t)$$

$$\Rightarrow \frac{\frac{d}{dt}\varphi_Y(t)}{\varphi_Y(t)} = -t$$

$\Rightarrow \frac{d}{dt} ln\varphi_Y(t) = -t$(in einer Nullumgebung wo $\varphi_Y(t) > 0$)
Außerdem gilt $ln\varphi_Y(0) = 0$

$$\Rightarrow ln\varphi_X(t) = -\frac{t^2}{2}$$

$\Rightarrow \varphi_X(t) = e^{-\frac{1}{2}t^2}$ (in einer Nullumgebung aber damit auf $\mathbb{R}$ da $e^{-\frac{1}{2}t^2} > 0 \forall t \in \mathbb{R}$

□

Hat μ Dichte f, so gilt
$\widehat{\mu}(t) = \int e^{itz}\mu(dz) = \int e^{itz} f(z)dz =: \widehat{f}(t)$. Die Fouriertransformierte des Maßes ist also die Fouriertransformierte (im klassischen analytischen Sinn) der Dichte.

4.5 Satz (Eindeutigkeitssatz)

Seien μ_1, μ_2 Wahrscheinlichkeitsmaße auf $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. Dann gilt
$\widehat{\mu_1} = \widehat{\mu_2} \Leftrightarrow \mu_1 = \mu_2$.
Für Zufallsvariablen X, Y gilt also
$X \stackrel{d}{=} Y \Leftrightarrow \varphi_X = \varphi_Y$.

4.6 Beispiel

Mithilfe der charakteristischen Funktion lassen sich Faltungen gegebenenfalls leicht berechnen.
Seien $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ unabhängig.
Dann gilt $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, da
$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t) = e^{i\mu_1 t - \frac{1}{2}\sigma_1^2 t^2} e^{i\mu_2 t - \frac{1}{2}\sigma_2^2 t^2}$
$= e^{i(\mu_1+\mu_2)t - \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$. Um Siatz 4.5 zu beweisen, zeigen wir zunächst:

4.7 Lemma

Seien μ, ν Wahrscheiunlichkeitsmaße auf $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. Dann gilt
$\int\limits_{\mathbb{R}} e^{-ixt}\widehat{\mu}(x)d\nu(x) = \int\limits_{\mathbb{R}} \widehat{\nu}(z-t)d\mu(z)$.

Beweis

$\int\limits_{\mathbb{R}} e^{-ixt}\widehat{\mu}(x)d\nu(x) = \int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}} e^{-ixt}e^{izx}\mu(dz)\nu(dx)$

$\overset{Fubini}{=\!=\!=} \int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}} e^{i(z-t)x}\nu(dx)\mu(dz) = \int\limits_{\mathbb{R}} \widehat{\nu}(z-t)\mu(dz)$

Beweis von Satz 4.5

„$\Leftarrow$ “ klar

„$\Rightarrow$ “

1. Sei $\nu = N(0, a^{-2})$ für $a > 0$ und μ ein beliebiges Wahrscheinlichkeitsmaß.
   Mit Lemma 4.7 gilt

   $\int\limits_{\mathbb{R}} e^{-ixt}\widehat{\mu}(x)d\nu(x) = \frac{a}{\sqrt{2\pi}} \int e^{-ixt}\widehat{\mu}(x)e^{-\frac{1}{2}x^2a^2} dx$

   $= \int\limits_{\mathbb{R}} e^{-\frac{1}{2}a^{-2}(z-t)^2} \mu(dz)$

   $\frac{1}{2\pi} \int e^{-ixt}\widehat{\mu}(x)e^{-\frac{1}{2}a^2x^2} dx = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2a^2}(z-t)^2} \mu(dz)$

   $=$ Dichte von $N(0, a^2) * \mu$ an der Stelle t.
   Somit gilt
   $(*)\widehat{\mu_1} = \widehat{\mu_2} \Rightarrow N(0, a^2) * \mu_1 = N(0, a^2) * \mu_2 \forall a > 0$

2. Wir zeigen nun: $(**) a_n \searrow 0 \Rightarrow N(0, a_n^2) \xrightarrow{\omega} \delta_0$
   Sei $X_n \sim N(0, a_n^2)$. Dann gilt mit Tschebyscheff
   $P(|X_n| \geq \varepsilon) \leq \frac{a_n^2}{\varepsilon^2} \to 0$ für $n \to \infty \forall \varepsilon > 0$
   $\Rightarrow X_n \xrightarrow{p} 0 \Rightarrow X_n \xrightarrow{d} 0.$

3. Seien $Y_1 \sim \mu_1, Y_2 \sim \mu_2$ und $X_n \sim N(0, a_n^2), n \in \mathbb{N}$, unabhängig.
   Dann gilt mit Slutzky $(***) X_n + Y_i \xrightarrow[n\to\infty]{d} Y_i$ für $i = 1, 2$
   $\Rightarrow \underbrace{P^{X_n+Y_1}}_{\xrightarrow{d}\mu_1} = P^{X_n} * P^{Y_1} = N(0, a_n^2) * \mu_1 = \underbrace{N(0, a_n^2) * \mu_2}_{\xrightarrow{d}\mu_2}$
   $\Rightarrow \mu_1 = \mu_2.$

$\square$

4.8 Satz (Stetigkeitssatz von Levy)

Seien μ_n, μ Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so gilt:

(a) $\mu_n \xrightarrow[n\to\infty]{\omega} \mu \Rightarrow \widehat{\mu}_n(t) \xrightarrow[n\to\infty]{} \widehat{\mu}(t) \forall t \in \mathbb{R}$.

(b) $\widehat{\mu}_n(t) \xrightarrow[n\to\infty]{} Y(t)$ punktweise mit $Y : \mathbb{R} \to \mathbb{C}$ stetig in Null $\Rightarrow \exists$ Wahrscheinlichkeitsmaß μ mit $\widehat{\mu} = Y$ und $\mu_n \xrightarrow{\omega} \mu$ für $n \to \infty$.

4.9 Bemerkung

8.Vorlesung, 07.05.2018

1. Seien X_n, X Zufallsvariablen, so gilt:
   $X_n \xrightarrow{d} X \Leftrightarrow \varphi_{X_n}(t) \to \varphi_X(t) \forall t \in \mathbb{R}$
2. Die Konvergenz in Satz 4.8 a) ist sogar lokal gleichmäßig

4.10 Lemma

Sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann gilt:
$\mu\left(\left\{x : |x| \geq \frac{1}{\lambda}\right\}\right) \leq \frac{7}{\lambda} \int\limits_0^\lambda 1 - Re(\widehat{\mu}(t))dt$

Beweis

$$\lambda^{-1} \int\limits_0^\lambda \underbrace{1}_{\int\limits_{\mathbb{R}} 1\mu(dx)} - \underbrace{Re(\widehat{\mu}(t))}_{\int\limits_{\mathbb{R}} cos(xt)\mu(dx)} \; dt = \lambda^{-1} \int\limits_0^\lambda \int\limits_{\mathbb{R}} (1 - cos(xt))\mu(dx)dt$$

$$\overset{Fubini}{=\!=\!=} \int\limits_{\mathbb{R}} \lambda^{-1} \int\limits_0^\lambda (1 - cos(xt))dt\mu(dx)$$

$$\geq \int\limits_{\{x:|x|\geq\lambda^{-1}\}} \lambda^{-1} \left(t - \frac{sin(xt)}{x}\right)\Bigg|_{t=0}^{\lambda} \mu(dx)$$

$$= \int\limits_{\{x:|x|\geq\lambda^{-1}\}} \left(1 - \frac{sin(\lambda x}{\lambda x}\right) \mu(dx)$$

$$\geq \int\limits_{\{x:|x|\geq\lambda^{-1}\}} (1 - sin(1))\mu(dx) \geq \frac{1}{7}\mu\left\{x : |x| \geq \lambda^{-1}\right\}$$

Denn $\frac{sin(t)}{t} \leq sin(1) \leq \frac{6}{7} \forall t \geq 1$

$\square$

Beweis von Satz 4.8

(a) Offensichtlich, da $sin(\cdot t), cos(\cdot t) \in \mathcal{C}_b(\mathbb{R}) \forall t$

(b) Wegen $\varphi(0) = 1$ und der Stetigkeit in 0 gilt:
$\forall \varepsilon > 0 \exists \lambda > 0$ mit
$1 - Re\varphi(t) \leq \frac{\varepsilon}{7} \forall t \in [-\lambda, \lambda]$ (da $Im\varphi(t) \xrightarrow[t\to 0]{} 0$)
und $Re\varphi(t) \xrightarrow[t\to 0]{} 1$. Mit Lemma 4.10 folgt

$$\limsup_{n\to\infty} \mu_n(\{x : |x| \geq \lambda^{-1}\}) \leq \limsup_{n\to\infty} \frac{7}{\lambda} \int\limits_0^\lambda \underbrace{(1 - Re\widehat{\mu}_n(t))}_{\substack{\to 1 - Re\varphi(t) \\ |\cdot|\leq 2}} dt$$

$$= \frac{7}{\lambda} \int\limits_0^\lambda \underbrace{(1 - Re(\varphi(t))}_{\leq \frac{\varepsilon}{7}} dt \leq \varepsilon$$

Somit gibt es $\forall \varepsilon > 0$ ein $n_0 \in \mathbb{N}$ so, dass $\mu_n((-\lambda, \lambda)^c) \leq 2\varepsilon$
Es folgt, dass $(\mu_n)_{n\in\mathbb{N}}$ straff ist.

Nach dem Satz von Prohorov gibt es eine Teilfolge $(n_l)_{l\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß μ mit $\mu_{n_l} \xrightarrow{\omega} \mu, l \to \infty$
$\overset{a)}{\Rightarrow} \widehat{\mu}_{n_l} \to \widehat{\mu}$ punktweise $\Rightarrow \widehat{\mu} = \varphi$
Angenommen $\mu_n \overset{\varphi}{\not\to} \mu$. Dann $\exists f \in \mathcal{C}_b(\mathbb{R})$ und eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $\left|\int f d\mu_{n_k} - \int f d\mu\right| > \varepsilon \forall k \in \mathbb{N}$.
Nun gibt es aber eine weitere Teilfolge $(n_{k_l})_{l\in\mathbb{N}}$ mit $\mu_{n_{k_l}} \xrightarrow{\omega} \mu$ ↯

□

4.11 Korollar

Seien $X_n, Y_n : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (für gleichen Index) unabhängige Zufallsvariablen für $n \in \mathbb{N}_0$ mit $X_n \xrightarrow{d} X_0, Y_n \xrightarrow{d} Y_0$.
Dann gilt $X_n + Y_n \xrightarrow{d} X_0 + Y_0$.
Beweis
$\varphi_{X_n+Y_n} = \varphi_{X_n}\varphi_{Y_n} \xrightarrow{pktweise} \varphi_{X_0}\varphi_{Y_0} = \varphi_{X_0+Y_0}$

□

4.12 Satz (Fourierinversion)

Sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $\int\limits_{\mathbb{R}} |\widehat{\mu}(x)| dx < \infty$
Dann ist μ absolutstetig (bzgl. des Lebesguemaßes) mit stetiger und beschränkter Dichte

$$f(t) = \int\limits_{\mathbb{R}} \frac{1}{2\pi} e^{-ixt} \widehat{\mu}(x) e^{-\frac{a^2x^2}{2}} dx$$

Beweis
Aus dem Beweis von Satz 4.5 wissen wir, dass $N(0, a^2) * \mu$ die Dichte

$f_a(t) = \frac{1}{2\pi} \int\limits_{\mathbb{R}} e^{-ixt} \widehat{\mu}(x) e^{-\frac{a^2x^2}{2}} dx$ hat (für $a > 0$)

$\Rightarrow |f(t)| \vee |f_a(t)| \leq \frac{1}{2\pi} \int\limits_{\mathbb{R}} |\widehat{\mu}(x)| dx < \infty \forall t \in \mathbb{R}$
Mittels dominierter Konvergenz folgt
$f_a(t) \xrightarrow{a \to 0} f(t)$
In allen Stetigkeitsstellen von μ gilt
$\int\limits_{(-\infty,x]} f_a(x) dx = N(0, a^2) * \mu((-\infty, x]) \xrightarrow[a\searrow 0]{} \mu((-\infty, x])(*)$
und andererseits gilt
$\int\limits_{(y,x]} f_a(x) dx \xrightarrow[\substack{dom.\\ konv.}]{a \to 0} \int\limits_{(y,x]} f(x) dx \overset{(*)}{=\!=} \mu((y, x])$
(falls y, x stetigkeitsstellen von μ)
$\xRightarrow{Monotonie} \int\limits_{(-\infty,x]} f(x) ds = \mu((-\infty, x])$ ($\forall$ Stetigkeitsstellen x von μ)
$\Rightarrow f$ ist Dichte von μ.
f ist beschränkt und stetig folgt aus $\widehat{\mu} \in \mathcal{L}^1$ und dominierter Konvergenz

$\square$

4.13 Proposition (Cramer-Wold-Device)

Seien $(X_n)_{n\in\mathbb{N}}, X : (\Omega, \mathcal{F}, P) \to \left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ Zufallsvariablen. Dann sind äquivalent:

(a) $X_n \xrightarrow{d} X, n \to \infty$

(b) $\langle \lambda, X_n \rangle \xrightarrow[n\to\infty]{d} \langle \lambda, X\rangle \forall \lambda \in \mathbb{R}^d$

Beweis

- *a)* $\Rightarrow$ *b)*
  folgt aus dem continuous mapping theorem, da $x \mapsto \langle \lambda, x\rangle$ stetig ist für $\lambda \in \mathbb{R}^d$
- *b)* $\Rightarrow$ *a)*
  Für alle $t \in \mathbb{R}^d$ gilt
  $\varphi_{X_n}(t) = E\left(e^{i\langle t, X_n\rangle}\right) \xrightarrow{b} E\left(e^{i\langle t, X\rangle}\right) = \varphi_X(t)$
  $\Rightarrow \varphi_{X_n} \to \varphi_X$ punktweise $\xRightarrow[satzv.Levy]{Stetigkeits-} X_n \xrightarrow{d} X$

$\square$

4.14 Satz

Sei $X \in \mathcal{L}^n$ für ein $n \in \mathbb{N}$. Dann ist φ_x n-fach stetig differenzierbar mit
$\varphi_x(t)^{(r)} = i^r \int\limits_{\mathbb{R}} e^{ixt} x^r dP^X(x)$
und es gilt
$E(X^r) = i^{-r} \varphi_X^{(r)}(0)$
für $r = 1, ..., n$
Beweis
n=1:
Sei $X \in \mathcal{L}^1$

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = E\left(\underbrace{e^{itX}}_{|\cdot|\le 1}\left(\underbrace{\frac{e^{ihX} - 1}{h}}_{\substack{\xrightarrow{h\to 0} iX \\ |\cdot|\le|X|}}\right)\right)$$

$$\xrightarrow[n\to 0]{dom.Konv.} E\big(e^{itX} iX\big) = i^1 \int\limits_{\mathbb{R}} e^{ixt} x dP^X(x)$$

$\Rightarrow$ differenzierbar mit stetiger Ableitung

$$\frac{d}{dt}\varphi_X(t)\Bigg|_{t=0} = i^1 \int x dP^X(x) = iE(X)$$

$n = 2, 3, ...$ Analog per Induktion

$\square$

Eng verbunden mit der charakteristischen Funktion ist die momenterzeugende Funktion / Laplacetransformierte eines Maßes / einer Zufallsvariable:
$\mathbb{R} \to \mathbb{R}, t \mapsto E(e^{tX}) =: M_X(t) \quad$ „$= \varphi_X(it)$“
Diese muss für kein t (außer 0) existieren! Für positive Zufallsvariablen existiert sie für $t \in (-\infty, 0]$.
Existiert die Laplacetransformierte in einer Nullumgebung, so sind alle Momente endlich $(X \in \mathcal{L}^(p) \forall p > 0)$.

5 Zentrale Grenzwertsätze

9. Vorlesung, 15.05.2018

Sei $(X_i)_{i\in\mathbb{N}}$ eine iid (independent identically distributed) Folge von Zufallsvariablen mit $X_1 \in \mathcal{L}^2$.
Setze $\mu = E(X_1), \sigma^2 = Var(X_1)$.
Wir nehmen von nun an an $\sigma^2 > 0$.
Für $n \in \mathbb{N}$ definieren wir
$S_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$.
Dann gilt $E(S_n) = 0, Var(S_n) = 1$.

5.1 Satz (ZGWS klassisch)

Unter obigen Voraussetzuzngen gilt
$S_n \xrightarrow[n\to\infty]{d} N(0,1)$
Beweis
Setze $Y_i = \frac{1}{\sigma}(X_i - \mu)$. Dann gilt $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$ und $Y_i \in \mathcal{L}^2 \Rightarrow S_n \in \mathcal{L}^2$.
Es bezeichne ν die Verteilung von Y_1.
Dann gilt mit Satz 4.14 und Taylor
$\varphi_{Y_1}(x) = \widehat{\nu}(x) = \widehat{\nu}(0) + \widehat{\nu}(0)'x + \frac{1}{2}\widehat{\nu}''(0)x^2 + o(x^2)$
$= 1 + \frac{1}{2}i^{-2}E(Y_1^2)x^2 + o(x^2) = 1 - \frac{1}{2}x^2 + o(x^2)$
$\varphi_{S_n}(x) = \widehat{\nu}\left(\frac{x}{\sqrt{n}}\right)^n = \left(1 - \frac{x^2}{2n} + o\left(\frac{x^2}{n}\right)\right)^n$
$= \left(1 + \frac{-\frac{1}{2}x^2 + no\left(\frac{x^2}{n}\right)}{n}\right)^n \xrightarrow{n\to\infty} e^{-\frac{1}{2}x^2} = \widehat{N(0,1)}(x)$

□

Mit mehr Aufwand und Voraussetzungen kann man die Konvergenzgeschwindigkeit (uniform für die Verteilungsfunktionen) bestimmen:

5.2 Satz (Berry-Esséen)

Unter den Voraussetzungen von Satz 5.1 und $X_1 \in \mathcal{L}^3$ gilt
$\sup_{t\in\mathbb{R}} |P(S_n \leq t) - \Phi(t)| \leq c\frac{E(|X_1|^3)}{\sqrt{n}\sigma} \forall c \geq c^*$ mit $0,4097 \leq c* \leq 0,4748$ wobei Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet
Mit dem starken Gesetz der großen Zahlen und Slutzky folgt sofort:

5.3 Korollar

Unter den Voraussetzungen von Satz 5.1 gilt

$$\frac{\sum\limits_{n=1}^{n}(X_i-\mu)}{\sqrt{n}\widehat{\sigma}_n} \xrightarrow{d} N(0,1)$$

mit $\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \frac{1}{n}\sum\limits_{i=1}^{n}X_i\right)^2$

Wir wenden uns nun der Verallgemeinerung des Zentralen Grenzwertsatzes auf Dreiecksschemata von Zufallsvariablen zu. Gegeben seien Zufallsvariablen X_{nl} : $(\Omega,\mathcal{F},P)\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ mit $n\in\mathbb{N}, l\in\{1,2,...,n\}$, die Zeilenweise unabhängig sind, d.H. $S_{n1},S_{n2},...,S_{nn}$ sind jeweils unabhängig.

5.4 Satz (Lindeberg-Feller)

Sei $(X_{nl})_{l\in\{1,...,n\}}, n\in\mathbb{N}$ ein Dreiecksschema zeilenweise unabhängiger Zufallsvariablen in $\mathcal{L}^2$ mit folgenden Eigenschaften:

(a) $\sum\limits_{l=1}^{n} E(X_{nl}) = 0 \quad \forall n\in\mathbb{N}$

(b) $\sum\limits_{l=1}^{n} Var(X_{nl}) \xrightarrow{n\to\infty} \sigma^2 \in (0,\infty)$

(c) (Lindeberg-Bedingung):
$L_n(\varepsilon) := \sum\limits_{l=1}^{n} E\left(X_{nl}^2 1_{\{|X_{nl}|>\varepsilon\}}\right) \xrightarrow{n\to\infty} 0 \quad \forall\varepsilon>0$

Dann gilt

$\sum\limits_{i=1}^{n} X_{nl} \xrightarrow{d} N(0,\sigma^2)$

Beweis

Technisch aufwendige Verallgemeinerung des Beweises von Satz 5.1

□

5.5 Lemma

Erfüllt ein Dreiecksschema von Zufallsvariablen die Lindeberg-Bedingung, so gilt:

(a) Das Dreiecksschema ist (uniform) asymptotisch vernachlässigbar, d.h.
$\max\limits_{l\in\{1,...,n\}} P(|X_{nl}|>\varepsilon) \xrightarrow{n\to\infty} 0 \quad \forall\varepsilon>0$

(b) Es erfüllt die Fellerbedingung, d.h.
$\max\limits_{l\in\{1,...,n\}} Var(X_{nl}) \xrightarrow{n\to\infty} 0$

Beweis

(a) Folgt aus b) und Tschebyscheff

(b) Für $\varepsilon > 0$ gilt
$Var(X_{nl}) \leq E(X_{nl}^2) = E(X_{nl}1_{\{|X_{nl]\leq\varepsilon}\}}) + E(X_{nl}1_{\{|X_{nl]>\varepsilon}\}})$
$\leq \varepsilon^2 + L_n(\varepsilon) \xrightarrow[\varepsilon\searrow 0]{n\to\infty} 0.$

$\square$

5.6 Lemma

Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von iid Zufallsvariablen in $\mathcal{L}^2$ mit $Var(X_1) > 0$. Dann erfüllt das Dreiecksschema

$$X_{nl} = \frac{X_l - E(X_1)}{\sqrt{n}}$$

die Voraussetzungen von Satz 5.4.
Beweis

(a) Zeilenweise unabhängigkeit ist Trivial

(b) $\sum\limits_{l=1}^{n} Var(X_{nl}) = \frac{1}{n}\sum\limits_{l=1}^{n} Var(X_l) = Var(X_1)$

(c) o.B.d.A sei $E(X_1) = 0$. Für $\varepsilon > 0$ gilt:

$$L_n(\varepsilon) = \frac{1}{n}\sum_{l=1}^{n} E\left(X_l^2 1_{\{|X_l|>\varepsilon\sqrt{n}\}}\right) = E\left(\underbrace{X_1^2 1_{\{|X_1|>\varepsilon\sqrt{n}\}}}_{\substack{\xrightarrow{n\to\infty} 0 pktweise \\ \leq X_1^2}}\right) \xrightarrow[n\to\infty]{dom.Konv} 0$$

$\square$

5.7 Lemma (Lyapunov-Bedingung)

Sei X_{nl} ein Dreiecksschema von Zufallsvariablen. Existiert ein $\delta > 0$ mit $X_{nl} \in \mathcal{L}^{2+\delta} \forall n, l$ und gilt
$\sum\limits_{l=1}^{n} E\left(|X_{nl}^{2+\delta}\right) \xrightarrow{n\to\infty} 0$, dann erfüllt das Dreiecksschema die Lindebergbedingung.
Beweis
Sei $\varepsilon > 0$.
$L_n(\varepsilon) = \sum\limits_{l=1}^{n} E(X_{nl}^2 \underbrace{\varepsilon^{\delta}}_{<|X_{nl}|^{\delta}} \varepsilon^{-\delta} 1_{\{|X_{n1}|>\varepsilon\}} \leq \varepsilon^{-\delta} \sum\limits_{l=1}^{n} E\left(|X_{nl}|^{2+\delta}\right) \xrightarrow{n\to\infty} 0$

5.8 Bemerkung

(a) Interpretiert man $N(0,0)$ als δ_0 so gilt Satz 5.4 auch, falls $\delta^2 = 0$.
Beweis
Dann gilt $Var\left(\sum\limits_{l=1}^{n} X_{nl}\right) \to 0$ und mit Tschebyschefffolgt $\sum\limits_{l=1}^{n} X_{nl} \xrightarrow{p} 0$

$\square$

(b) Es gibt Varianten des ZGWS,die keine Unabhängigkeit fordern, dafür aber geeignete asymptotische Unabhängigkeit (und mehr Momente).

(c) Es ist auch möglich, ähnliche Resultate für Zufallsvariablen ohne zweite Momente zu zeigen. Die Grenzverteilungen sind dann stabile Verteilungen.

5.9 Beispiel (Rekorde)

10.Vorlesung, 17.05.2018

Sei $(X_i)_{i\in\mathbb{N}}$ eine iid Folge von Zufallsvariablen mit $\mathcal{L}(X_1)$ stetig.

$$A_n := \left\{ \max_{1\le l\le n-1} X_l < X_n \right\} \quad \widehat{=}\text{neuer Rekord zur Zeit n})(A_1 = \Omega)$$

Dann gilt für $R_n = \sum\limits_{i=1}^{n} 1_(A_i)$

$$T_n = \frac{R_n - ln(n)}{\sqrt{ln(n)}} \xrightarrow{d} N(0,1)$$

Beweis
Aus der Stetigkeit von X_i und X_i iid folgt $P(A_n) = \dfrac{1}{n}$ (unter den ersten n Zufallsvariablen wird das Maximum an den Stellen $1, ..., n$ mit der gleichen Wahrscheinlichkeit angenommen)
$\Rightarrow 1_(A_n) \sim \text{Bernoulli}\left(\dfrac{1}{n}\right)$
Für $i < j$ gilt $P(A_i \cap A_j) = P(A_i|A_j)P(A_j) = \dfrac{1}{ij}$
Induktiv folgt $(1_{A_n})_{n\in\mathbb{N}}$ sind unabhängig.

$$E\left(1_{A_n}\right) = \frac{1}{n}, \quad Var\left(1_{A_n}\right) = \frac{1}{n}\left(1 - \frac{1}{n}\right) \sigma_n^2 := \sum_{l=1}^{n} Var\left(1_{A_l}\right) = \sum_{l=1}^{n} \frac{1}{l}\left(1 - \frac{1}{l}\right) \xrightarrow{n\to\infty} \infty$$

Setze $X_{nl} = \dfrac{1_{A_l} - \frac{1}{l}}{\sigma_n}$.
Dann gilt $E(X_{nl}) = 0, \quad \sum\limits_{i=1}^{n} Var\left(X_{nl}\right) = 1$
und wegen $|X_{nl}| \le \dfrac{2}{\delta_n} \xrightarrow{n\to\infty} 0$ gilt Lindeberg
$\Rightarrow \sum\limits_{l=1}^{n} X_{nl} \xrightarrow{d} N(0,1)$

Es gilt:

$$T_n = \frac{\sum\limits_{l=1}^{n} \left(1_{A_l} - \frac{1}{l} \right)}{\sigma_n} \cdot \frac{\sigma_n}{\sqrt{log(n)}} + \frac{\sum\limits_{l=1}^{n} \frac{1}{l} - log(n)}{\sqrt{log(n)}}$$

- $\frac{\sum\limits_{l=1}^{n} \frac{1}{l} - log(n)}{\sqrt{log(n)}} \to 0$, da gilt $log(n) = \int\limits_{1}^{n} \frac{1}{n} dx \leq \sum\limits_{l=1}^{n} \frac{1}{l} \leq log(n) + 1$
- $\frac{\sigma_n^2}{log(n)} \to 1$, denn

$$\sigma_n^2 = \sum_{l=1}^{n} \frac{1}{l} - \underbrace{\sum_{l=1}^{n} \frac{1}{l^2}}_{\xrightarrow{n \to \infty} c \in (0,\infty)}$$

$\square$

5.10 Satz (Lindeberg-Feller multivariat)

Sei $X_{nl} : (\Omega, \mathcal{F}, P) \left(\mathbb{R}^d, \mathcal{B} \left(\mathbb{R}^d \right) \right)$ ein Dreiecksschema zeilenweise unabhängiger Zufallsvektoren mit $X_{nl} \in \mathcal{L}^2 \quad \forall n, l$. Setze $\Sigma_{nl} = Cov(X_{nl})$ $(d \times d$ symmetrische strikt positiv definite Matrix). Wenn

(a) $\sum\limits_{l=1}^{n} E(X_{nl} = 0 \in \mathbb{R}^d$

(b) $\sum\limits_{l=1}^{n} \Sigma_{nl} \xrightarrow{n \to \infty} \Sigma(d \times d$ symmetrische strikt positiv definite Matrix)

(c) $L_n(\varepsilon) := \sum\limits_{l=1}^{n} E \left(\|Xnl\|^2 \, 1_{\|X_{nl}\| > \varepsilon} \right) \xrightarrow{n \to \infty} 0 \quad \forall \varepsilon > 0$

dann gilt

$$\sum_{l=1}^{n} X_{nl} \xrightarrow{d} N(0, \Sigma)$$

Beweis
Sei $X \in \mathbb{R}^d \setminus \{0\}$ (beachte $\langle 0, \sum\limits_{l=1}^{n} X_{nl} \rangle \xrightarrow{d} 0$ ist klar)
Und setze $Y_{nl} = \langle X, X_{nl} \rangle$. Dann gilt

- $\sum\limits_{l=1}^{n} E(Y_{nl}) = \langle X, E \left(\sum\limits_{l=1}^{n} X_{nl} \right) \rangle = 0$
- $\sum\limits_{l=1}^{n} Var(Y_{nl}) = X^T \left(\sum\limits_{l=1}^{n} Cov(X_{nl}) \right) X \to X^T \Sigma X \in (0, \infty)$

- $\sum\limits_{l=1}^n E\left(Y_{nl}^2 1_{\{>_{nl}>\varepsilon\}}\right) \overset{CSU}{\leq} \sum\limits_{l=1}^n E\left(\|X\|^2 \|X_{nl}\|^2 1_{\{|X^T X_{nl}|>\varepsilon\}}\right)$
$\overset{CSU}{\leq} \|X\|^2 \sum\limits_{l=1}^n E\Bigg(\|X_{nl}\|^2 \underbrace{1_{\{\|X\|\|X_{nl}\|>\varepsilon\}}}_{=1_{\left\{\left\|X_{nl}>\dfrac{\varepsilon}{\|X\|}\right\|\right\}}} \Bigg) \xrightarrow{n\to\infty} 0$

Somit erfüllt Y_{nl} die Voraussetzungen des eindimensionalen Lindeberg-Feller-Satzes.
Sei $Y \sim N(0,\Sigma)$, dann gilt
$\underbrace{\sum\limits_{n]l=1Y_{[nl}}^{\infty}}_{=X^T \sum\limits_{l=1}^n X_{nl}} \xrightarrow{d} X^T Y \sim N(0, X^T\Sigma X)$ Mit Cramer-Wold-Device folgt die Behauptung.

□

5.11 Bemerkung

Lässt man auch degenerierte Normalverteilungen zu (definiert über die charakteristische Funktion $z \mapsto e^{i\langle\mu,z\rangle-\frac{1}{2}\langle z,\Sigma z\rangle}$) so genügt es wegen Bemerkung 5.8 a) Σ als positiv semidefinit vorauszusetzen.

5.12 Korollar

Seien $(X_i)_{i\in\mathbb{N}}$ iid Zufallsvariablen mit $E(X_1) = \mu$ und $Cov(X_1) = \Sigma \in M_d(\mathbb{R})$ symmetrisch strikt positiv definit.
Dann gilt:
$\frac{1}{\sqrt{n}} \sum\limits_{i=1}^n X_i - \mu \xrightarrow{d} Y \sim N_d(0,\Sigma)$.
Eine weitere wichtige Aussage über das Wachstumsverhalten von iid Partialsummen üver iid Zufallsvariablen gibt der folgende Satz

5.13 Satz (Gesetz vom iterierten Logarithmus)

Sei $(X_n)_{n\in\mathbb{N}}$ eine iid Folge von Zufallsvariablen mit $E(X_1) = 0$ und $Var(X_1) = 1$.
Dann gilt

$$\limsup_{n\to\infty} \frac{\sum\limits_{i=1}^n X_i}{\sqrt{2nlog(log(n))}} = 1 \quad P-f.s.$$

$$\liminf_{n\to\infty} \frac{\sum\limits_{i=1}^n X_i}{\sqrt{2nlog(log(n))}} = -1 \quad P-f.s.$$

6 Bedingte Erwartungswerte

11.Vorlesung, 24.05.2018

Erinnerung an Elementare Wahrscheinlichkeitsrechnung und Statistik:
Seien $X : \Omega \to \mathbb{R}^n$, $Y_\Omega \to \mathbb{R}^m$ Zufallsvariablen mit gemeinsamer Wahrscheinlichkeitsdichte $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^+$.
Dann ist die bedingte Dichte
$f_{X|Y}(x,y) = \dfrac{f(x,y)}{\int\limits_{\mathbb{R}^d} f(z,y)dz}$
(eine messbare Funktion in (x,y))
und die bedingte Erwartung von X gegeben $Y = y$
$E(X|Y=y) = \int\limits_{\mathbb{R}^m} x f_{X|Y}(x,y)dx$
eine messbare Funktion von y.
Entsprechend ist die bedingte Erwartung
$E(X|Y) = \int\limits_{\mathbb{R}^m} x f_{X|Y}(x,Y)dx$
eine Zufallsvariable (messbare Funktion in ω).
$E(X|Y)$ ist offenbar sogar $\sigma(Y) - \mathcal{B}(\mathbb{R}^n)$-messbar, also eine Zufallsvariable auf $(\Omega, \sigma(Y), P|_{\sigma(Y)})$.
Sei $h : \mathbb{R}^m \to \mathbb{R}$ messbar und beschränkt, $X \in \mathcal{L}^1(P)$, dann gilt
$E(h(Y)E(X|Y)) = \int\limits_{\mathbb{R}^m} h(y) \int\limits_{\mathbb{R}^n} x \dfrac{f(x,y)}{f_Y(y)} dx f_Y(y) dy$
$= \int\limits_{\mathbb{R}^m} \int\limits_{\mathbb{R}^n} h(y) x f(x,y) dx dy = E(h(Y)|X)$

6.1 Definition und Satz

Sei $X \in \mathcal{L}^1(P)$ oder $X \geq 0$ und $\mathcal{G} \subseteq \mathcal{F} \quad \sigma-$Algebra.

(a) Dann existiert $Z : \Omega \to \overline{\mathbb{R}}$ mit

  (i) Z ist $\mathcal{G}$-Messbar

  (ii) $E(X 1_A) = E(Z 1_A) \quad \forall A \in \mathcal{G}$

  Man nennt Z die bedingte Erwartung von X gegeben $\mathcal{G}$, i.Z. $Z = E(X|\mathcal{G})$.

(b) Z ist durch 1) und 2) f.s. eindeutig bestimmt.

Seien μ, ν zwei Maße auf (Ω, F).
μ heißt absolutstetig bezüglich ν, (i.Z. $\mu \ll \nu$), wemm $\nu(A) = 0 \Rightarrow \mu(A) = 0$
μ und ν heißen äquivalent, wenn $\mu \ll \nu$ und $\nu \ll \mu$

Satz von Radon-Nikodym
Wenn $\mu \ll \nu$, dann gibt es $Z : \Omega \to \mathbb{R}^+,\ Z \geq 0$ und messbar mit $\int 1_B d\mu = \int 1_B Z d\nu \quad \forall B \in \mathcal{F}$
„ $\underbrace{\dfrac{d\mu}{d\nu}}_{\text{Radon-Nikodym-Ableitung}}$ “$= Z$ **Beweis von Satz 6.1**

(a) (i) Sei $X \gg 0$
$Q : \mathcal{G} \to \mathbb{R}^+,\ B \mapsto \int\limits_B X dP = E(1_B X)$

ein Maß auf $(\Omega, \mathcal{G})$ mit $Q \ll P|_\mathcal{G}$
Nach Radon-Nikodym existiert eine Dichte Z von Q bzgl. $P|_\mathcal{G}\mathcal{G}$ d.h. $Z \geq 0$, $\mathcal{G}$-messbar mit
$Q(B) = E_P(!_B X) = \int 1_B Z dP|_\mathcal{G} = E_P(1_B Z)$

(ii) Sei $X \in \mathcal{L}^1(P)$. Zerlege $X = X^+ - X^-$. Mit a) existieren Z^+, Z^- $\mathcal{G}$-messbar so, dass
$E(X^{+/-}) = E(Z^{+/-} 1_B) \quad \forall B \in \mathcal{G}$
Die linearität der Erwartungswerte liefert die Behauptung.

(b) Seien Z_1, Z_2 Zufallsvariablen die i), ii) erfüllen. Dann gilt
$B := \{Z_1 > Z_2\} \in \mathcal{G}$
$E)!_B(\underbrace{Z_1 - Z_2}_{>0 auf B})) \overset{ii)}{=\!=} E(1_B(X - X)) = 0$
$\Rightarrow P(B) = 0$
Analog zeigt man $P(\{Z_2 > Z_1\}) = 0 \Rightarrow P(Z_1 = Z_2) = 1$

$\square$

6.2 Satz

Seien $X, Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ Zufallsvariablen mit $X, Y \in \mathcal{L}^1$ oder $X, Y \geq 0$ und $\mathcal{G} \subseteq \mathcal{H} \subseteq \mathcal{F}$ σ-Algebren. Dann gelten:

(a) $E(\alpha X + \beta Y|\mathcal{G}) = \alpha E(X|\mathcal{G}) + \beta E(Y|\mathcal{G}) \quad \forall \alpha, \beta \geq 0$
(bzw. $\forall \alpha, \beta \in \mathbb{R}$, falls $X, Y \in \mathcal{L}^1$)

(b) $E(E(X|\mathcal{G})) = E(X)$ und $E(E(X|\mathcal{G})|\mathcal{G}) = E(X|\mathcal{G})$

(c) $E(E(X|\mathcal{H})|\mathcal{G}) = E(X|\mathcal{G})$ (tower priority)

(d) $X = Y$ f.s. $\Rightarrow E(X|\mathcal{G}) = E(Y|\mathcal{G})$ f.s. und
$X \leq Y$ f.s. $\Rightarrow E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$ f.s.

(e) Ist Y $\mathcal{G}$-messbar und $XY \in \mathcal{L}^1$, dann gilt
$E(XY|\mathcal{G}) = YE(X|\mathcal{G})$ und $E(Y|\mathcal{G}) = Y = E(Y|\underbrace{Y}_{\sigma(Y)})$

(f) Sind $\sigma(X)$ und $\mathcal{G}$ unabhängig, dann gilt
$E(X|\mathcal{G}) = E(X)$. Insbesondere $E(X) = E(X|\{\emptyset, \Omega\})$

(g) Bedingte Jensen-Ungleichung:

(i) $f : \mathbb{R} \to \mathbb{R}$ konvex, $f(X) \in \mathcal{L}^1(P) \Rightarrow f(E(X|\mathcal{G})) \leq E(f(X)|\mathcal{G})$

(ii) Ist f strikt konvex, so gilt in i) "="$\Leftrightarrow X = E(X|\mathcal{G})$

Beweis

(a) Folgt unmittelbar aus der Definition und der Linearität des Erwartungswertes

(b) Folgt aus c) und f)

(c) Sei $B \in \mathcal{G}$, dann gilt
$E(1_B \underbrace{E(E(X|\mathcal{H})|\mathcal{G})}_{\mathcal{G}-messbar}) = E(1_B E(X|\mathcal{H})) \overset{\mathcal{G}\subseteq\mathcal{H}}{=\!=\!=} E(1_B X)$

(d) Übung

(e) Sei zunächst $Y = 1_A$ mit $A \in \mathcal{G}$. Sei $B \in \mathcal{G}$.
$E(E(X1_A|\mathcal{G})1_B) = E(X1_A 1_B) = E(X1_{A\cap B}) = E(\underbrace{E(X|\mathcal{G})1_{A\cap B}}_{\mathcal{G}-messbar})$
$\Rightarrow E(X1_A|\mathcal{G} = 1_A E(X|\mathcal{G})$
Der allgemeine Fall folgt nun durch maßtheoretische Induktion.

(f) Sei $B \in \mathcal{G}$: Dann ist 1_B unabhängig von X und
$E(1_B(X) = E(X)E(1_B) = E(\underbrace{E(X)1_B}_{\mathcal{G}-messbar})$
$\Rightarrow E(X|\mathcal{G}) = E(X)$

(g) (i) Konvexe Funktionen sind stetig im Inneren ihres Definitionsbereiches. Eine konvexe Funktion ist das Supremum ihrer Subgradienten. Deshalb $\exists a_n, b_n$ für $n \in \mathbb{N}$ so, dass

$$f(x) = \sup_{n\in\mathbb{N}}(a_n x + b_n) \quad \forall x \in \mathbb{R}$$

Für jedes $n \in \mathbb{N}$ gilt
$E(f(X)|\mathcal{G}) \geq E(a_n x + b_n|\mathcal{G}) = a_n E(X|\mathcal{G} + b_n \quad f.s.$
$\Rightarrow E(f(X)|\mathcal{G}) \geq \sup n \in \mathbb{N}(a_n E(X|\mathcal{G}) + b_n) = f(E(X|\mathcal{G})) \quad f..s.$

(ii) ausgelassen

6.3 Satz (bedingte monotone Konvergenz)

12.Vorlesung, 28.05.2018

Es seien $(X_n)_{n\in\mathbb{N}_0}$ Zufallsvariablen und $\mathcal{G} \subseteq \mathcal{F}$ σ-Algebra. Dann gilt:

(a) $0 \leq X_n \nearrow X_0 \Rightarrow 0 \leq E(X_n|\mathcal{G}) \nearrow E(X_0|\mathcal{G})$

(b) $|X_n| \leq Y \in \mathcal{L}^1(P)$ und $X_n \xrightarrow[n\to\infty]{f.s.} X_0 \Rightarrow E(X_n|\mathcal{G}) \xrightarrow[n\to\infty]{f.s.} E(X_0|\mathcal{G})$

Beweis
Folgt jeweils mit geeigneter Nutzung der unbedingten Sätze.

□

6.4 Bemerkung

Sei $X \in \mathcal{L}^2$ ein „Schätzer“und $\mathcal{G} \in \mathcal{F}$. Dann gilt:

- $E(X) = E(E(X|\mathcal{G}))$

- $Var(E(X|\mathcal{G})) = E(E(X|\mathcal{G})^2) - E(X)^2 \leq E(X^2) - E(X)^2 = Var(X)$

Das heißt $E(X|\mathcal{G}) \in \mathcal{L}^2$ ist ein genauso (un)verzerrter Schätzer mit kleinerer Varianz, also effizienter.

6.5 Satz (Orthogonalprojektion)

Sei $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ ein Hilbertraum, $\mathcal{U} \subseteq \mathcal{H}$ ein abgeschlossener Unterraum und $X \in \mathcal{H}$. Dann existiert eine eindeutige Bestapproximation (Orthogonalprojektion) $X_\mathcal{U}$ von X in $\mathcal{U}$, d.h. $X_\mathcal{U}$ ist das einzige Element von $\mathcal{U}$ so, dass

$$\|X - X_\mathcal{U}\| = \inf_{U \in \mathcal{U}} \|X - U\|$$

$X_\mathcal{U}$ ist auch das einzige Element von $\mathcal{U}$ mit

$$\langle X - X_\mathcal{U}, U\rangle = 0 \quad \forall U \in \mathcal{U}$$

6.6 Satz

Sei $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ und $\mathcal{G} \subseteq \mathcal{F}$ σ-Algebra.
Dann gilt

$$E((X - E(X|\mathcal{G}))\mathcal{U}) = 0 \quad \forall \mathcal{U} \in \mathcal{L}^2(\Omega, \mathcal{G}, P|_\mathcal{G})$$

$E(X|\mathcal{G})$ ist also die Orthogonalprojektion von X auf $\mathcal{L}^2(\Omega, \mathcal{G}, P|_\mathcal{G})$.
Beweis
Jede $\mathcal{G}$-messbare Funktion ist $\mathcal{F}$-messbar und damit ist $\mathcal{L}^2(\Omega, \mathcal{G}, P|_\mathcal{G})$ ein Untervektorraum von $\mathcal{L}^2(\Omega, \mathcal{F}, P)$. Da $\mathcal{L}^2$-Räume vollständig sind, ist es ein abgeschlossener Unterraum. Sei $\mathcal{U} \in \mathcal{L}^2(\Omega, \mathcal{G}, P|_\mathcal{G})$, dann gilt

$$E((X - E(X|\mathcal{G}))\mathcal{U}) \overset{6.2e)}{=\!=\!=} E(X\mathcal{U}) - E(E(X\mathcal{U}|\mathcal{G})) \overset{6.2b)}{=\!=\!=} 0$$

□

6.7 Satz (Faktorisierungssatz)

Sei $X \geq 0$ oder $X \in \mathcal{L}^1(P)$ und $Y : \Omega \to (\Omega_Y, \mathcal{F}_Y)$ messbar (Ω_1-wertige Zufallsvariable). Dann gibt es eine messbare Funktion $f_x : (\Omega_Y, \mathcal{F}_Y) \to (\overline{\mathbb{R}}, B(\overline{\mathbb{R}}))$ mit $E(X|Y) = f_X(Y)$.
Die Funktion f_X ist P-f.s. eindeutig und

$$\int_A f dP^Y = \int_{Y^{-1}} (A) X dP \quad \forall A \in \mathcal{F}_Y$$

Man schreibt $f(y) = E(X|Y = y) \quad \forall y \in \Omega_Y$ und nennt $f(y)$ den bedingten Erwartungswert von X unter $Y = y$
Beweis
ausgelassen (auch in der Vorlesung)

□

7 Bedingte Verteilungen

Für Zufallsvariablen X, Y wollen wir sinnvoll bedingte Wahrscheinlichkeitsmaße $P(X \in \cdot | Y = y)$ erklären.

7.1 Definition

Seien $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ messbare Räume. Eine Abbildung
$K : \Omega_1 \times \mathcal{F}_2 \to [0,1]$ heißt stochastischer Kern (oder Markovkern) von Ω_1 nach Ω_2, falls

(i) $A_2 \mapsto K(\omega_1, A_2)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega_2, \mathcal{F}_2) \quad \forall \omega_1 \in \Omega_2$

(ii) $\omega_1 \mapsto k(\omega_1, A_2)$ ist messbar $\forall A_2 \in \mathcal{F}_2$.

Interpretation als zweistufiges Zufallsexperiment:

1. Ziehe $\omega_1 \in \Omega_1$ mit Wahrscheinlichkeitsmaß P_1 auf $(\Omega_1, \mathcal{F}_1)$
2. Ziehe $\omega_2 \in \Omega_2$ mit Wahrscheinlichkeitsmaß $K(\omega_1, \cdot)$ auf $(\Omega_2, \mathcal{F}_2)$

Betrachten wir als Beispiel das zweifache ziehen ohne Zurücklegen aus $\{1,2,3\}$

$$\Omega_1 = \Omega_2 = \{1,2,3\}$$

- Der erste Zug ist gleichverteilt auf Ω_1:
$$P_1 = \frac{1}{3}(\delta_1 + \delta_2 + \delta_3)$$
- Der zweite Zug ist gleichverteilt auf $\Omega_2 \setminus \{\omega_1\}$ d.h.
$$K(\omega_1, \cdot) = \frac{1}{2} \sum_{i \in \Omega_2 \setminus \{\omega_1\}} \delta_i$$

7.2 Satz

Seien $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ messbare Räume, P_1 ein Wahrscheinlichkeitsmaß auf $(\Omega_1, \mathcal{F}_1)$ und k ein stochastischer Kern. Dann wird durch

$$P_1 \otimes K(A) = \int_{\Omega_1} \int_{\Omega_2} 1_A(\omega_1, \omega_2) K(\omega_1, d\omega_2) P_1(d\omega_1)$$

für $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ ein Wahrscheinlichkeitsmaß auf $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ definiert.
Schreibweise:
$P_1 \times K(d\omega_1, d\omega_2) = \int K(\omega_1, d\omega_2) P_1(d\omega_1)$
Auf Rechteckmengen $A = A_1 \times A_2$ (bestimmen Maß eindeutig) gilt

$$P_1 \times K(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2) dP_1(\omega_1)$$

Beweis
Nachrechnen

$\square$

7.3 Satz (Fubini)

Sei $f \in \mathcal{L}^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \times K)$. Dann gilt:

(i) Die Abbildung

$$\omega_1 \mapsto \int\limits_{\Omega_2} f(\omega_1, \omega_2) K(\omega_1, d\omega_2)$$

ist messbar und P-f.s. definiert und in $\mathcal{L}^1(P_1)$

(ii) $\int\limits_{\Omega_1 \times \Omega_2} f dP_1 \times K = \int\limits_{\Omega_1} \int\limits_{\Omega_2} f(\omega_1, \omega_2) K(\omega_1, d\omega_2) dP(\omega_1)$

Beweis
Analog zum klassischen Satz von Fubini

□

Seien nun 13.Vorlesung,
$Y : (\Omega, \mathcal{F}, P) \to (\Omega_Y, \mathcal{F}_Y)$ 07.06.2018
$Z : (\Omega, \mathcal{F}, P) \to (\Omega_Z, \mathcal{F}_Z)$
Zufallsvariablen. Wir wollen die gemeinsame Verteilung $P^{(Y,Z)}$ von (Y, Z) auf $(\Omega_Y \times \Omega_Z, \mathcal{F}_Y \otimes \mathcal{F}_Z)$ beschreiben basierend auf der Randverteilung P^Y.

7.4 Definition

Eine Abbildung $K : \Omega_Y \times \mathcal{F}_Z \to \mathbb{R}$ heißt (reguläre)bedingte Verteilung von Z unter $Y = y$, falls K ein Kern ist und $P^{(Y,Z)} = P^Y \times K$, d.h. falls

(i) $y \mapsto K(y, A)$ ist messbar $\forall A \in \mathcal{F}_Z$

(ii) $A \mapsto K(y, A)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega_Z, \mathcal{F}_Z) \quad \forall y \in \Omega_y$

(iii) $P(y \in B, Z \in A) = \int\limits_B K(y, A) dP^Y(y) \quad \forall A \in \mathcal{F}_2, B \in \mathcal{F}_Y$

Wir schreiben $P^{Z|Y=y}(A)$ für $K(y, A)$.

7.5 Bemerkung

Sind Y, Z unabhängig, so ist $P^{Z|Y=y} = P^Z \quad \forall y \in \Omega_Y$

7.6 Satz

Sei $g : \Omega_Y \times \Omega_Z \to \overline{\mathbb{R}}$ eine nichtnegative messbare Funktion oder $g(y, z) \in \mathcal{L}^1(P)$.
Ferner existiere eine reguläre bedingte Verteilung $P^{Z|Y=y} = K(y, \cdot)$.
Dann gilt

$$E(g(Y, Z)|Y) = \int\limits_{\Omega_Z} g(Y, z) K(Y, dz) \tag{1}$$

Für $z \geq 0$ oder $Z \in \mathcal{L}^1$ gilt also

$$E(Z|Y) = \int\limits_{\Omega_2} g(Y,z) K(Y,dz)$$

Genauso macht es jetzt Sinn

$$E(g(Y,Z)|Y=y) = \int\limits_{Omega_Z} g(y,z) K(y,dz)$$

zu betrachten.
Beweis
Rechte Seite von (1) ist $\sigma(Y)$-messbar. Sei $B \in \sigma(Y)$
$E(1_B \int\limits_{\Omega_2} g(Y,z)K(Y,dz)) = \int\limits_{\Omega_1} \int\limits_{\Omega_2} 1_B g(y,z) K(y,dz) dP^Y(y)$
$= \int\limits_{\Omega_1} \int\limits_{\Omega_2} 1_B g(y,z) dP^{(Y,Z)}(y,z) = E(1_B g(Y,Z))$

$\square$

Es liegt nahe zu denken, dass durch

$$P^{Z|Y=y}(A) := E(1_A(Z)|Y=y) \quad \text{für } A \in \mathcal{F}_Z$$

eine reguläre bedingte Verteilung definiert wird.

DIES IST ABER FALSCH!

Nur wenn z.B. Ω_2 polnischer Raum (vollständig, metrisch und separabel) und $\mathcal{F}_Z$ die Borel-σ-Algebra (aus der Topologie der Metrik) ist, stimmt das.
$\Rightarrow$ Für $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$-wertige Zufallsvariablen existieren stets bedingte Verteilungen.

Teil 2: Statistik

8 Deskriptive Statistik

8.1 Folien

14.Vorlesung, 11.6.2018

9 Einführung in die induktive Statistik

9.1 Definition

15.Vorlesung, 14.06.2018

(i) Sei $(\Omega, \mathcal{F})$ ein messbarer Raum, θ eine Menge (Parametermenge) und

$$\mathcal{P} = \{P_\vartheta \ \text{ Wahrscheinlichkeitsmaß auf } (\Omega, \theta) : \vartheta \in \theta\}$$

eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{F})$. Dann heißt $(\Omega, \mathcal{F}, P)$ ein statistisches Experiment (SE).

(ii) Ist $\theta \subseteq \mathbb{R}^d$ für ein $d \in \mathbb{N}$, so spricht man von einem parametrischen Experiment/Modell

(iii) Gilt $P_{\vartheta_1} \neq P_{\vartheta_2} \quad \forall \vartheta_1, \vartheta_2 \in \theta$ mit $\vartheta_1 \neq \vartheta_2$, so heißt das Modell identifizierbar

9.2 Annahme

Wir nehmen stets an, dass die betrachteten SE identifizierbar sind.

9.3 Definition

Sei $(\Omega, \mathcal{F}, P)$ SE und D eine Menge (Entscheidungsraum).

(a) Eine Abbildung $g : \mathcal{P} \to D$ heißt statistisches Funktional

(b) Sei $\mathcal{D}$ eine σ-Algebra auf D. Dann heißt eine messbare Abbildung $T : (\Omega, \mathcal{F}) \to (D, \mathcal{D})$ (Punkt)schätzer

(c) Sei $D \subseteq \mathbb{R}^d$. Ein Punktschätzer T heißt erwartungstreu/unverzerrt/unbiased für ein statischtisches Funktional g, falls $T \in \mathcal{L}^1(P_\vartheta) \quad \forall \vartheta \in \theta$ und $E_\vartheta(T) = E_{P_\vartheta}(T) = g(\vartheta) = g(P_\vartheta)$

(d) Für einen integrierbaren Punktschätzer T heißt

$$Bias_\vartheta(T) =:= E_\vartheta(T) - g(\vartheta)$$

der Bias/die Verzerrung von T in ϑ.
Für (quadratintegrierbare) Schätzer heißt

$$MSE_\vartheta(T) := E_\vartheta(\|T_g(\vartheta)\|^2)$$

der mean-squared-error von T in ϑ.

9.4 Lemma

$$MSE_\vartheta(T) = Var_\vartheta(T) + Bias_\vartheta(T)^2$$

9.5 Beispiel

Sei $(\mathbb{R}^n, B(\mathbb{R}^n), \mathcal{P} = P^{\theta n})$ mit $P \in M_1(\mathbb{R}, B(\mathbb{R}))$ (= Menge aller Wahrscheinlichkeitsmaße auf $(\mathbb{R}, B(\mathbb{R}))$) ein n-faches Produktmodell und $g : \mathcal{P} \to M_1(\mathbb{R}, B(\mathbb{R})) =: D$
$P^{\otimes n} \mapsto P$.
Dann ist das empirische Maß $T(x_1, ..., x_n) := \frac{1}{n} \sum_{i=1}^{n} \delta_{x-i}$ ein Schätzer für g. (Wir verzichten absichtlich auf Details zu $\mathcal{D}$ und zur Messbarkeit)

9.6 Definition

Ein SE $(\Omega, \mathcal{F}, P)$ heißt dominiert, falls ein σ-endliches Maß auf $(\Omega, \mathcal{F})$ existiert mit $P \ll \mu \quad \forall P \in \mathcal{P}$ (i.Z. $\mathcal{P} \ll \mu$).
In einem dominierten Modell können alle P_ϑ durch Dichten bezüglich μ beschrieben werden

9.7 Bemerkung

(i) μ kann o.B.d.A. als Wahrscheinlichkeitsmaß angenommen werden. Sonst wähle $E_i \nearrow \Omega$ mit endlichem Maß und betrachte

$$A \mapsto \tilde{\mu}(A) := \sum_{i=1}^{\infty} \frac{\mu(A \cap E_i)}{2^i \mu(E_i)}$$

(ii) Man kann sogar zeigen:
Ist $\mathcal{P} \ll \mu, \mu$ σ-endlich, so existiert eine Folge $(P_n)_{n\in\mathbb{N}}$ in $\mathcal{P}$ mit

$$\mathcal{P} \ll \sum_{n=1}^{\infty} \frac{P_n}{2^n}$$

(iii) Gilt für $\mathcal{P}_i := \{P_{i,\vartheta} : \vartheta \in \theta_i\}$ f+r i=1,2
$\mathcal{P}_i \ll \mu_i$ mit μ_i σ-endlich
So gilt auch

$$\mathcal{P}_1 \otimes \mathcal{P}_2 := \{P_{1,\vartheta_1} \otimes P_{2,\vartheta_2} : \vartheta \in \theta_i\} \ll \mu_1 \otimes \mu_2$$

(iv) Sei $Y = a + Z,\ a \in \mathbb{R}$ und $Z \sim P \in M_1(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ fest.
Sei $\mathcal{P} := \{P * \delta_a : a \in \mathbb{R}\}$
Dann gilt:
$\mathcal{P} \ll \mu,\ \mu\ \sigma - endlich \Leftrightarrow \mathcal{P} \ll \lambda$
Beweis
„$\Leftarrow$“ $P * \delta_a \ll \lambda * \delta_a = \lambda$ aufgrund der Translationsinvarianz. Setze $\mu = \lambda$
„$\Rightarrow$“o.B.d.A μ Wahrscheinlichkeitsmaß. Sei $A \in \mathcal{B}(\mathbb{R})$ mit
$0 = \mu * \lambda(A) = \int_{\mathbb{R}} \underbrace{\mu(A-y)}_{\geq 0} \lambda(dy)$
$\Rightarrow \exists y_0 \in \mathbb{R}$ mit $\mu(A - y_0) = 0 \quad \lambda-$f.ü.
$\xRightarrow{\mathcal{P}\ll\mu} 0 = P * \delta_{a-y_0}(A - y_0) = P \times \delta_a(A) = 0 \quad \forall a$
$\Rightarrow \mathcal{P} \ll \mu * \lambda$
$\mu * \lambda(A) = \int \lambda(A-y) d\mu(y) = \lambda(A) \underbrace{\int d\mu(y)}_{=1} = \lambda(A)$

10 Suffizienz

$X_1, ..., X_n$ seien Realisierungen unabhängiger Versuchswiederholungen nach einer unbekannten Verteilung P auf (Ω, F).

n-faches Produktmodell: $\mathcal{P} = \left\{ P_\vartheta^{\otimes n} : \vartheta \in \theta \right\}$.
Häufig besitzen Teile der Daten $X_1, ..., X_n$ keine zusätzliche Information über das wahre ϑ!
Ziel:

- Transformation der Daten ohne Informationsverlust
- Datenkompression ohne Informationsverlust

Sei zum Beispiel
$T : (\Omega^n, \mathcal{F}^{\otimes n}) \to (\Omega_T, \mathcal{F}_T)$ (polnisch) eine Transformation.
Wenn die bedingte Verteilung ($P \in \mathcal{P}$ unbekannt)

$$(P_\vartheta^{\otimes n})^{id|T=t}$$

für alle t unabhängig von $\vartheta \in \theta$ wählbar ist, sollte $T(X_1, ..., X_n)$ alle Informationen der Daten über θ enthalten.

10.1 Beispiele

(i) Betrachte $(\{0,1\}^n, \mathcal{P}(\{0,1\}^n), Binom(1, \vartheta)^{\otimes n})$.
Dann sollte die Komprimierung
$T : (x_1, ..., x_n) \mapsto \sum_{i=1}^{n} x_i$ ohne Informationsverlust sein.
Es gilt $\forall 0 \leq k \leq n$ und $\forall \{x_1, ..., x_n\} + \{0,1\}^n$ mit $\sum_{i=1}^{n} X_i = K$:

$$P_\vartheta^n((x_1, ..., x_n)|T = K) = \frac{\prod_{i=1}^{n} \vartheta^{x_i}(1-\vartheta)^{1-x_i}}{\binom{n}{K} \vartheta^K (1-\vartheta)^{n-K}} = \binom{n}{K}^{-1}$$

Was nicht von ϑ abhängt.

(ii) Gegeben sei ein Produktmodell $(\Omega^n, \mathcal{F}^{\otimes n}, (P_\vartheta^{\otimes n})_{\vartheta \in \theta})$
Dann sollte die Reihenfolge der Beobachtungen irrelevant sein.
Insbesondere sollten die Ordnungsstatistiken dieselben Informationen beinhalten wie die Daten. In der Tat kann man zeigen (siehe Skript Pauly)
$P_\vartheta^{\otimes n\ id|(x_{(1)}, ..., x_{(n)}) = (y_1, ..., y_n)}$
hängt nicht von ϑ ab.
Da bedingte Verteilungen nicht immer existieren müssen, nutzen wir für die allgemeine Definition bedingte Erwartungswerte.

10.2 Definition

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ ein statistisches Experiment.

(a) Eine Teil-σ-Algebra $\tau \subseteq \mathcal{F}$ heißt suffizient für $\mathcal{P}$, (bzw. ϑ), falls für alle $A \subset \mathcal{F}$ eine von ϑ unabhängige Version $E_\bullet(1_a|\tau)$ des bedingten Erwartungswertes $E_{P_\vartheta}(1_A|\tau)$ existiert

(b) Sei $T : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ eine Statistik.
T heißt suffizient für $\mathcal{P}$, falls $\tau = T^{-1}(\mathcal{F}')(= \sigma(T))$ suffizient ist.

Interpretation: Wir werden sehen:

1. Bei der Suffizienz enhält das SE $(\Omega, \tau, \left\{P_{\vartheta|\tau}\right\}_{\vartheta\in\theta})$ genau soviel Information über ϑ wie das Ausgangsexperiment
2. Die Reduktion der Daten wird einen Effizienzgewinn (niedrigere Varianz) mit sich bringen

d.h. Suffizienz ist sozusagen „Datenredutkion ohne Informationsverlust “.

10.3 Lemma

Sei X eine Statistik auf $(\Omega, \mathcal{F})$, $X \in \mathcal{L}^1(P_\vartheta) \quad \forall\vartheta \in \theta$ und $\tau \subseteq \mathcal{F}$ suffizient.
Dann existiert eine Version $E_\bullet(X|\tau)$ von $E_{P_\vartheta}(X|\tau)$ die unabhängig von ϑ ist.
Beweis
Offensichtlich für $X = 1_A \quad \forall A \in \mathcal{F}$. Dann maßtheoretische Induktion.

$\square$

10.4 Bemerkung (Extremfälle)

- Ist $T^{-1}(\mathcal{F}') = \mathcal{F}$, so ist T suffizient.
- $\tau = \{\emptyset, \Omega\}$: Suffizienz hieße $E_\vartheta(1_A|\tau) = P_\vartheta(A) \quad \forall A \in \mathcal{F}$ wäre unabhängig von ϑ. Das ist nur in trivialen Fällen möglich. $\Rightarrow$ suffiziente σ-Algebren/Statistiken können nicht „zu klein“ sein

10.5 Lemma

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ ein statistisches Experiment.

(a) T ist suffizient $\Leftrightarrow$ $\forall A \in \mathcal{F}$ existiert eine von P_ϑ unabhängige Version $E_\bullet(1_A|T = \cdot)$ von $E_{P_\vartheta}(1_A|T = \cdot)$.

(b) Existiert eine von $P_\vartheta \in \mathcal{P}$ unabhängige Version
$K(t, \cdot) = P^{id|T=t}$
so ist T suffizient für $\mathcal{P}$.

10.6 Satz (Rao-Blackwell)

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ ein SE und S ein erwartungstreuer Schätzer für $g(\vartheta)$. Weiter sei τ suffizient und $h := E_\bullet(S|\tau)$. Dann gilt:

(i) h ist τ-messbarer erwartungstreuer Schätzer für $g(\nu)$

(ii) $MSE_\vartheta(h) = Var_\vartheta(h) \leq Var_\vartheta(S) = MSE_\vartheta(S) \quad \forall \vartheta \in \theta$ mit "=", falls $E_\vartheta(S|\tau) = SP_\vartheta - f.s.\forall\vartheta$
Beweis
Bedingte Jensensche Ungleichung

□

10.7 Satz (Neyman-Kriterium)

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ ein statistisches Experiment, $\mathcal{P} \ll \mu$ mit μ σ-endlich. Für eine Statistik $T : \Omega \to \Omega'$ sind äquivalent:

(a) T ist $\mathcal{P}$-suffizient

(b) $\exists \mathcal{F}'$- messbare Funktionen $g_p : \Omega' \to [0, \infty)$ und eine $\mathcal{F}$-messbare Funktion $h : \Omega \to [0, \infty)$ mit

$$\frac{dP}{d\mu} = (g_p \circ T) \cdot h \quad \forall P \in \mathcal{P} \ (\mu - \text{f.ü.})$$

10.8 Bemerkung

16.Vorlesung, 18.06.2018

Es genügt in b) $\frac{dP}{d\mu} = g_P \circ T$ zu fordern (ersetzte μ durch $\widehat{\mu}$ mit $\widehat{\mu}(A) = \int\limits_A h d\mu$
Beweis
Wir zeigen nur den diskreten Fall.
„$\Rightarrow$ "$\forall x \in \Omega,\ \forall P \in \mathcal{P}\ \exists$ Funktionen f_x mit

$$P(\{x\}\,|T = t) = f_x(t) \quad P^T - f.s. \quad \text{(d.h. auf P(T=t)>0)}$$

Setze $h(x) = f_x(T(x))$ und $g_P(t) = P(T = t)$, so folgt
$P(\{x\}) = P(\{x\} \cap \{T = T(x)\}) = P(\{x\}\,|T = T(x))P(T = T(x))$
$= h(x)g_P(T(x))$
(Gilt offensichtlich auch für $P(T = T(x)) = 0$.
„$\Leftarrow$ "Es gilt

$$P(\{x\}\,|T = t) = \frac{P(\{x\} \cap \{T = t\})}{P(T = t)} \quad P^T - f.s \tag{10.1}$$

$$P(\{x\} \cap \{T = t\}) = \begin{cases} 0, & \text{falls } T(x) \neq t \\ P(\{x\}) & \text{sonst} \end{cases} \tag{10.2}$$

Sei also $T(x) = t$ (sonst klar), dann gilt

$$P(T = t) = \sum_{z:T(z)=t} P(\{z\}) \tag{10.3}$$

Einsetzen von b) in (10.2) und (10.3) liefert für (10.1)

$$P(\{x\}\,|T = t) = \frac{g_P(T(x))h(x)}{\sum\limits_{z:T(z)=t} g_P(T(z))h(z)} = \frac{h(x)}{\sum\limits_{z:T(z)=t} h(z)}$$

was von P unabhängig ist.

$\square$

10.9 Korollar

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ SE, $\mathcal{P} \ll \mu,\ \mu$ σ-endlich und $\tau \subseteq \mathcal{F}$ σ-Algebra. Dann sind äquivalent:

(a) τ ist $\mathcal{P}$-suffizient

(b) $\forall P \in \mathcal{P}\ \exists \tau$-messbare Funkion $f_P : \Omega \to \mathbb{R}$ und eine von $P \in \mathcal{P}$ unabhängige $\mathcal{F}$-messbare Abbildung $h : \Omega \to \mathbb{R}$ mit

$$\frac{dP}{d\mu} = f_P \cdot \quad \mu - f.s.$$

10.10 Korollar

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ SE, $\mathcal{P} \ll \mu, \mu$ σ-endlich und τ eine suffiziente σ-Algebra. Dann ist jede σ-Algebra τ' mit $\tau \subseteq \tau' \subseteq \mathcal{F}$ ebenfalls suffizient.
Beweis
In Korollar 10.9 ist f_p τ- und damit τ'-mmessbar.

$\square$

10.11 Korollar

Seien $(\Omega_i, \mathcal{F}_i, \mathcal{P}_i), i = 1, 2$ dominierte statistische Experimente und $\tau_i \subseteq \mathcal{F}_i$ seien suffiziente σ-Algebren.
Dann ist $\tau_1 \otimes \tau_2 = \sigma(\tau_1 \times \tau_2)$ suffizient für
$\mathcal{P}_1 \otimes \mathcal{P}_2 := \{P_1 \otimes P_2 : P_i \in \mathcal{P}_i, i = 1, 2\}$.
Beweis

$$\frac{dP_1 \otimes P_2}{d\mu_1 \otimes \mu_2}(x, y) = \frac{dP_1}{d\mu_1}(x)\frac{dP_2}{d\mu_2}(y) = \underbrace{\overbrace{f_{P_1}(x)}^{\tau_1 - messbar}\ \overbrace{f_{P_2}(y)}^{\tau_2 - messbar}}_{\tau_1 \otimes \tau_2 - messbar} \cdot \underbrace{h_1(x)h_2(y)}_{h(x,y)}$$

Mit Korollar 10.9 folgt die Behauptung

$\square$

10.12 Korollar

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ dominierte SE und $T : \Omega \to \Omega',\ f : \Omega' \to \Omega''$ Statistiken.
Ist $T' = f \circ t$ suffizient, so ist T suffizient.
Beweis
Folgt aus Korollar 10.11, da $\sigma(T) \supseteq \sigma(f \circ T)$.

$\square$

10.13 Definition

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ ein SE. $\mathcal{P} = \{P_\vartheta : \vartheta \in \theta\}$ heißt Exponentialfamilie, falls μ σ-endlich existiert mit $\mathcal{P} \ll \mu$ und es Funktionen $g_1, ..., g_k : \theta \to \mathbb{R}$ für $k \in \mathbb{N}$ sowie Statistiken $T_1, ..., T_k : \Omega \to \mathbb{R}$ gibt mit

$$\frac{dP_\vartheta}{d\mu} = C(\vartheta) \exp\left(\sum_{i=1}^{k} g_i(\vartheta) T_i(x)\right) h(x) \tag{10.4}$$

wobei $h : \Omega \to \mathbb{R}$ messbar und

$$C(\vartheta) = \left(\int_\Omega \exp\left(\sum_{i=1}^{k} g_i(\vartheta) T_i(x)\right) h(x) d\mu(x)\right)^{-1}$$

10.14 Satz

Für eine Exponentialfamilie $\mathcal{P}$ wie in (10.4) ist $(T_1, T_2, ..., T_k)$ eine suffiziente Statistik.
Beweis
Nutze Theorem 10.7 mit
$g_P : \mathbb{R}^k \to \mathbb{R}$
$(x_1, ..., x_n) \mapsto \exp\left(\sum_{i=1}^{k} g_i(\vartheta) x_i\right)$

$\square$

10.15 Beispiel (Suffizienz und Exponentialfamilien)

(a) $\theta \in (0,1), P_\vartheta = Binom(n, \vartheta), \mu = \sum_{n \in \mathbb{N}_0} \delta_n$ (Zählmaß)
Dann gilt

$$\begin{aligned} P_\vartheta(\{x\}) &= \binom{n}{k} \vartheta^x (1-\vartheta)^{n-x} 1_{\{0,\ldots,n\}}(x) &(2)\\ &= \underbrace{(1-\vartheta)^n}_{C(\vartheta)} \exp\left(\underbrace{x}_{T_1(x)} \underbrace{log \frac{\vartheta}{1-\vartheta}}_{g_1(\vartheta)}\right) \underbrace{\binom{n}{x} 1_{\{0,\ldots,n\}}(x)}_{h(x)} &(3)\end{aligned}$$

Somit liegt eine Exponentialfamilie vor und $T(x) = x$ ist suffizient für $\mathcal{P} = \{Binom(n, \vartheta : \vartheta \in (0,1)\}$

(b) Offensichtlich sind Produkte von Exponentialfamilien wieder Exponentialfamilien, z.B. betrachte a) mit $n = 1, \theta = (0,1)$, und davon das N-fache Produktexperiment. Dann gilt: $P_\vartheta = Binom(1, \vartheta)^{\otimes N}$
$P_\vartheta(\{x\}) = \vartheta^{\sum_{i=1}^{N} x_i} (1-\vartheta)^{N - \sum_{i=1}^{N} x_i}$

$$= \underbrace{(1-\vartheta)}_{C(\vartheta)}^N \exp\left(\underbrace{\sum_{i=1}^N x_i}_{T_1(x)} \underbrace{log\left(\frac{\vartheta}{1-\vartheta}\right)}_{g_1(\vartheta)}\right) \underbrace{1_{\{0,1\}}(x)}_{h(x)}$$

Somit ist $T(x) = \sum_{i=1}^N x_i$ suffizient.

(c) Sei $\vartheta = (\mu, \sigma^2) \in \theta = \mathbb{R} \times \mathbb{R}^+ \setminus \{0\}$, $\mathcal{P}_\vartheta = N(\mu, \sigma^2)^{\otimes n}$ und $\mu = \lambda^n$. Dann gilt

$$\frac{dP_\vartheta}{d\mu}(x_1, ..., x_n) = \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{n\mu^2}{2\sigma^2}}}_{C(\vartheta)} \exp bigg(\underbrace{\frac{\mu}{\sigma^2}}_{g_1(\vartheta)} \underbrace{\sum_{i=1}^N x_i}_{h(x)} - \underbrace{\frac{1}{2\varsigma^2}}_{g_2(\vartheta)} \underbrace{\sum_{i=1}^N x_i^2}_{T_2(i)} \Big)$$

Also liegt eine Exponentialfamilie mit $k = 2$ und suffiziente Statistik

$$T(x) = (T_1(x), T_2(x)) = \left(\sum_{i=1}^N x_i, \sum_{i=1}^N x_i^2\right)$$

vor. 17.Vorlesung, 21.06.2018

Leicht sieht man, dass eine messbare Abbildung $f : \mathbb{R}^2 \to \mathbb{R}^2$ existiert, so dass

$$f\left(\overline{x_n}, \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x_n})^2\right) = T(x)$$

Aus Korollar 10.12 folgt somit, dass $\left(\overline{x_n}, \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x_n})^2\right)$(ebenfalls suffizient für $\mathcal{P} = \left\{N(\mu, \sigma^2)^{\otimes n} : \mu \in \mathbb{R}, \sigma^2 > 0\right\}$ ist.

10.16 Beispiel (Suffizienz ohne Exponentialfamilie)

Sei $\mathcal{U}_{(a,b)} = \frac{\lambda|_{(a,b)}}{b-a}$ die Gleichverteilung auf $(a, b), b > a$. Sei $\mathcal{P} = \left\{\mathcal{U}_{(a,b)}^{\otimes n} : \vartheta = (a, b) \in \mathbb{R}^2, a < b\right\}$. Aus Definition 10.13 folgt sofort, dass in einer Exponentialfamilie gilt

$$P_{\vartheta_1} \sim P_{\vartheta_2} \forall \vartheta_1, \vartheta_2 \in \theta$$

Somit ist $\mathcal{P}$ sicher keine Exponentialfamilie. Aber $T(x) = \left(\min_{i=1,...,n} x_i, \max_{i=1,...,n} x_i\right)$ ist suffizient, denn:

$$\frac{dP_\vartheta}{d\lambda^n}(x) = \frac{1}{(b-a)^n} \prod_{i=1}^n 1_{(a,b)}(x_i) = \frac{1}{(b-a)^n} 1_{(a,\infty)}(\min x_i) 1_{(-\infty,b)}(\max x_i)$$

womit das Neyman-Kriterium erfüllt ist.

11 Vollständigkeit und UMVU Schätzer

Der Satz von Rao-Blackwell bietet ein Verfahren zur Verbesserung erwartungstreuer Schätzer durch Nutzung sufizienter σ-Algebren an. Wir werden nun versuchen, optimale Schätzer zu erhalten.

11.1 Definition

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ ein SE.

(a) Ein statistisches Funktional $g : \mathcal{P} \to \mathbb{R}$ heißt <u>erwartungstreu schätzbar</u>, falls eine erwartungstreuer Punktschätzer für g existiert.

(b) Sei g erwartungstreu schätzbar. Dann heißt ein erwartungstreuer Schätzer $h^* : \Omega \to \mathbb{R}$ <u>gleichmäßig bester erwartungstreuer (UMVU) Schätzer</u> für g, falls $Var_\vartheta(h^*) < \infty \forall \vartheta \in \theta$ und

$$Var_\vartheta(h^*) = \min_{\substack{\text{h erw.treu} \\ \text{für g}}} Var_\vartheta(h) \quad \forall \vartheta \in \Theta$$

gilt.

11.2 Bemerkung

(i) Statistische Funktionale müssen nicht erwartungstreu schätzbar sein.

(ii) Konvexkombinationen erwartungstreuer Schätzer sind erwartungstreu

(iii) Die Varianz kann man prinzipiell in der folgenden Theorie durch sogenannte Verlustfunktionen ersetzen

11.3 Satz (Eindeutigkeit)

Seien h_1, h_2 UMVUE für g. Dann gilt

$$P_\vartheta(h_1 \neq h_2) = 0 \quad \forall \vartheta \in \theta$$

Beweis
Siehe elementare Wahrscheinlichkeitsrechnung und Statistik

$\square$

11.4 Satz (Raosche Kovarianzmethode)

Sei h^* erwartungstreu für g und $Var_\vartheta(h^*) < \infty \quad \forall \vartheta \in \theta$. Dann sind äquivalent:

(a) h^* ist UMVUE

(b) $\forall$ Nullschätzer d, d.h. Schätzer mit $E_\vartheta(d) = 0$ und $Var_\vartheta(d) < \infty \quad \forall\vartheta$, gilt $Cov(h^*, d) = 0 \quad \forall\vartheta$

Beweis

„$\Leftarrow$"

Sei h erwartungstreu mit $Var_\vartheta(h) < \infty$. Dann $d = h - h^*$ Nullschätzer und
$Var_\vartheta(h) = Var_\vartheta(d + h*) = \underbrace{Var_\vartheta(d)}_{\geq 0} + Var_\vartheta(h^*) + 2\underbrace{Cov(h^*, d)}_{=0} \geq Var_\vartheta(h^*)$

$\Rightarrow h^*$ ist UMVUE.

„$\Rightarrow$ "

Offenbar ist $d_t := h^* + td$ erwartungstreu $\forall f \in \mathbb{R}$.
Somit $Var_\vartheta(h^*) \leq Var_\vartheta(d_t) = Var_\vartheta(h^*) + t^2 Var_\vartheta(d) + 2tCov_\vartheta(d, h^*)$
Die rechte Seite hat ein Minimum in $t = 0$ (und ist ≥ 0)
$\Rightarrow Cov_\vartheta(d, h^*) = 0$

$\square$

11.5 Definition (Vollständige σ-Algebra)

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ SE.

(a) Eine σ-Algebra $\tau \subseteq \mathcal{F}$ heißt vollständig (für $\mathcal{P}$), falls gilt:
$\forall f : \Omega \to \mathbb{R}, \tau - messbar, f \in \mathcal{L}^1 P_\vartheta) \quad \forall\vartheta$ mit
$E\vartheta(f) = 0 \quad \forall\vartheta$ ist

$$f = 0 \quad P_\vartheta|_\tau - f.s. \quad \forall\vartheta$$

(b) Eine Statistik $T : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ heißt vollständig, falls $T^{-1}(\mathcal{F}') = \sigma(T)$ vollständig ist.

(c) Eine σ-Algebra $\tau \subseteq \mathcal{F}$ (eine Statistik T) heißt beschränkt vollständig, falls (11.1) für alle beschränkten messbaren f gilt.

Vollständigkeit bedeutet also, dass alle τ-messbaren Nullschätzer trivial (also f.s. identisch Null) sind. Vollständige σ-Algebren sind also eher klein.

11.6 Bemerkung

(i) $T : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ ist genau dann vollständig, wenn $\forall h : \Omega' \to \mathbb{R}$ messbar mit $h \circ T \in \mathcal{L}^1(P_\vartheta)$ und $E_{P_\vartheta}(h \circ T) = 0 \quad \forall\vartheta \in \theta$ folgt $h \circ T \equiv 0 \quad P_\vartheta - f.s.$

(ii) Ein Resultat Bahadour:
Ist T suffizient und beschränkt vollständig für $\mathcal{P}$, so existiert für alle suffizienten S eine messbare Abbildung φ mit $T = \varphi \circ S$.
Somit gilt $\sigma(T) \subseteq \sigma(S)$.
Man spricht von Minimalsuffizienz für T und T ergibt sozusagen eine maximal mögliche Datenkompression ohne Informationsverlust.

11.7 Beispiel

(a) Sei $\mathcal{P} = \{P \in M_1(\mathbb{R}, \mathcal{B}(\mathbb{R})) : P(A) = P(-A) \quad \forall A \in \mathcal{B}(\mathbb{R})\}$.
Dann ist $\mathcal{B}(\mathbb{R})$ nicht vollständig für $\mathcal{P}$, denn

$$\int_{\mathbb{R}} \underbrace{1_{(0,\infty)} - 1_{(-\infty,0)}}_{\neq 0} dP = 0 \quad \forall P \in \mathcal{P}$$

(b) Sei $\mathcal{P} = \{Binom(1,\vartheta)^{\otimes n} : \vartheta \in (0,1)\}$ und
$T : \{0,1\}^n \to \{0,...,n\}$ mit $(x_1,...,x_n) \mapsto \sum_{i=1}^{n} x_i$. Dann ist T vollständig für $\mathcal{P}$.

Beweis
Betrachte $h : \{0,1,...,n\} \to \mathbb{R}$ mit $E_{P_\vartheta}(h \circ T) = 0 \quad \forall \vartheta$

$$\Rightarrow 0 = \sum_{k=0}^{n} h(k) \binom{n}{k} \vartheta^k (1-\vartheta)^{(n-k)}$$

$$= (1-\vartheta)^n \sum_{k=0}^{n} h(k) \Bigg(\underbrace{\frac{\vartheta}{1-\vartheta}}_{=:\varsigma \in (0,\infty)} \Bigg)^k$$

$$\Rightarrow \sum_{k=0}^{n} h(k) \binom{n}{k} \varsigma^k = 0 \quad \forall \varsigma \in (0,\infty)$$

$\Rightarrow$ Dies muss das Nullpolynom in ς sein
$\Rightarrow h(k) = 0 \quad \forall k \in \mathbb{N}_0$

11.8 Satz (Lehmann-Scheffe)

18.Vorlesung, 28.06.2018

Sei $g : \mathcal{P} \to \mathbb{R}$ erwartungstreu schätzbar und h ein erwartungstreuer Schätzer für g mit $Var_\vartheta(h) < \infty \quad \forall \vartheta$. Dann gilt

(a) Ist $\tau \subseteq \mathcal{F}$ suffizient und vollständig für $\mathcal{P}$, dann ist

$$h^* := E_\bullet(h|z)$$

der (f.s. eindeutige) UMVUE für g.

(b) Ist $T : \Omega \to \Omega'$ eine suffiziente und vollständige Statistik, dann ist $f^* \circ T$ mit $f^*(t) = E_\bullet(H|T=t)$ der (f.s. eindeutige) UMVUE für g.

Beweis
b) folgt offensichtlich aus a)
Ad a)
Offensichtlich ist h^* unverzerrt. Sei $\tilde{h}$ erwartungstreu mit $Var_\vartheta(\tilde{h}) < \infty$. Mit Rao-Blackwell gilt

$$Var_\vartheta(\tilde{h}) \geq Var_\vartheta(E_\bullet(\tilde{h}|\tau)) \quad \forall \vartheta \tag{$*$}$$

Aus der Erwartungstreue folgt

$$E_\vartheta(\underbrace{h^* - E_\bullet(\tilde{h}|z)}_{\tau-messbar}) = g(\vartheta) - g(\vartheta) = 0$$

Da τ vollständig ist, folgt $h^* = E_\bullet(\tilde{h}|\tau) \quad P_\vartheta - f.s. \quad \forall\vartheta$. Nun folgt aus $(*)$ die Varianzminimalität. Die Eindeutigkeit folgt aus Satz 11.3.

□

11.9 Beispiel

Aus den Beispielen 11.7 b) und 10.15 b) folgt, dass $T(x) = \sum_{i=1}^{n} x_i$ suffizient und vollständig ist für $\{Binom(1,\vartheta)^{\otimes n} : \vartheta \in (0,1)\}$
Erwartungstreuer Schätzer ist $\widehat{\vartheta}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$. Dieser ist offensichtlich $\sigma(T)$-messbar. Mit Lehmann-Scheffe ist $\widehat{\vartheta}$ also UMVUE für ϑ.

Hat man eine suffiziente und vollständige Statistik T, so kann man folgende zwei Möglichkeiten nutzen, um einen UMVUE zu finden:

(i) Finde einen beliebigen erwartungstreuen Schätzer h und berechne $h^* = E_\bullet(h|T)$

(ii) Finde eine messbare Funktion f mit

$$E_\vartheta(f \circ T) = g(\vartheta) \quad \forall\vartheta \in \theta$$

11.10 Beispiel

Sei $\mathcal{P} = \left\{P_b^{\otimes n} : P_b = \mathcal{U}_{(0,b)}, b > 0\right\}$. Dann ist $T(x) = \max\limits_{1\le i\le n} x_i$ suffizient und vollständig.
Beweis
Suffizienz analog zu Beispiel 10.16.
Es gilt $P_b^{\otimes n}(T \le t) = P(x_1 \le t)^n = \left(\frac{t}{n}\right)^b \quad \forall 0 \le t \le b$
Sei nun $b > 0$ und f so, dass $f \circ T \in \mathcal{L}^1(P_b) \quad \forall b$ und
$0 = E_b(f \circ T) = \int\limits_{\mathbb{R}} f(t) d(P_b^{\otimes n})^T = \int\limits_0^b f(t)\frac{nt^{n-1}}{b^n}dt$
Zerlege $f = f^+ - f^-$ in Positiv- und Negativteil:

$$\int\limits_0^b f^+(t)t^{n-1}dt = \int\limits_0^b f^-(t)t^{n-1}dt \quad \forall b > 0$$

Betrachte die Maße μ^+ und μ^- auf $B((0,\infty))$ mit
$\mu^+(B) = \int\limits_B f^+(t)t^{n-1}dt,\ \mu^-(B) = \int\limits_B f^-(t)t^{n-1}dt.$
Mit dem Eindeutigkeitssatz für Maße gilt $\mu^+ = \mu^-$ und beides sind σ-endliche Maße.
Aus Radon-Nikodym (f.ü. Eindeutigkeit der Dichte) folgt
$\frac{d\mu^+}{d\lambda} = \frac{d\mu^-}{d\lambda} \quad \lambda -$ f.ü.

$\Rightarrow f^+ = f^- \quad \lambda$-f.ü. $\Rightarrow f = 0 \quad \lambda$-f.ü. und damit $(P_b^{\otimes n})^T - f.s. \quad \forall b > 0$.
Mit Bemerkung 11.6 (i) folgt die Behauptung.

□

$$h(x) = \frac{n+1}{n} \max_{1\leq i\leq n} x_i \text{ ist UMVUE, da}$$

$$E_\vartheta(h(x)) = \frac{n+1}{n} \int_0^b t \frac{nt^{n-1}}{b^n} dt = \left.\frac{t^{n+1}}{b^n}\right|_0^b = b$$

□

Wir wenden uns nun wieder Exponentialfamilien zu, Sei

$$\mathcal{P} = \{P_\vartheta : \vartheta \in \theta\} \ll \mu \quad (\mu - \sigma - endlich)$$

von der Form

$$\frac{dP_\vartheta}{d\mu}(x) = C(\vartheta) \exp\left(\sum_{i=1}^k g_i(\vartheta) T_i(x)\right) h(x)$$

Dann heißt die Menge

$$\Xi^* := \left\{ s \in \mathbb{R}^k : \int_\Omega exp\left(\sum_{i=1}^k s_i T_i)\right) d\mu \right\}$$

natürlicher Parameterraum der Exponentialfamilie. Offensichtlich gelten:

- $\Xi := \{(g_1(\vartheta), ..., g_n(\vartheta)) : \vartheta \in \theta\} \subseteq \Xi^*$
- $\frac{dQ_s}{d\mu} := \tilde{C}(s) \exp\left(\sum_{i=1}^k s_i T_i(x)\right) h(x)$ definiert eine Exponentialfamilie $\{Q_s : s \in \Xi^*\}$ in natürlicher Parametrisierung mit

$$\tilde{C}(s) = \frac{1}{\int_\Omega \exp\left(\sum_{i=1}^k s_i T_i(x)\right) h(x) \mu(dx)}$$

11.11 Lemma (Momente bei Exponentialfamilien)

Sei $(\Omega, \mathcal{F}, \mathcal{P})$ SE mit $\mathcal{P}$ Exponentialfamilie in natürlicher Parametrisierung.

(a) Ist $s \in \Xi^*$ so existieren unter Q_s für $T_1, ..., T_k$ Momente beliebiger Ordnung, d.h. $\forall (\mu_1, ..., \mu_n) \in (\mathbb{N}_0)^k$ gilt $\prod_{j=1}^k T_j^{\mu_j} \in \mathcal{L}^1(Q_s)$ und

$$\underbrace{\int_\Omega \prod_{j=1}^k T_j^{\mu_j} dQ_s}_{E_{Q_s}(T|...)} = \tilde{C}(s) \frac{\partial^{\mu_1 + ... + \mu_k}}{\partial s_1^{\mu_1} \partial s_2^{\mu_2} ... \partial s_k^{\mu_k}} \int_\Omega e^{\langle s, T\rangle} h d\mu$$

(b) Ist $\varphi : \Omega \to \mathbb{R}$ messbar und beschränkt, so gilt

$$\frac{\partial}{\partial s_i} E_{Q_s}(\varphi) = E_{Q_s}(\varphi T_i) - E_{Q_s}(\varphi) E_{Q_s}(T_i) \tag{4}$$

$$= Cov_{Q_s}(\varphi, T_i) \tag{5}$$

(c) $\frac{\partial}{\partial_{s_i}} \tilde{C}(s) = -\tilde{C}(s) E_{Q_s}(T_i)$

Beweis