Stochastik II

Vorlesung von Prof. Grübel
LaTeX-Version von Dirk Hillbrecht

Wintersemester 1994/95

Inhaltsverzeichnis

Teil III

Allgemeine Wahrscheinlichkeitsräume

11. Oktober 1994

10 Mengensysteme

Im folgenden: $\omega \neq \emptyset$.

Definition 10.1
(a) $\mathfrak{R} \subset \wp(\omega)$ heißt *Ring*, wenn:

(i) $\emptyset \in \mathfrak{R}$;

(ii) $A, B \in \mathfrak{R} \implies A \setminus B \in \mathfrak{R}$;

(iii) $A, B \in \mathfrak{R} \implies A \cup B \in \mathfrak{R}$.

(b) $\mathfrak{D} \subset \wp(\omega)$ heißt *Dynkin-System*, wenn

(i) $\emptyset \in \mathfrak{D}$;

(ii) $A \in \mathfrak{D} \implies A^c \in \mathfrak{D}$;

(iii) $A_1, A_2, \dots \in \mathfrak{D}$ mit $A_i \cap A_j = \emptyset$ für $i \neq j \implies \cup_{i=1}^{\infty} A_i \in \mathfrak{D}$.

Mit $A \cap B = A \setminus (A \setminus B)$ folgt, daß Ringe $\cap$-stabil sind. Induktion liefert für Ringe $\mathfrak{R}$ die Eigenschaft

$$A_1, \dots, A_n \in \mathfrak{R} \implies A_1 \cup \dots \cup A_n \in \mathfrak{R}.$$

Dynkin-Systeme haben die Eigenschaft

$$A, B \in \mathfrak{D}, A \subset B \implies B \setminus A \in \mathfrak{D}, \tag{$*$}$$

denn: $B \cap A^c = (A + B^c + \emptyset + \emptyset + \dots)^c$.

Klar: Jede σ-Algebra ist ein Ring und auch ein Dynkin-System. Dynkin-Systemen „fehlt die $\cap$-Stabilität".

Satz 10.2
Ein $\cap$-stabiles Dynkin-System ist eine σ-Algebra.

Beweis

Sei $A_1, A_2, \dots$ eine nicht notwendigerweise disjunkte Folge von Elementen von $\mathfrak{D}$. Zu zeigen:

$$\bigcup_{i=1}^{\infty} A_i \in \mathfrak{D}.$$

Setze $B_1 = A_1$, $B_n = A_n \cap A_1^c \cap \dots \cap A_{n-1}^c$ für $n > 1$. Die $\cap$-Stabilität liefert $B_n \in \mathfrak{D}\ \forall n \in \mathbb{N}$. Die B_n sind disjunkt, also $\bigcup_{n=1}^{\infty} B_n \in \mathfrak{D}$. Verwende nun $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. *q.e.d.*

Verwandte Begriffe: einen Ring $\mathfrak{R}$ mit $\omega \in \mathfrak{R}$ nennt man *Algebra*.

Beispiel 10.3
Es sei ω überabzählbar. Dann ist

- $\{A \subset \omega \mid \#A < \infty\}$ ein Ring, aber keine Algebra oder gar σ-Algebra;
- $\{A \subset \omega \mid \#A < \infty$ oder $\#A^c < \infty\}$ eine Algebra, aber keine σ-Algebra;
- $\{A \subset \omega \mid A$ oder A^c abzählbar$\}$ eine σ-Algebra ($\neq \wp(\omega)$).

Der Durchschnitt von beliebig vielen Ringen (Algebren, Dynkin-System, σ-Algebren) ist wieder ein Ring (eine Algebra, ein Dynkin-System, eine σ-Algebra).

Definition 10.4
Sei $\mathfrak{E} \subset \wp(\omega)$. Dann heißen

$$\sigma(\mathfrak{E}) := \bigcap_{\substack{\mathfrak{A}\ \sigma\text{-Algebra} \\ \mathfrak{A} \supset \mathfrak{E}}} \mathfrak{A}$$

$$\delta(\mathfrak{E}) := \bigcap_{\substack{\mathfrak{D}\ \text{Dynkin-System} \\ \mathfrak{D} \supset \mathfrak{E}}} \mathfrak{D}$$

die von $\mathfrak{E}$ *erzeugte σ-Algebra* bzw. das von $\mathfrak{E}$ *erzeugte Dynkin-System*.

Satz 10.5
Ist $\mathfrak{E}$ $\cap$-stabil, so gilt

$$\sigma(\mathfrak{E}) = \delta(\mathfrak{E}).$$

Beweis

„$\subset$“: $\delta(\mathfrak{E}) \subset \sigma(\mathfrak{E})$ ist klar.

„$\supset$“: Sei für $D \in \delta(\mathfrak{E})$

$$\mathfrak{D}_D := \{A \subset \omega \mid A \cap D \in \delta(\mathfrak{E})\}$$

$\mathfrak{D}_D$ ist ein Dynkin-System:

$\emptyset \in \mathfrak{D}_D$ ist klar, $\sum A_i \in \mathfrak{D}_D$ ebenso.

$D \in \delta(\mathfrak{E})$, $A \cap D \in \delta(\mathfrak{E}) \implies A^c \cap D = D \setminus (A \cap D) \in \delta(\mathfrak{E})$ (wg. $(*)$ und $A \cap D \in D$), d.h. mit A liegt auch A^c in $\mathfrak{D}_D$.

Da $\mathfrak{E}$ $\cap$-stabil ist, gilt $E' \in \mathfrak{D}_E$ $\forall E, E' \in \mathfrak{E}$, also $\mathfrak{E} \subset \mathfrak{D}_E$ und damit $\delta(\mathfrak{E}) \subset \mathfrak{D}_E$ $\forall E \in \mathfrak{E}$. Damit: $D \in \delta(\mathfrak{E})$, $E \in \mathfrak{E} \implies D \cap E \in \delta(\mathfrak{E})$, also $E \in \mathfrak{D}_D$ $\forall E \in \mathfrak{E}$, $D \in \delta(\mathfrak{E})$. Dies liefert $\mathfrak{E} \subset \mathfrak{D}_D$, also $\delta(\mathfrak{E}) \subset \mathfrak{D}_D$ $\forall D \in \delta(\mathfrak{E})$, d.h.

$$A \in \delta(\mathfrak{E}),\ D \in \delta(\mathfrak{E}) \implies A \cap D \in \delta(\mathfrak{E}).$$

Also: $\delta(\mathfrak{E})$ ist $\cap$-stabil und damit σ-Algebra und damit $\delta(\mathfrak{E}) \supset \sigma(\mathfrak{E})$.

q.e.d.

Mengensysteme über $\mathbb{R}$

Es sei

$$\mathfrak{I} := \{(a,b] \mid a,b \in \mathbb{R}\}.$$

$\mathfrak{I}$ ist offensichtlich $\cap$-stabil.

Satz 10.6
Die Menge $\mathfrak{R}$ aller endlichen Vereinigungen von Elementen von $\mathfrak{I}$ ist ein Ring.

Beweis

$\emptyset \in \mathfrak{R}$ klar.

$\cup$-Stabilität: z.z. $A \in \mathfrak{R}$, $B \in \mathfrak{R} \implies A \setminus B \in \mathfrak{R}$. Seien $A = \bigcup_{k=1}^n I_k$, $B = \bigcup_{i=1}^m I_l$, $\quad I_k, I_l \in \mathfrak{I}$. Dann gilt $A \setminus B = \bigcup_{k=1}^n \bigcap_{l=1}^m I_k \setminus I_l$. Die Aussage würde also folgen aus

(i) $I_1, I_2 \in \mathfrak{I} \implies I_1 \setminus I_2 \in \mathfrak{R}$

(ii) $\mathfrak{R}$ $\cap$-stabil.

Zu (i):

$$I_1 = (a,b], \quad I_2 = (c,d], \quad I_1 \setminus I_2 = \begin{cases} \emptyset, & \text{falls } c \le a \text{ und } d \le b \\ (a,c], & \text{falls } c > a \text{ und } d \ge b \\ (a,d], & \text{falls } c \le a \text{ und } d < b \\ (a,c] + (d,b] & \text{sonst} \end{cases}$$

Man erhält also in allen Fällen ein Element aus $\mathfrak{R}$.

Eigenschaft (ii) folgt aus

$$\left(\bigcup_{k=1}^m I_k\right) \cap \left(\bigcup_{l=1}^n I'_l\right) = \bigcup_{k=1}^m \bigcup_{l=1}^n I_k \cap I'_l \quad \in \mathfrak{R}.$$

q.e.d.

Lemma 10.7
Zu jedem $A \in \mathfrak{R}$ existieren in $n \in \mathbb{N}$ und $I_1, \dots, I_n \in \mathfrak{I}$ mit $I_i \cap I_j = \emptyset$ für $i \ne j$ derart, daß gilt

$$A = \bigcup_{j=1}^n I_j.$$

Beweis

Sei $A = \bigcup_{j=1}^n I_j$ mit (nicht notwendigerweise disjunkten) $I_1, \dots, I_n \in \mathfrak{I}$. Es gilt

$$A = I_1 \cup (I_2 \setminus I_1) \cup (I_3 \setminus (I_1 \cup I_2)) \cup \dots \cup (I_n \setminus (I_1 \cup \dots \cup I_{n-1}))$$

d.h. A ist disjunkte Vereinigung von Mengen der Form

$$I \setminus (I_1 \cup \dots \cup I_k), \quad I, I_1, \dots, I_k \in \mathfrak{I}$$

Es reicht daher, zu zeigen, daß sich solche Mengen als disjunkte Vereinigung von $\mathfrak{I}$-Intervallen schreiben lassen. Dies folgt aus:

(i) $I \setminus (I_1 \cup \cdots \cup I_n) = (I \setminus I_1) \cap \cdots \cap (I \setminus I_k)$.

(ii) Sind A und B disjunkte Vereinigungen von $\mathfrak{I}$-Intervallen, so gilt dies auch für $A \cap B$:

$$\left(\sum_{k=1}^{m} I_k\right) \cap \left(\sum_{l=1}^{n} I'_l\right) \stackrel{\mathfrak{I}\ \cap\text{-stabil}}{=} \sum_{k=1}^{m}\sum_{l=1}^{n} I_k \cap I_l$$

(iii) $I_1, I_2 \in \mathfrak{I} \implies I_1 \setminus I_2$ sind disjunkte Vereinigung von $\mathfrak{I}$-Intervallen (siehe Beweis zu 10.6)

q.e.d.

Die von $\mathfrak{I}$ erzeugte σ-Algebra heißt die *σ-Algebra der Borel-Mengen* von $\mathbb{R}$ und wird mit $\mathfrak{B}$ bezeichnet. $\mathfrak{B}$ enthält „alle für die Praxis relevanten“ Teilmengen von $\mathbb{R}$, z.B.

- Einpunktmengen: $\{a\} = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, a\,]$, damit auch
- alle abzählbaren Mengen (z.B. $\mathbb{Q}$) und auch deren Komplemente, z.B.
- die irrationalen Zahlen: $[\,a,b\,] \in \mathfrak{B}$, denn $[\,a,b\,] = \{a\} \cup (a,b\,]$.
- $(-\infty, a\,] \in \mathfrak{B}$, denn $(-\infty, a\,] = \bigcup_{n=1}^{\infty} (a-n, a\,]$.

13. Oktober 1994

Andere Erzeugendensysteme für $\mathfrak{B} = \sigma(\{(a,b\,] \mid a,b \in \mathbb{R}\})$.

Satz 10.8

$$\mathfrak{B} = \sigma(\{(-\infty, a\,] \mid a \in \mathbb{R}\}) = \sigma(\{U \subset \mathbb{R} \text{ offen}\})$$

Beweis

„$\supset$“: $(-\infty, a\,] = \bigcup (a-n, a\,]$, also $(-\infty, a\,] \in \mathfrak{B} \quad \forall a \in \mathbb{R}$, also

$$\mathfrak{I}_0 := \{(-\infty, a\,] \mid a \in \mathbb{R}\}, \text{ also: } \sigma(\mathfrak{I}_0) \subset \mathfrak{B}.$$

Wegen $(a,b\,] = (-\infty, b\,] \cap (-\infty, a\,]^c$ gilt $\mathfrak{I} \subset \sigma(\mathfrak{I}_0)$, also $\mathfrak{B} = \sigma(\mathfrak{I}) \subset \sigma(\mathfrak{I}_0)$ und damit insgesamt $\sigma(\mathfrak{I}_0) = \mathfrak{B}$.

Sei $\mathfrak{U} := \{U \in \mathbb{R} \mid U \text{ offen}\}$. Behauptung: $\mathfrak{U} \subset \mathfrak{B}$.

Sei hierzu $U \subset \mathbb{R}$ offen. Dann existiert zu jedem $x \in U$ eine offene Umgebung U_x von x mit $U_x \subset U$, also ein $\varepsilon > 0$ mit $(x-\varepsilon, x+\varepsilon) \subset U$, also $a, b \in \mathbb{Q}$ mit $x \in (a,b\,]$ und $(a,b\,] \in U$.

Bemerkung: *Falsch* ist jetzt $U = \bigcup_{x \in U} \{x\}$, da diese Menge *nicht* abzählbar ist.

Also:

$$U = \bigcup_{\{(a,b) \in \mathbb{Q}^2 \;\mid\; (a,b\,] \subset U\}} (a,b\,],$$

wobei $(a,b) \in \mathbb{Q}^2$ ein Paar und $(a,b\,] \subset U$ eine abzählbare Vereinigung ist. Dies liefert $\mathfrak{U} \subset \mathfrak{B}$, also $\sigma(\mathfrak{U}) \subset \mathfrak{B}$.

„$\subset$“:

Andererseits folgt aus $(a,b\,] = \underbrace{\bigcap_{n=1}^{\infty}}_{\text{abzählbarer Durchschnitt}} \underbrace{\left(a, b + \frac{1}{n}\right)}_{\text{offen}}$, daß $\mathfrak{I} \subset \sigma(\mathfrak{U})$ und damit

$$\mathfrak{B} = \sigma(\mathfrak{J}) \subset \sigma(\mathfrak{U}).$$

q.e.d.

Sei $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, $\overline{\mathfrak{B}} := \{A \subset \overline{\mathbb{R}} \mid A \cap \mathbb{R} \in \mathfrak{B}\}$.

Einschub: Rechnen mit ∞:

- $a + \infty = \infty \quad \forall a \in \mathbb{R}$,
- $a + (-\infty) = -\infty$,
- $a \cdot \infty = \infty$,
- $a \cdot (-\infty) = -\infty$,
- $0 \cdot \infty = 0$, $0 \cdot (-\infty) = 0$,
- $\infty + (-\infty)$ ist nicht definiert!

11 Additive Mengenfunktionen

11.1 ???

Sei $\Omega \neq \emptyset$, $\mathfrak{E} \subset \wp(\Omega)$.

Definition 11.1
$\mu : \mathfrak{E} \to [0, \infty]$ heißt *additiv*, wenn $\mu(A \cup B) = \mu(A) + \mu(B)$ für alle $A, B \in \mathfrak{E}$ mit $A \cap B = \emptyset$, $A \cup B \in \mathfrak{E}$.

μ heißt σ-*additiv*, wenn

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

für disjunkte Mengenfolgen (A_i) mit $\bigcup_{i=1}^{\infty} \in \mathfrak{E}$.

Ist $\mathfrak{R}$ ein Ring, $\mu : \mathfrak{R} \to [0, \infty]$ additiv, $\mu(\emptyset) = 0$, so heißt μ ein *Inhalt*; ist μ dabei noch σ-additiv, so heißt μ *Prämaß*. Ein Prämaß auf einer σ-Algebra heißt *Maß*.

Ein Paar $(\Omega, \mathfrak{A})$, mit Ω Menge ($\neq \emptyset$), $\mathfrak{A}$ eine σ-Algebra über Ω heißt *meßbarer Raum*, ein Tripel $(\Omega, \mathfrak{A}, \mu)$ mit $(\Omega, \mathfrak{A})$ meßbarer Raum, $\mu : \mathfrak{A} \to [0, \infty]$ Maß, heißt *Maßraum*, die Elemente von $\mathfrak{A}$ *meßbare Mengen*.

Beispiel 11.2

(i) Sei $x \in \Omega$. Dann ist

$$\delta_x : \wp(\Omega) \to [0, \infty], \delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & \text{sonst} \end{cases}$$

ein Maß, das *Einpunkt-Maß* oder *Dirac-Maß* in x.

(ii) Sei Ω überabzählbar, $\mathfrak{A} = \{A \subset \Omega \mid A \text{ oder } A^c \text{ abzählbar}\}$. Setzt man $\mu(a) = \begin{cases} 0, & A \text{ abzählbar} \\ 1, & \text{sonst} \end{cases}$, so erhält man wieder ein Wahrscheinlichkeitsmaß.

(iii) Ist $\mu_1, \mu_2, \ldots$ eine Folge von Inhalten (Prämaßen) auf einem Ring $\mathfrak{R}$, $\alpha_1, \alpha_2, \cdots \geq 0$, so ist auch

$$\mu : \mu(A) = \sum_{i=1}^{\infty} \alpha_i \mu_i(A) \qquad (\text{wobei } \mu = \sum \alpha_i \mu_i)$$

wieder ein Inhalt (bzw. Prämaß) auf $\mathfrak{R}$.

Lemma 11.3
Sei $\mathfrak{R}$ Ring, $\mu : \mathfrak{R} \to [0, \infty]$ Inhalt. Dann gilt für alle $A, B, A_1, \cdots \in \mathfrak{R}$

(i) $A \subset B \implies \mu(A) \leq \mu(B)$ (Isotonie)

(ii) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$

(iii) $\mu(A_1 \cup \cdots \cup A_n) \leq \mu(A_1) + \cdots + \mu(A_n)$ (Subadditivität)

(iv) $A_1, A_2, \cdots \in \mathfrak{R}$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{R} \implies \sum_{i=1}^{\infty} \mu(A_i) \leq \mu(\bigcup_{i=1}^{\infty} A_i)$.

Beweis

(nur (iv))

Additivität und Isotonie liefern

$$\sum_{i=1}^{n} \mu(A_i) = \mu\left(\bigcup_{i=1}^{n} A_i\right) \leq \mu\left(\bigcup_{i=1}^{\infty} A_i\right).$$

Nun $n \to \infty$. *q.e.d.*

σ-Additivität kann als Stetigkeitseigenschaft aufgefaßt werden. Vorher Terminologie:

Man nennt eine Folge $A_1, A_2, \ldots$ von Mengen

- *isoton*, wenn $A_n \subset A_{n+1} \forall n \in \mathbb{N}$,
- *antiton* wenn $A_{n+1} \subset A_n \forall n$.

$A_n \uparrow A$ steht für: $(A_n)_{n\in\mathbb{N}}$ isoton, $A = \bigcup_{n=1}^{\infty} A_n$;
$A_n \downarrow A$ steht für $(A_n)_{n\in\mathbb{N}}$ antiton, $A = \bigcap_{n=1}^{\infty} A_n$.

Satz 11.4
Es sei μ ein Inhalt auf dem Ring $\mathfrak{R}$. Dann gilt (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (iv), wobei

(i) μ Prämaß (also: μ σ-additiv),

(ii) $\lim_{n\to\infty} \mu(A_n) = \mu(A)$, falls $A_n \uparrow A \quad (\in \mathfrak{R})$ (μ heißt dann *stetig von unten*, beachte: ∞ als Grenzwert zugelassen.)

(iii) $\lim_{n\to\infty} \mu(A_1) = \mu(A)$, falls $A_n \downarrow A$ und $\mu(A_n) < \infty$ $\forall n \in \mathbb{N}$ (μ stetig von oben).
??? Bild

(iv) $\lim \mu(A_n) = 0$, falls $A_n \downarrow \emptyset$ und $\mu(A_n) < \infty \quad \forall n \in \mathbb{N}$ (μ ist $\emptyset$-stetig).

Genügt μ der Bedingung $\mu(A) < \infty \quad \forall A \in \mathfrak{R}$???, so gilt

$$\text{(i)} \iff \text{(ii)} \iff \text{(iii)} \iff \text{(iv)}.$$

Beweis

„(i) $\Longrightarrow$ (ii)“:

Sei $B_1 := A_1$, $B_n := A_n \setminus A_{n-1}$, $n > 1$. $(B_n)_{n\in\mathbb{N}}$ ist dann eine disjunkte Folge von Mengen in $\mathfrak{R}$ mit

$$A_n = B_1 \cup \cdots \cup B_n \;\forall n, \quad A = \bigcup_{n=1}^{\infty} B_n.$$

Als Prämaß ist μ σ-additiv, also gilt

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n\to\infty} \mu\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \mu(A_n).$$

„(ii) $\Longrightarrow$ (i)“:

Ist $(A_n)_{n\in\mathbb{N}}$ eine disjunkte Mengenfolge in $\mathfrak{R}$, $A = \bigcup_{i=1}^{\infty} A_i$, so gilt

$$B_n \uparrow A \quad \text{mit} \quad B_n := A_1 \cup \cdots \cup A_n,$$

also folgt

$$\begin{aligned}
\mu(A) &= \lim_{n\to\infty} \mu(B_n) \\
&= \lim_{n\to\infty} \mu(A_1 \cup \cdots \cup A_n) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A_i) \\
&= \sum_{i=1}^{\infty} \mu(A_i).
\end{aligned}$$

„(ii) $\Longrightarrow$ (iii)“:

Sei $B_n := A_1 \setminus A_n$. Aus $A_n \downarrow A$ folgt $B_n \uparrow A_1 \setminus A$. Additivität liefert $\mu(B_n) + \mu(A_n) = \mu(A_1)$, also $\mu(B_n) = \mu(A_1) - \mu(A_n)$ (hier geht $\mu(A_n) < \infty$ ein).

Damit folgt nun

$$\mu(A_1 \setminus A) = \lim_{n\to\infty} \mu(B_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n).$$

Da $A \subset A_1$ und $\mu(A_1) < \infty$ gilt $\mu(A) < \infty$ (Isotonie), also $\mu(A_1 \setminus A) = \mu(A_1) - \mu(A)$.

„(iii) $\Longrightarrow$ (iv)“:

ist trivial ((iv) ist Spezialfall von (iii)).

Es gelte nun zusätzlich

$$\mu(A) < \infty \quad \forall A \subset \mathfrak{R}.$$

„(iv) $\Longrightarrow$ (ii)“:

Sei $A_n \uparrow A \in \mathfrak{R}$. Setze $B_n := A \setminus A_n$, dann gilt $B_n \downarrow \emptyset$. Die Zusatzbedingung liefert $\mu(B_n) < \infty \;\forall n$, also $\lim_{n\to\infty} \mu(B_n) = 0$. (*) erlaubt auch den Schluß

$$\mu(B_n) = \mu(A) - \mu(A_n)$$

also

$$0 = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} (\mu(A) - \mu(A_n)) = \mu(A) - \lim_{n\to\infty} \mu(A_n).$$

q.e.d.

11.2 Fortsetzung von Prämaßen zu Maßen

Gibt es zu einem Prämaß μ auf einem Ring $\mathfrak{R}$ stets eine Fortsetzung auf $\sigma(\mathfrak{R})$, also ein Maß

$$\overline{\mu}:\ \sigma(\mathfrak{r}) \to [\,0,\infty\,] \quad \text{mit } \overline{\mu}(A) = \mu(A) \quad \forall A \in \mathfrak{R}?$$

18. Oktober 1994

Definition 11.5
$\mu^* : \wp(\Omega) \to [\,0,\infty\,]$ heißt *äußeres Maß*, wenn

(i) $\mu(\emptyset) = 0$

(ii) $A \subset B \implies \mu^*(A) \leq \mu^*(B)$

(iii) $\mu^*(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$ (σ-Additivität)

Lemma 11.6
Sei $\mu^* : \wp(\Omega) \to [0,\infty]$ ein äußeres Maß und

$$\mathfrak{A} := \{A \subset \Omega \mid \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) \quad \forall B \subset \Omega\}.$$

Dann ist $\mathfrak{A}$ eine σ-Algebra und die Restriktion von μ^* auf $\mathfrak{A}$ ist ein Maß.

Beweis

(i) $\Omega \in \mathfrak{A} : \mu^*(B \cap \Omega) + \mu^*(B \cap \Omega^c) = \mu^*(B)$

(ii) $A \in \mathfrak{A} \implies A^c \in \mathfrak{A}$ trivial.

(iii) $A, B \in \mathfrak{A} \implies A \cup B \in \mathfrak{A}$: Für alle $C \subset \Omega$ gilt $\mu^*(C) = \mu^*(C \cap B) + \mu^*(C \cap B^c)$.
Mit $C \cap A$ und $C \cap A^c$ anstelle von C erhält man $\forall C \subset \Omega$

$$\begin{aligned}
\mu^*(C \cap A) &= \mu^*(C \cap A \cap B) + \mu^*(C \cap A \cap B^c) \quad \text{und} \\
\mu^*(C \cap A^c) &= \mu^*(C \cap A^c \cap B) + \mu^*(C \cap A^c \cap B^c), \quad \text{also} \\
\mu^*(C) &= \mu^*(C \cap A) + \mu^*(C \cap A^c) \\
&= \mu^*(C \cap A \cap B) + \mu^*(C \cap A \cap B^c) + \mu^*(C \cap A^c \cap B) + \mu^*(\underbrace{C \cap A^c \cap B^c}_{(A\cup B)^c}).
\end{aligned}$$

Im Spezialfall $C \subset (A \cup B)$ für C ergibt sich, da $A, B \subset A \cup B$

$$\mu^*(C \cap (A \cup B)) = \mu^*(C \cap A \cap B) + \mu^*(C \cap A \cap B^c) + \mu^*(C \cap A^c \cap B). \tag{1}$$

Aus den letzten beiden Gleichungen folgt

$$\mu^*(C) = \mu^*(C \cap (A \cup B)) + \mu^*(C \cap (A \cup B)^c) \quad \forall C \subset \Omega,$$

und damit gilt (iii).

(iv) $A, B \in \mathfrak{A} \implies A \cap B \in \mathfrak{A}$ folgt unmittelbar aus (ii) und (iii).

(v) Sei $(A_n)_{n\in\mathbb{N}}$ eine disjunkte Mengenfolge in $\mathfrak{A}$, $A := \bigcup_{n=1}^{\infty} A_n$. Setze $A = A_1, B = A_2$ in (1):

$$\mu^*(C \cap (A_1 \cup A_2)) = \mu^*(C \cap A_1) + \mu^*(C \cap A_2) \quad \forall C \subset \Omega.$$

Induktion liefert

$$\mu^*(C \cap (\bigcup_{i=1}^{n} A_i)) = \sum_{i=1}^{n} \mu^*(C \cap A_i) \quad \forall C \subset \Omega,\ n \in \mathbb{N}. \tag{2}$$

Bereits aus Teil (iv) bekannt ist $B_n := \bigcup_{i=1}^{n} A_i \in \mathfrak{A}$. Mit $C \cap B_n^c \supset C \cap A^c$ und der Isotonie von μ_n^* folgt

$$\mu^*(C \cap B_n^c) \geq \mu^*(C \cap A^c),$$

also

$$\begin{aligned} \mu^*(C) &= \mu^*(C \cap B_n) + \mu^*(C \cap B_n^c) \\ &= \sum_{i=1}^{n} \mu^*(C \cap A_i) + \mu^*(C \cap A^c) \quad \forall n \in \mathbb{N},\ \forall c \subset \Omega. \end{aligned}$$

Die σ-Additivität liefert nun

$$\begin{aligned} \mu^*(C) &\geq \sum_{i=1}^{\infty} \mu^*(C \cap A_i) + \mu^*(C \cap A^c) \\ &\geq \mu^*(C \cap A) + \mu^*(C \cap A^c) \quad \forall C \subset \Omega. \end{aligned} \tag{3}$$

Nun wieder mit σ-Subadditivität

$$\begin{aligned} \mu^*(C) &= \mu^*(C \cap A + C \cap A^c + \emptyset + \emptyset + \dots) \\ &\leq \mu^*(C \cap A) + \mu^*(C \cap A^c), \end{aligned} \tag{4}$$

also insgesamt

$$\mu^*(C) = \mu^*(C \cap A) + \mu^*(C \cap A^c) \quad \forall C \subset \Omega,$$

d.h. $A \in \mathfrak{A}$, $\mathfrak{A}$ enthält abzählbar viele disjunkte Vereinigungen.

„So, und jetzt holen wir wieder den Herrn Dynkin aus dem Schrank ... “

Aus (i), (ii), (iii) und (v) folgt, daß $\mathfrak{A}$ ein $\cap$-stabiles Dynkin-System ist, also eine σ-Algebra (Satz 10.2).

Noch zu zeigen: Auf $\mathfrak{A}$ ist μ^* ein Maß, d.h. σ-additiv. Sei hierzu $(A_n)_{n\in\mathbb{N}}$ eine disjunkte Mengenfolge in $\mathfrak{A}$. Wählt man $C = \bigcup_{i=1}^{\infty} A_i \quad (= A)$ in (3) und beachtet man, daß die Ungleichungen in (3) wegen (4) sogar Gleichungen sind, so folgt

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu^*(A_i),$$

denn $C \cap A^c = \emptyset$ und $C \cap A_i = A_i$.

„Damit haben wir ein hochgradig nicht-triviales Lemma bewiesen.“

q.e.d.

Satz 11.7 (Maßerweiterungssatz, Fortsetzungssatz für Maße, Satz von Carathéodory)
Es sei μ_0 ein Prämaß auf dem Ring $\mathfrak{R}$. Dann existiert ein Maß μ auf $\sigma(\mathfrak{R})$ mit $\mu(A) = \mu_0(A) \quad \forall A \in \mathfrak{R}$.

Beweis

Wir definieren $\mu^* : \wp(\Omega) \to [\,0, \infty\,]$ durch

$$\mu^*(C) := \inf\left\{\sum_{n=1}^{\infty} \mu_0(A_n) \mid (A_n)_{n\in\mathbb{N}} \subset \mathfrak{R}, C \subset \bigcup_{n=1}^{\infty} A_n\right\},$$

wobei wir $\inf \emptyset = \infty$ vereinbaren.

Offensichtlich gilt $\mu^*(\emptyset) = 0$, $\mu^*(A) \le \mu^*(B)$ für $A \subset B$ (jede Überdeckung von B überdeckt auch A).

Außerdem gilt

$$\mu^*\left(\bigcup_{n=1}^{\infty} C_n\right) \le \sum_{n=1}^{\infty} \mu^*(C_n)\ \forall (C_n)_{n\in\mathbb{N}} \subset \wp(\Omega),$$

denn wir können $\mu^*(C_n) < \infty$ voraussetzen ($\forall n \in \mathbb{N}$).

Sei $\varepsilon > 0$. Dann existiert zu jedem $n \in \mathbb{N}$ eine Folge $(A_{nm})_{m\in\mathbb{N}} \subset \mathfrak{R}$ mit $\bigcup_{m=1}^{\infty} A_{nm} \supset C_n$ und

$$\sum_{m=1}^{\infty} \mu_0(A_{nm}) \le \mu^*(C_n) + \varepsilon \cdot 2^{-n}.$$

Wegen $\bigcup_{n,m=1}^{\infty} A_{nm} \supset \bigcup_{n=1}^{\infty} C_n$ gilt nun

$$\begin{aligned} \mu^*\left(\bigcup_{n=1}^{\infty} C_n\right) &\le \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mu_0(A_{nm}) \\ &\le \sum_{n=1}^{\infty} \left(\mu^*(C_n) + \varepsilon \cdot 2^{-n}\right) \\ &= \sum_{n=1}^{\infty} \mu^*(C_n) + \varepsilon \end{aligned}$$

Da $\varepsilon > 0$ beliebig war und die linke Seite nicht von ε abhängt, folgt die σ-Subadditivität mit $\varepsilon \downarrow 0$. Damit ist gezeigt: μ^* ist äußeres Maß. Wir zeigen nun, daß μ^* auf $\mathfrak{R}$ mit μ_0 übereinstimmt.

Seien hierzu $A, A_1, A_2, \dots \in \mathfrak{R}$ mit $A \subset \bigcup_{n=1}^\infty A_n$. Setze $A'_n := A_n \cap A$ $(\in \mathfrak{R})$. Dann gilt

$$A = \bigcup_{n=1}^\infty A'_n.$$

Sei $B_1 := A'_1$, $B_n := A'_n \cap (A'_1 \cup \dots \cup A'_{n-1})$ für $n > 1$. Diese bilden eine disjunkte Zerlegung von A und sind Elemente von $\mathfrak{R}$. Da μ_0 ein Prämaß ist, folgt

$$\begin{aligned}
\mu_0(A) &= \sum_{n=1}^\infty \mu_0(B_n) \\
&\leq \sum_{n=1}^\infty \mu_0(A'_n) \\
&\leq \sum_{n=1}^\infty \mu_0(A_n).
\end{aligned}$$

Dies gilt für alle Folgen $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{R}$ mit $\bigcup_{n=1}^\infty A_n \supset A$, also folgt $\mu_0(A) \leq \mu^*(A)$.

Andererseits folgt mit der Überdeckung $A_1 = A$, $A_2 = A_3 = \dots = \emptyset$, daß auch $\mu^*(A) \leq \mu_0(A)$. Insgesamt gilt also

$$\mu^* = \mu_0 \quad \text{auf } \mathfrak{R}.$$

Wir behaupten nun, daß alle $A \in \mathfrak{R}$ die Teilungseigenschaft aus Lemma 11.6 haben, d.h.

$$\mu^*(C) = \mu^*(C \cap A) + \mu^*(C \cap A^c) \quad \forall A \in \mathbb{R},\ C \subset \Omega.$$

Wegen der σ-Additivität reicht es, „$\geq$“ zu zeigen. Wir können $\mu^*(c) < \infty$ annehmen. Für jede Folge $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{R}$ mit $\bigcup_{n=1}^\infty A_n \supset C$ gilt

$$\begin{aligned}
\bigcup_{n=1}^\infty (A_n \cap A) &\supset C \cap A, \\
\bigcup_{n=1}^\infty (A_n \cap A^c) &\supset C \cap A^c, \\
A_n \cap A &\in \mathfrak{R}, \\
A_n \cap A^c &\in \mathfrak{R},
\end{aligned}$$

also

$$\begin{aligned}
\sum_{n=1}^\infty \mu_0(A_n \cap A) &\geq \mu^*(C \cap A), \\
\sum_{n=1}^\infty \mu_0(A_n \cap A^c) &\geq \mu^*(C \cap A^c),
\end{aligned}$$

und damit wegen der endlichen Additivität von μ_0

$$\begin{aligned}\mu^*(C\cap A)+\mu^*(C\cap A^c) &\le \sum_{n=1}^{\infty}(\mu_0(A_n\cap A)+\mu_0(A_n\cap A^c))\\ &= \sum_{n=1}^{\infty}\mu_0(A_n).\end{aligned}$$

Bildet man nun das Infimum über die zugelassenen Folgen $(A_n)_{n\in\mathbb{N}}$, so folgt

$$\mu^*(C\cap A)+\mu^*(C\cap A^c)\le\mu^*(C).$$

Wende nun Lemma 11.6 an: Es gibt ein Maß μ auf $\sigma(\mathfrak{R})$, das auf $\mathfrak{R}$ mit μ^* übereinstimmt (denn $\mathfrak{A}\supset\mathfrak{R}$, also $\mathfrak{A}\supset\sigma(\mathfrak{R})$). Da μ^* und μ_0 auf $\mathfrak{R}$ übereinstimmen, ist μ eine Fortsetzung von μ_0. *q.e.d.*

20. Oktober 1994

Wann ist die Fortsetzung eindeutig?

Satz 11.8
Es sei $\varphi\subset\wp(\Omega)$ $\cap$-stabil. Weiter sei $(\varepsilon_n)_{n\in\mathbb{N}}\subset\varphi$ mit $E_n\uparrow\Omega$. Sind μ_1,μ_2 Maße auf $\sigma(\varphi)$ mit

$$\mu_1(E)=\mu_2(E)\quad\forall E\in\varphi\quad\text{und}\quad\mu_1(E_n)<\infty\quad\forall n\in\mathbb{N},$$

so folgt

$$\mu_1=\mu_2.$$

Beweis
Es sei

$$\mathfrak{D}_n:=\{A\in\sigma(\varphi)\mid\mu_1(A\cap E_n)=\mu_2(A\cap E_n)\}.$$

Dann ist $\mathfrak{D}_n$ ein Dynkin-System. Gilt $A\in\mathfrak{D}_n$, so folgt

$$\begin{aligned}\mu_1(A^c\cap E_n) &= \mu_1(E_n)-\mu_1(A\cap E_n)\\ &= \mu_2(E_n)-\mu_2(A\cap E_n)\\ &= \mu_2(A^c\cap E_n),\end{aligned}$$

also

$$A^c\in\mathfrak{D}_n.$$

Hierbei haben wir von $\mu_1(E_n)<\infty$ Gebrauch gemacht. — Die anderen Eigenschaften von Dynkin-Systemen sind auch leicht zu überprüfen.

Da φ $\cap$-stabil ist, gilt $\varphi\subset\mathfrak{D}_n$, also wegen Satz 10.5:

$$\mathfrak{D}_n=\sigma(\varphi).$$

Sei nun $A\in\sigma(\varphi)$. Mit $E_n\cap A\uparrow A$ und der Stetigkeit von unten folgt

$$\mu_1(A)=\lim_{n\to\infty}\mu_1(A\cap E_n)=\lim_{n\to\infty}\mu_2(A\cap E_n)=\mu_2(A).$$

q.e.d.

Korollar 11.9
Sind P und Q Wahrscheinlichkeitsmaße auf $(\Omega, \mathfrak{A})$, die auf einem $\cap$-stabilen Erzeuger von $\mathfrak{A}$ übereinstimmen, so gilt $P = Q$.

Beweis

Ergänze, wenn nötig, φ durch Ω (zerstört die $\cap$-Stabilität nicht), wähle $E_n = \Omega \; \forall n$ und beachte $P(\Omega) = Q(\Omega) = 1$. *q.e.d.*

Definition 11.10
Ein Inhalt μ auf einem Ring $\mathfrak{R}$ über Ω heißt *σ-endlich*, wenn es eine Folge $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{R}$ gibt mit $A_n \uparrow \Omega$, $\mu(A_n) < \infty \; \forall n$.

Korollar 11.11
Zu einem σ-endlichen Prämaß auf $\mathfrak{R}$ existiert genau eine Fortsetzung zu einem Maß auf $\sigma(\mathfrak{R})$.

Beweis

Setze $\varphi = \mathfrak{R}$ in Satz 11.8 und verwende Definition 11.10. *q.e.d.*

11.3 Das Lebesgue-Maß

Es sei wieder $\mathfrak{J} = \{(a, b] \mid a, b \in \mathbb{R}\}$ und $\mathfrak{R}$ der Ring der endlichen Vereinigungen von $\mathfrak{J}$-Intervallen (siehe auch 10.3).

Lemma 11.12
Es gibt genau einen Inhalt $\ell_0 : \; \mathfrak{R} \to [\,0, \infty\,]$ mit $\ell_0((a, b\,]) = b - a$ für alle $a, b \in \mathbb{R}$ mit $a \leq b$.

Beweis

Jedes $\mathfrak{R}$-Element läßt sich als disjunkte Vereinigung von $\mathfrak{J}$-Intervallen schreiben (Lemma 10.7), wegen der Additivität kann es also höchstens ein ℓ_0 geben.

Nachweis der Existenz: Definiere ℓ_0 zunächst auf $\mathfrak{J}$ durch

$$\ell_0((a, b\,]) = b - a \qquad (a \leq b).$$

Ist dies mit der Additivität verträglich? Wenn $(a, b] = (a, c_1] + (c_1, c_2] + \cdots + (c_n, b]$, so folgt $b - a = (c_1 - a) + (c_2 - c_1) + \cdots + (c_n - c_{n-1}) + (b - c_n)$, also: Additivität gewahrt.

Sei nun $A = \sum_{i=1}^n I_i \in \mathfrak{R}$, $I_1, \ldots, I_n \in \mathfrak{J}$. Setze

$$\ell_0(A) := \sum_{i=1}^n \ell_0(I_j).$$

Ist diese Fortsetzung wohldefiniert? Angenommen $A = I_1 + I_2 + \cdots + I_n = I_1' + \cdots + I_m'$, $I_j, I_k \in \mathfrak{J}$. Dann gilt: $I_j = I_j \cap A = \sum_{k=1}^n I_j \cap I_k'$, also folgen, da ℓ_0 auf $\mathfrak{J}$ additiv und $\mathfrak{J}$ $\cap$-stabil ist

$$\begin{aligned}
\ell_0(I_j) &= \sum_{k=1}^{m} \ell_0(I_j \cap I'_k) \qquad \text{und analog} \\
\ell_0(I'_k) &= \sum_{j=1}^{n} \ell_0(I_j \cap I'_k) \qquad \text{und damit} \\
\sum_{j=1}^{n} \ell_0(I_j) &= \sum_{j=1}^{n}\sum_{k=1}^{m} \ell_0(I_j \cap I'_k) \\
&= \sum_{k=1}^{m}\sum_{j=1}^{n} \ell_0(I_j \cap I'_k) \\
&= \sum_{k=1}^{m} \ell_0(I'_k),
\end{aligned}$$

beide Darstellungen führen also auf denselben Wert.

Es bleibt zu zeigen, daß ℓ_0 additiv ist. Hat man $A = \sum_{j=1}^{n} I_j$, $B = \sum_{k=1}^{m} I'_k$, $A \cap B = \emptyset$, so folgt (mit $A \cap B = \emptyset \implies I_j \cap I'_k = \emptyset \quad \forall j, k$)

$$\begin{aligned}
A \cup B &= \sum_{j=1}^{n} I_j + \sum_{k=1}^{m} I'_k, \qquad \text{d.h.} \\
\ell_0(A \cup B) &= \sum_{j=1}^{n} \ell_0(I_j) + \sum_{k=1}^{m} \ell_0(I'_k) \\
&= \ell_0(A) + \ell_0(B).
\end{aligned}$$

q.e.d.

Lemma 11.13
ℓ_0 ist ein Prämaß.

Beweis

Da ℓ_0 auf $\mathfrak{R}$ endlich ist, reicht der Nachweis der $\emptyset$-Stetigkeit (Satz 11.4). Sei hierzu $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{R}$ antiton mit $\varepsilon = \liminf {}_0(A_n) > 0$. Zu zeigen: $\bigcup_{n=1}^{\infty} \neq \emptyset$.

Jedes A_n läßt sich schreiben als

$$A_n = (a_{n_1}, b_{n_1}] + (a_{n_2}, b_{n_2}] + \cdots + (a_{n_{k_n}}, b_{n_{k_n}}].$$

Setze

$$\begin{aligned}
c_{n_i} &= \begin{cases} a_{n_i} + \frac{1}{k_n}\varepsilon 2^{-n-1}, & \text{falls } b_{n_i} - a_{n_1} > \frac{1}{k_n}\varepsilon 2^{-n-1} \\ b_{n_i} & \text{sonst,} \end{cases} \\
B_n &= [c_{n_1}, b_{n_1}] + [c_{n_2}, b_{n_2}] + \cdots + [c_{n_{k_n}}, b_{n_{k_n}}], \\
C_n &= (c_{n_1}, b_{n_1}) + (c_{n_2}, b_{n_2}) + \cdots + (c_{n_{k_n}}, b_{n_{k_n}}).
\end{aligned}$$

Dann gilt: B_n abgeschlossen, $C_n \subset B_n \subset A_n$, $\ell_0(A_n \cap {C_n}^c) \leq \varepsilon 2^{-n-1}$. Aus $A_n \downarrow$ folgt

$$\ell_0(A_n \cap (C_1 \cup \cdots \cup C_n)^c) \leq \ell_0\left(\bigcup_{k=1}^n (A_k \cap {C_k}^c)\right) \leq \varepsilon \sum_{k=1}^n 2^{-k-1} \leq \frac{\varepsilon}{2}.$$

Da $\bigcap_{k=1}^n C_k \subset A_n$ und $\ell_0(A_n) \geq \varepsilon$, folgt hieraus

$$\ell_0\left(\bigcap_{k=1}^n {C_k}^c\right) \geq \frac{\varepsilon}{2}, \quad \text{insbesondere } \bigcap_{k=1}^n C_k \neq \emptyset.$$

Sei nun $K_n := \bigcap_{k=1}^n B_k$. Wegen $\bigcap_{k=1}^n C_k \subset \bigcap_{k=1}^n B_k$ gibt es zu jedem n ein $x_n \in K_n$.

Betrachte die Folge $(x_n)_{n\in\mathbb{N}}$: alle Glieder liegen in der beschränkten Menge A_1, als existiert ein Häufungspunkt x_0. Da $(x_{n+m})_{m\in\mathbb{N}} \subset K_n$ (K_n antiton) und $K-n$ abgeschlossen, gilt $x_0 \in K_n$. Also:

$$x_o \in \bigcap_{n=1}^\infty K_n \subset \bigcap_{n=1}^\infty A_n \qquad \text{und damit}$$
$$\bigcap_{n=1}^\infty A_n \neq \emptyset.$$

q.e.d.

Satz 11.14
Es existiert genau ein Maß $\ell : \mathfrak{B} \to [0,\infty]$ mit $\ell((a,b]) = b-a$ für alle $b \geq a$. Wir nennen ℓ das *Lebesgue-Maß.* ℓ ist σ-endlich.

Beweis

Satz 11.7 und Lemma 11.13 ergeben die Existenz. Korollar 11.11 liefert die Eindeutigkeit, denn wegen $(-n,n] \uparrow \mathbb{R}$, $\ell_0((-n,n]) = 2n < \infty$ ist ℓ_0 (und damit ℓ) σ-endlich. *q.e.d.*

Mit der Stetigkeit von oben folgt

$$\ell(\{x\}) = \lim_{n\to\infty} \ell\left(\left(x-\frac{1}{n},x\right]\right) = \lim_{n\to\infty}\frac{1}{n} = 0.$$

Des Lebesgue-Maß ordnet also Einpunktmengen und damit allen endlichen und abzählbar unendlichen Mengen (wie z.B. $\mathbb{Q}$) den Wert 0 zu. Diese Mengen nennt man *Lebesgue-Nullmengen.* Hieraus folgt auch

$$\ell((a,b]) = \ell([a,b]) = \ell([a,b)) = \ell((a,b)).$$

Übungsaufgabe 7 gibt ein Beispiel für eine überabzählbare ℓ-Nullmenge.

Satz 11.15
Das Lebesgue-Maß ist translationsinvariant, d.h. $\ell(x+A) = \ell(A)$ für alle $x \in \mathbb{R}$, $A \in \mathfrak{B}$, wobei $x + A = \{x+y \mid y \in A\}$.

Beweis

Definiere $\ell' : \mathfrak{B} \to [0,\infty]$ durch $\ell'(A) = \ell(x+A)$. Dann ist ℓ' ein Maß, das auf dem $\cap$-stabilen Erzeuger $\mathfrak{I}$ von $\mathfrak{B}$ mit ℓ übereinstimmt:

$$\ell'((a,b]) = \ell((a+x, b+x]) = b - a = \ell((a,b]).$$

Satz 11.8 liefert $l = l'$. *q.e.d.*

Man kann sogar zeigen daß ℓ das *einzige* translationsinvariante Maß auf $\mathfrak{B}$ ist, das dem Einheitsintervall den Wert 1 zuordnet.

Bemerkung 11.16
Die Spur von $\mathfrak{B}$ in $(0,1]$ ist die σ-Algebra

$$\mathfrak{B}|_{[0,1)} = \{B \cap [0,1) \mid B \in \mathfrak{B}\} = \{B \subset [0,1) \mid B \in \mathfrak{B}\},$$

siehe auch Übungsaufgabe 1. Sei nun

$$P : \mathfrak{B}|_{[0,1)} \to [0,1], \quad P(A) = \ell(A),$$

die Einschränkung von ℓ auf $[0,1)$. Dann ist $([0,1), \mathfrak{B}|_{[0,1)}, P)$ ein Wahrscheinlichkeitsraum. Dieser Raum ist unser Modell für ein „Laplace-Experiment über $[0,1)$“. Beachte: Alle Elementarereignisse $\{w\}$ haben die Wahrscheinlichkeit 0.

Translationsinvarianz gilt nun modulo 1. Damit ist das nach Kapitel 9 beschriebene Problem gelöst. Der dort bewiesene Satz hat die (interessante) Konsequenz $\mathfrak{B}|_{[0,1)} \neq \wp([0,1))$, d.h.

$$\mathfrak{B} \neq \wp(\mathbb{R}).$$

25. Oktober 1994

Ganz allgemein gilt: Ist $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum, $A \in \mathfrak{A}$, so wird durch

$$\mu_A : \mathfrak{A} \to [0, \infty], \quad \mu_A(B) := \mu(A \cap B)$$

ein Maß auf $(\Omega, \mathfrak{A})$ definiert. Im Falle $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}, \mathfrak{B}, \ell)$, $\infty > \ell(A) > 0$, nennen wir das Maß

$$\mu : \begin{cases} \mathfrak{B} \to [0,1] \\ B \mapsto \frac{1}{\ell(A)} \cdot \ell(A \cap B) \end{cases}$$

die *Gleichverteilung* auf der Menge A, Notation:

$$U(a,b) \text{ oder } \mathrm{Unif}(a,b) \text{ im Falle } A = [a,b] \text{ (oder } (a,b) \text{ etc.)}$$

(engl: uniform distribution)

12 Meßbare Abbildungen: Zufallsgrößen, Bildmaße, Verteilungen

12.1 Grundlegendes

Definition 12.1
Es seien $(\Omega, \mathfrak{A})$, $\Omega', \mathfrak{A}')$ meßbare Räume. Eine Abbildung $T : \Omega \to \Omega'$ heißt $(\mathfrak{A}, \mathfrak{A}')$-meßbar, wenn $T^{-1}(A') = \{\omega \in \Omega \mid T(\omega) \in A'\} \in \mathfrak{A} \quad \forall A' \in \mathfrak{A}'$ gilt. Im Falle $(\Omega, \mathfrak{A}) = (\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$ nennen wir T *borel-meßbar*. Ist $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ $(\mathfrak{A}, \mathfrak{B})$-meßbar, so heißt X Zufallsvariable (ZV).

Hintereinanderschaltungen meßbarer Abbildungen sind meßbar:

Satz 12.2
Es seien $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$, $(\Omega'', \mathfrak{A}'')$ meßbare Räume, $S : \Omega \to \Omega'$ $(\mathfrak{A}, \mathfrak{A}')$-meßbar und $T : \Omega' \to \Omega''$ $(\mathfrak{A}', \mathfrak{A}'')$-meßbar. Dann ist $T \circ S$ $(\mathfrak{A}, \mathfrak{A}'')$-meßbar.

Beweis

$$\underbrace{\Omega \overset{S}{\to} \Omega' \overset{T}{\to}}_{T \circ S} \Omega''$$

Es gilt

$$(T \circ S)^{-1}(A) = S^{-1}(T^{-1}(A)),$$

also

$$\begin{aligned} A \in \mathfrak{A} \quad &\Longrightarrow \quad T^{-1}(A) \in \mathfrak{A}' \qquad (\text{da } T \ (\mathfrak{A}', \mathfrak{A}'')\text{-meßbar}) \\ &\Longrightarrow \quad S^{-1}(T^{-1}(A)) \in \mathfrak{A} \qquad (\text{da } S \ (\mathfrak{A}, \mathfrak{A}')\text{-meßbar}) \\ &\Longrightarrow \quad (T \circ S)^{-1}(A) \in \mathfrak{A}. \end{aligned}$$

q.e.d.

Beim Nachweis der Meßbarkeit kann man sich auf Erzeugendensysteme beschränken:

Satz 12.3
Es seien $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$ meßbare Räume, $T : \Omega \to \Omega'$ eine Abbildung und $\mathfrak{E} \subset \wp(\Omega')$ ein Erzeugendensystem.

Gilt dann $T^{-1}(\mathfrak{E}) \in \mathfrak{A} \quad \forall E \in \mathfrak{E}$, so ist T $(\mathfrak{A}, \mathfrak{A}')$-meßbar.

Beweis

Wir zeigen zunächst, daß $\mathfrak{A}_0 := \{A \subset \Omega' \mid T^{-1}(A) \in \mathfrak{A}\}$ eine σ-Algebra über Ω' ist:

$T^{-1}(\Omega') = \Omega \in \mathfrak{A}$, also gilt: $\Omega' \in \mathfrak{A}_0$.

Für alle $A \subset \Omega'$ gilt

$$T^{-1}(A^c) = \{\omega \in \Omega \mid T(\omega) \notin A\} = (\{\omega \in \Omega \mid T(\omega) \in A\})^c = (T^{-1}(A))^c$$

also:

$$\begin{aligned} A \in \mathfrak{A}_0 \quad &\Longrightarrow \quad T^{-1}(A) \in \mathfrak{A} \\ &\Longrightarrow \quad (T^{-1}(A))^c \in \mathfrak{A} \\ &\Longrightarrow \quad T^{-1}(A^c) \in \mathfrak{A} \\ &\Longrightarrow \quad A^c \in \mathfrak{A}. \end{aligned}$$

Analog erhält man mit $T^{-1}(\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} T^{-1}(A_n)$ die dritte definierende Eigenschaft einer σ-Algebra.

Da $\mathfrak{E} \subset \mathfrak{A}_0$ nach Voraussetzung, und $\mathfrak{A}' = \sigma(\mathfrak{E})$, folgt $\mathfrak{A}' \subset \mathfrak{A}_0$, d.h. $T^{-1}(A) \in \mathfrak{A} \quad \forall A \in \mathfrak{A}'$. *q.e.d.*

Korollar 12.4
Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X : \Omega \to \mathbb{R}$ ist genau dann eine Zufallsvariable, wenn $\{X \leq a\} \quad (:= \{\omega \in \Omega \mid X(\omega) \leq a\}) \quad \in \mathfrak{A}$ für alle $a \in \mathbb{R}$.

Beweis

Satz 12.3 und Satz 10.8. *q.e.d.*

Korollar 12.5
Stetige Abbildungen $f : \mathbb{R} \to \mathbb{R}$ sind Borel-meßbar.

Beweis

Stetigkeit heißt $f^{-1}(U)$ offen $\forall U$ offen. Verwende nun wieder Satz 12.3 und Satz 10.8. *q.e.d.*

12.2 Abbildungen mit Werten in $\mathbb{R}$ oder $\overline{\mathbb{R}}$

Erinnerung: $1_A : \Omega \to \mathbb{R}, 1_A(\omega) = \begin{cases} 1 & \omega \in A??? \\ 0 & \text{sonst} \end{cases}$.

Offensichtlich gilt: 1_A ist $(\mathfrak{A}, \mathfrak{B})$-meßbar $\iff$ $A \in \mathfrak{A}$.

Endliche Linearkombinationen meßbarer Indikatorfunktionen nennt man primitive Funktionen; dies ist die Klasse der meßbaren Funktionen, die nur endlich viele Werte annehmen. Wir schreiben $\mathrm{Prim}_+(\Omega, \mathfrak{A})$ für die Menge der nicht-negativen primitiven Funktionen auf $(\Omega, \mathfrak{A})$.

Lemma 12.6
Zu jeder $(\mathfrak{A}, \mathfrak{B})$-meßbaren, nicht-negativen Funktion $f : \Omega \to \mathbb{R}$ existiert eine Folge

$$(f_n)_{n \in \mathbb{N}} \subset \mathrm{Prim}_+(\Omega, \mathfrak{A})$$

mit der Eigenschaft $f_n \uparrow f$.

Beweis

Setze beispielsweise

$$f_n = \sum_{i=0}^{n2^n - 1} i 2^{-n} 1_{A_{ni}}$$

mit

$A_{ni} := \{\omega \in \Omega \mid i2^{-n} \leq f(\omega) < (i+1)2^{-n}\}$. *q.e.d.*

(wird bei der Konstruktion des Lebesgue-Integrals verwendet.)

Bemerkung 12.7
Es ist oft sinnvoll, die Werte $\pm\infty$ zuzulassen, d.h. den Wertebereich $\overline{\mathbb{R}}$; $f : \Omega \to \overline{\mathbb{R}}$ ist genau dann $(\mathfrak{A}, \overline{\mathfrak{B}})$-meßbar, wenn $\{f \leq a\} \in \mathfrak{A} \quad \forall a \in \mathbb{R}$ gilt.

Satz 12.8

(i) Sind $f, g : \Omega \to \mathbb{R}$ $(\mathfrak{A}, \mathfrak{B})$-meßbar, so liegen die Mengen $\{f < g\}$, $\{f \leq g\}$, $\{f = g\}$, $\{f \neq g\}$ in $\mathfrak{A}$. Hierbei steht $\{f < g\}$ für $\{\omega \in \Omega \mid f(\omega) < g(\omega)\}$ usw.

(ii) Seien $f, g : \Omega \to \mathbb{R}$ $(\mathfrak{A}, \mathfrak{B})$-meßbar und $\alpha, \beta \in \mathbb{R}$. Dann sind auch die Funktionen $\alpha f + \beta$, $f + g$, $f \cdot g$, $\max(f, g)$, $\min(f, g)$ $(\mathfrak{A}, \mathfrak{B})$-meßbar.

(iii) Ist $(f_n)_{n\in\mathbb{N}}$ eine Folge $(\mathfrak{A},\mathfrak{B})$-meßbarer Abbildungen, so sind $\sup_{n\in\mathbb{N}} f_n$, $\inf_{n\in\mathbb{N}} f_n$, $\limsup_{n\in\mathbb{N}} f_n$ und $\liminf_{n\in\mathbb{N}} f_n$ $(\mathfrak{A},\overline{\mathfrak{B}})$-meßbar (bzw. $(\mathfrak{A},\mathfrak{B})$-meßbar, wenn die jeweilige Funktion keinen der Werte $\pm\infty$ annimmt).

Gilt $f_n(\omega) \to f(\omega) \quad \forall\omega$ mit einem $f:\Omega\to\mathbb{R}$ (bzw. $f:\Omega\to\overline{\mathbb{R}}$ bei „bestimmter Divergenz"), so ist auch f $(\mathfrak{A},\mathfrak{B})$-meßbar (bzw. $(\mathfrak{A},\overline{\mathfrak{B}})$-meßbar).

Beweis

(i)

$$\{f<g\} = \underbrace{\bigcup_{q\in\mathbb{Q}}}_{\text{abzählbare Vereinigung}} \underbrace{f<q \quad \cap \quad g>q}_{f^{-1}((-\infty,q))\in\mathfrak{A},\text{ da } f \text{ meßbar}} .$$

Wegen $\{f\le g\} = \{g<f\}^c$, $\{f=g\} = \{f\le g\}\cap\{f<g\}^c$, $\{f\ne g\} = \{f=g\}^c$ liegen auch die anderen Mengen in $\mathfrak{A}$.

(ii) Die Abbildung $x\to\alpha x+\beta$ ist stetig, also ist $\alpha+\beta f$ als Hintereinanderschaltung meßbarer Abbildugen wieder meßbar (Satz 12.2 und Korollar 12.5).

Weiter gilt

$$\{f+g\le a\} = \{f \le \underbrace{a-g}_{\text{meßbar, da affine Transformation}}\} \in\mathfrak{A} \quad \forall a\in\mathbb{R}.$$

Mit $f\cdot g = \frac{1}{4}((f+g)^2-(f-g)^2)$ folgt dann die Meßbarkeit von $f\cdot g$ (hierbei verwendet: $x\to x^2$ stetig, also Borel-meßbar).

Mit $\{\max(f,g)\le a\} = \{f\le a\}\cap\{g\le a\}$, $\{\min(f,g)\le a\} = \{f\le a\}\cup\{g\le a\}$ folgt die Meßbarkeit von $\max(f,g)$ und $\min(f,g)$.

(iii) Wir verwenden Bemerkung 12.7.

$$\{\sup_{n\in\mathbb{N}} f_n \le a\} = \bigcap_{n=1}^{\infty}\{f_n\le a\}\in\mathfrak{A}$$

liefert die Meßbarkeit von $\sup_{n\in\mathbb{N}} f_n$. Die Meßbarkeit der anderen Abbildungen folgt mit

$$\inf_{n\in\mathbb{N}} f_n = -\sup_{n\in\mathbb{N}}(-f_n),$$
$$\limsup_{n\in\mathbb{N}} f_n = \inf_{n\in\mathbb{N}}\sup_{m\ge n} f_m,$$
$$\liminf_{n\to\infty} f_n = \sup_{n\in\mathbb{N}}\inf_{m\ge n} f_m.$$

Konvergiert f_n gegen f, so gilt $f=\limsup_{n\to\infty} f_n$.

q.e.d.

Bei der Verallgemeinerung solcher Aussagen auf $\overline{\mathbb{R}}$ ist wieder zu beachten, daß durch geeignete Voraussetzungen Ausdrücke der Form $\infty-\infty$ oder $\infty\cdot(-\infty)$ vermieden werden.

27. Oktober 1994

12.3 Bildmaße, Verteilungen und Verteilungsfunktionen

Verallgemeinerung von Satz & Def. 6.3

Satz & Definition 12.9
Es seien $(\Omega, \mathfrak{a}, \mu)$ ein Maßraum, $(\Omega', \mathfrak{A}')$ ein meßbarer Raum und $T : \omega \to \Omega'$ eine $(\mathfrak{A}, \mathfrak{A}')$-meßbare Abbildung.

Dann ist $\mu' : \mathfrak{A}' \to [0, \infty]$, $\mu'(\mathfrak{A}') := \mu(T^{-1}(A')) \quad \forall A' \in \mathfrak{A}'$ ein Maß auf $(\Omega', \mathfrak{A}')$, das Bildmaß von μ unter T, Bezeichnung: $\mu' = \mu^T$.

Beweis

$\mu'(\emptyset) = \mu(T^{-1}(\emptyset)) = \mu(\emptyset) = 0.$

Ist $(A'_n)_{n\in\mathbb{N}}$ eine disjunkte Folge in $\mathfrak{A}'$, so ist wegen der Meßbarkeit von T und $T^{-1}(A \cap B) = T^{-1}(A) \cap T^{-1}(B)$ die Folge $(A_n)_{n\in\mathbb{N}}$ mit $A_n := T^{-1}(A'_n)$ eine disjunkte Mengenfolge in $\mathfrak{A}'$, also

$$\begin{aligned}
\mu'(\bigcup_{n=1}^{\infty} A'_n) &= \mu(T^{-1}(\bigcup_{n=1}^{\infty} A'_n)) \\
&= \mu(\bigcup_{n=1}^{\infty} \underbrace{T^{-1}(A'_n)}_{A_n}) \\
&= \sum_{n=1}^{\infty} \mu(T^{-1}(A'_n)) \\
&= \sum_{n=1}^{\infty} \mu'(A'_n).
\end{aligned}$$

q.e.d.

Beispiel 12.10
Sei $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}, \mathfrak{B}, \ell)$, $(\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$, $T : \mathbb{R} \to \mathbb{R}$ werde def. durch $T(x) = ax + b$ ($a \neq 0$, b fest).

Dann gilt für alle $c, d \in \mathbb{R}$ mit $c \leq d$, im Falle $a > 0$:

$$\ell^T((c,d]) = \ell(\{x \in \mathbb{R} \mid c < ax + b \leq d\} = \ell((\frac{c-b}{a}, \frac{d-b}{a}]) = \frac{d-c}{a} = \frac{1}{a}\ell(c,d])$$

und im Falle $a < 0$:

$$\ell^T((c,d]) = \frac{1}{-a}\ell((c,d]),$$

also insgesamt

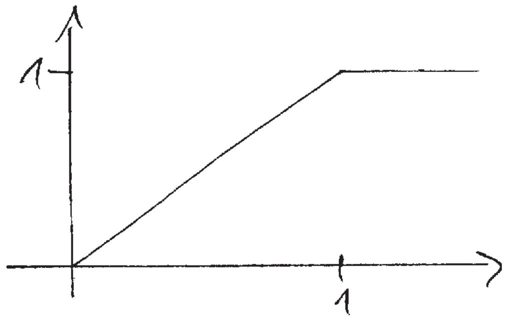
$$\ell^T(I) = \frac{1}{|a|}\ell(I)$$

für alle I aus einem $\cap$-stabilen Erzeugendensystem von $\mathfrak{B}$ etc.: Satz 11.8 liefert $\ell^T = \frac{1}{|a|}\ell$.

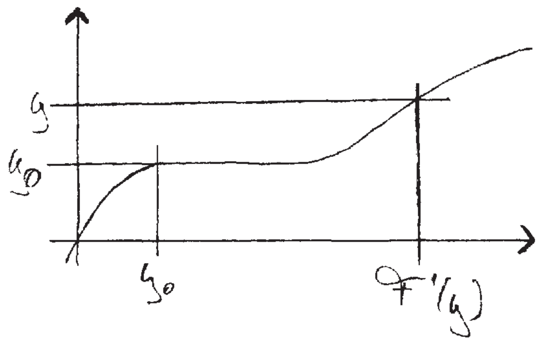
Definition 12.11
$(\Omega, \mathfrak{A}, P)$ sei Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}$ Zufallsvariable. Dann heißt P^X die *Verteilung* von X die Funktion $F : \mathbb{R} \to \mathbb{R}$, $F(x) = P(X \leq x) \quad (= P^X((-\infty, x]))$ heißt die *Verteilungsfunktion* zu X.

„Verteilung" häufig synonym für „Wahrscheinlichkeitsmaß", man spricht auch von der Verteilungsfunktion eines Wahrscheinlichkeitsmaßes auf $(\mathbb{R}, \mathfrak{B})$.

Da $\{(-\infty, x] \mid x \in \mathbb{R}\}$ ein $\cap$-stablies Erzeugendensystem von $\mathfrak{B}$ ist, wird die Verteilung einer Zufallsvariablen bzw. ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$ durch die zugehörige Verteilungsfunktion eindeutig bestimmt (Korollar 11.9).



Beispiel 12.12
Ein Stab der Länge 1 wird so zerbrochen, daß die Position der Bruchstelle über der Länge des Stabes gleichverteilt ist. Welche Verteilung hat die Länge des kürzeren Stücks?



$Y = \min(X, 1 - X)$.

$0 < x \leq \frac{1}{2}$, dann gilt:

$$\begin{aligned} F_Y(x) = P(Y \leq x) &= P(\min(X, 1-X) \leq x) \\ &= \ell(\{\omega \in (0,1) \mid \omega \leq x\} \cup \{\omega \in (0,1) \mid 1 - \omega \leq x\}) \\ &= \ell((0, x] + [1 - x, 1)) \\ &= 2x. \end{aligned}$$

Außerdem gilt $P(Y \leq x) = 0$ für $x < 0$, 1 für $x > 1$, also:

$$F_Y(x) = \begin{cases} 1, & \text{wenn } x > \frac{1}{2} \\ 2x, & \text{wenn } 0 \leq x \leq \frac{1}{2} \\ 0, & \text{wenn } x < 0. \end{cases}$$

Dies ist die Verteilungsfunktion zu $U(0, \frac{1}{2})$, wir haben also bewiesen: $Y \sim U(0, \frac{1}{2})$.

Satz 12.13
Eine Verteilungsfunktion F hat die Eigenschaften

(i) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$,

(ii) F ist (schwach) monoton steigend,

(iii) F ist stetig von rechts.

Beweis

Sei P Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$ mit Verteilungsfunktion F.

(i) Sei $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ mit $x_n \to -\infty$. Setze $y_n := \sup_{m\geq n} x_m$, dann $y_n \downarrow -\infty$, also

$$0 \leq F(x_n) \leq F(y_n) = P((-\infty, \underbrace{y_n}_{\text{geht antiton gegen } \emptyset}]) \overset{n\to\infty}{\to} P(\emptyset) = 0.$$

Die andere Aussage folgt analog mit der Stetigkeit von unten.

(ii) Sei $x \leq y$. Dann:

$$F(x) = P((-\infty, x]) \leq P((-\infty, x]) = F(y).$$

(iii) Ist $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ mit $x_n > 0\ \forall n$, $x_n \to x$, so gilt $y_n \downarrow x$ mit $y_n := \sup_{m\geq n} x_m$, also

$$\begin{aligned} F(x) &= P((-\infty, x]) \leq P((-\infty, x_n]) \\ &= F(x_n) \leq F(y_n) \\ &= P(\underbrace{(-\infty, y_n]}_{\downarrow(-\infty,x]}) \to P((-\infty, x]) \\ &= F(x). \end{aligned}$$

q.e.d.

Definition 12.14
Die zu einer Verteilungsfunktion F gehörende Quantil-Funktion F^{-1} wird definiert durch

$$F^{-1} : (0,1) \to \mathbb{R}, F^{-1}(y) := \inf\{x \in \mathbb{R} \mid F(x) \geq y\}.$$

Lemma 12.15

$$y \leq F(x) \iff F^{-1}(y) \leq x.$$

Beweis
$y \leq F(x) \implies F^{-1}(y) \leq x$ folgt unmittelbar aus der Definition von F^{-1}. Da außerdem $F(x) < y \implies F(x + \frac{1}{n}) < y$ für ein $n \in \mathbb{N}$ (da F stetig von rechts) $\implies F^{-1}(y) \geq x + \frac{1}{n}$ (da F schwach monoton steigend) $\implies F^{-1}(y) > x$, gilt $F^{-1}(y) \leq X \implies F(x) \geq y$. *q.e.d.*

Satz 12.16
Es sei $F : \mathbb{R} \to \mathbb{R}$ schwach monoton steigend, stetig von rechts und es gelte $F(x) \to 0$ mit $x \to -\infty$, $F(x) \to 1$ mit $x \to +\infty$.

Dann existiert ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathfrak{B})$ derart, daß F die Verteilungsfunktion zu P ist.

Beweis
Sei $\Omega = (0,1)$, $\mathfrak{A} = \mathfrak{B}|_{(0,1)}$, $P_0 = \ell|_{(0,1)}$. $X : \Omega \to \mathbb{R}$ werde definiert durch $X(\omega) := F^{-1}(\omega)$. (Meßbarkeit folgt aus Monotonie: Übungsaufgabe.)
Lemma 12.15 liefert für $P := P_0^X$:

$$\begin{aligned} P((-\infty, x]) &= P_0(X \le x) \\ &= P_0(\{\omega \in \Omega \mid \underbrace{F^{-1}(\omega) \le x}_{\iff \quad \omega \le F(x)} \}) \\ &= \ell((0, F(x)]) = F(x). \end{aligned}$$

P hat also Verteilungsfunktion F. *q.e.d.*

Wir haben sogar etwas mehr gezeigt: Zu jedem F mit den aufgeführten Eigenschaften existiert eine Zufallsvariable X mit Verteilungsfunktion F.

Bemerkung 12.17
Die in Abschnitt 11.3 durchgeführte Konstruktion des Lebesgue-Maßes ℓ auf $(\mathbb{R}, \mathfrak{B})$ läßt sich wie folgt verallgemeinern: Ist $G : \mathbb{R} \to \mathbb{R}$ eine schwach monoton steigende und rechtsstetige Funktion, so existiert genau ein Maß μ_G auf $(\mathbb{R}, \mathfrak{B})$ mit

$$\mu_G((a,b]) = G(b) - G(a) \quad \forall a, b \in \mathbb{R}, a \le b$$

Man nennt μ_G das *Lebesgue-Stieltjes-Maß* mit maßdefinierender Funktion G. Satz 12.16 ist ein Spezialfall dieser Aussage, der hier gewählte Weg über Bildmaße und die Quantiltransformation ist für Simulationsverfahren von großer Bedeutung.

1. November 1994

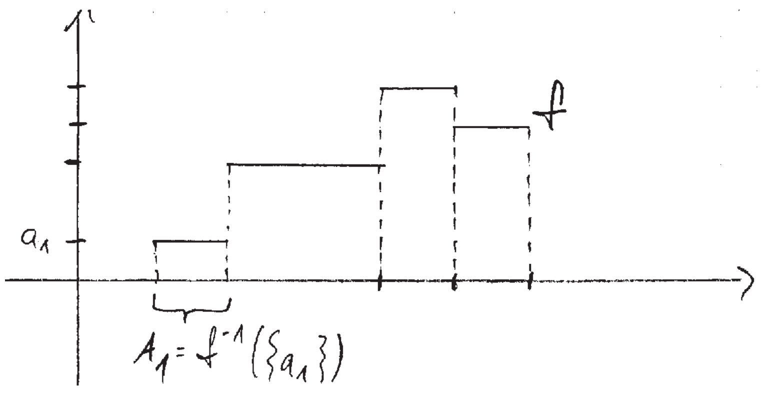
13 Integrale und Erwartungswerte

13.1 Das Lebesgue-Integral

Es sei $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum. Für eine möglichst große Klasse $(\mathfrak{A}, \mathfrak{B})$-meßbarer Funktionen $f : \Omega \to \mathbb{R}$ soll $\int f \, d\mu$ das (Lebesgue-)Integral von f bezüglich μ definiert werden.

Definition 13.1 (Stufe 1)
Für $f \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$ sei

$$\int f d\mu := \sum_{x \in f(\Omega)} x \cdot \mu(f^{-1}(\{x\}))$$



$f = \sum a_i 1_{A_i}$
$\int f \, d\mu = \sum a_i \mu(A_i)$

Beachte:

(i) Die Summe ist wegen $\# f(\Omega) < \infty$ endlich.

(ii) $0 \cdot \infty = 0$,

(iii) $f^{-1}(\{x\}) \in \mathfrak{A}$. Im Falle $f = 1_A$ erhält man $\int f \, d\mu = \mu(A)$. Insbesondere: $\int c \, d\mu = c \cdot \mu(\Omega)$, $\int 0 \, d\mu = 0$.

Lemma 13.2
Es seien $f, g \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$ und $a \geq 0$. Dann gilt:

$$af, \ f + g \in \mathrm{Prim}_+(\Omega, \mathfrak{A}) \text{ sowie}$$
$$\int (af) d\mu = a \int f d\mu,$$
$$\int (f+g) d\mu = \int f d\mu + \int g d\mu,$$
$$f \leq g \implies \int f d\mu \leq \int g d\mu.$$

Beweis
Die Funktionen af, $f + g$ sind wieder nicht-negativ, $(\mathfrak{A}, \mathfrak{B})$-meßbar und haben endlichen Wertebereich, sind also in $\mathrm{Prim}_+(\Omega, \mathfrak{A})$. Im Falle $a = 0$ gilt $af \equiv 0$, also

$$\int (af) \, d\mu = \int 0 \, d\mu = 0 = 0 \cdot \int f \, d\mu.$$

Im Falle $a > 0$, $f(\Omega) = \{b_1, \ldots, b_n\}$ gilt $(af)(\Omega) = \{ab_1, \ldots, ab_n\}$, also

$$\begin{aligned} \int (af) d\mu = \sum_{i=1}^{n} (ab_i) \mu((af)^{-1}(\{ab_i\})) &= a \sum_{i=1}^{n} b_i \mu(f^{-1}(\{b_i\})) \\ &= a \int f d\mu. \end{aligned}$$

Hat man $f(\Omega) = \{a_1, \ldots, a_n\}$, $g(\Omega) = \{b_1, \ldots, b_m\}$, so gilt mit $A_i := f^{-1}(\{a_i\})$ und $B_i := g^{-1}(\{b_i\})$:

$$f := \sum_{i=1}^{n} a_i 1_{A_i}, \quad g := \sum_{j=1}^{m} b_j 1_{B_j}$$

mit $A_1, \ldots, A_n$ disjunkt, $B_1, \ldots, B_m$ disjunkt, $\sum_{i=1}^{n} (A_i) = \Omega$, $\sum_{j=1}^{m} (B_j) = \Omega$.
Auf $A_i \cap B_j$ nimmt $f + g$ den Wert $a_i + b_j$ an, also

$$\begin{aligned} \int f \, d\mu + \int g \, d\mu = \sum_{i=1}^{n} a_i \mu(A_i) + \sum_{j=1}^{m} b_j \mu(B_j) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j) \mu(A_i \cap B_j) \\ &= \sum_{k=1}^{l} c_k \mu(C_k) = \int (f+g) \, d\mu \end{aligned}$$

mit $\{c_1, \ldots, c_l\} = (f+g)(\Omega) = \{a_i + b_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$, $C_k = (f+g)^{-1}(\{c_k\})$, denn

$$\mu((f+g)^{-1}(\{c_k\})) = \sum_{\substack{1\leq i\leq n,\ 1\leq j\leq m \\ a_i+b_j=c_k}} \mu(A_i \cap b_j).$$

Hat man schließlich $f = \sum_{i=1}^n a_i 1_{A_i} \leq g = \sum_{j=1}^m b_j 1_{B_j}$, so gilt für alle i, j:

$$A_i \cap B_j \neq \emptyset \implies a_i \leq b_j.$$

Damit:

$$\begin{aligned} \int f\, d\mu = \sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^n \sum_{j=1}^m a_i \mu(A_i \cap B_j) \quad &\leq \quad \sum_{i=1}^n \sum_{j=1}^m b_j \mu(A_i \cap B_j) \\ &= \quad \sum_{j=1}^m b_j \mu(B_j) = \int g\, d\mu \end{aligned}$$

q.e.d.

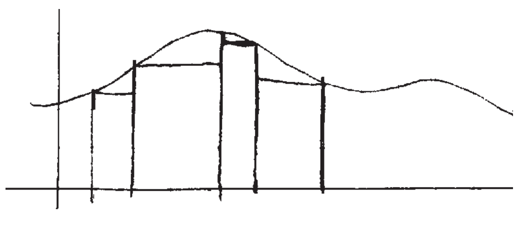
Das Lemma zeigt, daß $f \to \int f\, d\mu$ auf $\mathrm{Prim}_+(\Omega, \mathfrak{A})$ linear und isoton ist. Insbesondere gilt für alle Funktionen f der Form $f = \sum_{i=1}^n a_i 1_{A_i}$ mit $a_i \geq 0$, $A_i \in \mathfrak{A}$

$$\int f\, d\mu = \sum_{i=1}^n a_i \mu(A_i),$$

auch wenn $A_1, \ldots, A_n$ nicht disjunkt sind und die a_i nicht voneinander verschieden sind, d.h. alle solche Darstellungen können direkt bei der Berechnung von $\int f\, d\mu$ verwendet werden.

Definition 13.3 (Stufe 2)
Das μ-Integral einer $(\mathfrak{A}, \mathfrak{B})$-meßbaren nicht-negativen Funktion $f : \Omega \to \mathbb{R}$ wird definiert durch

$$\int f\, d\mu := \sup \left\{ \int g\, d\mu \mid g \in \mathrm{Prim}_+(\Omega, \mathfrak{A}),\ g \leq f \right\}$$



Beachte: $g \equiv 0 \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$, $g \leq f$, also „sup nicht-leer".

Außerdem ist wegen der Isotonie des Integrals auf $\mathrm{Prim}_+(\Omega, \mathfrak{A})$ Stufe 2 mit Stufe 1 verträglich: Im Falle $f \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$ taucht f selbst in der Menge auf, über die das Supremum gebildet wird.

Aus Lemma 12.6 ist bekannt: Zu jedem f aus Definition 13.3 existiert $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Prim}_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f$.

Lemma 13.4
Ist $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Prim}_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f$, so gilt

$$\int f\, d\mu = \lim_{n\to\infty} \int f_n\, d\mu.$$

Beweis

„$\geq$“: folgt unmittelbar aus der Definition.

„$\leq$“: Sei $g \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$, $g(\Omega) = \{a_1, \dots, a_k\}$, $g = \sum_{i=1}^k a_i 1_{A_i}$, $g \leq f$. Sei $0 < \eta < 1$, $B_n := \{\omega \in \Omega \mid \eta g(\omega???) \leq f_n(\omega)\}$.

Aus $\eta \cdot g < f$ auf $\{f > 0\}$ und $f_n = f$ auf $\{f = 0\}$ sowie $f_n \uparrow f$ folgt $B_n \uparrow \Omega$, als mit der Stetigkeit von unten:

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

Also mit $\eta \uparrow 1$

$$\int g \, d\mu \leq \lim_{n\to\infty} \int f_n \, d\mu.$$

Bildet man das Supremum über die zugelassenen Funktionen g, so folgt $\int f \, d\mu \leq \lim_{n\to\infty} \int f_n \, d\mu$. *q.e.d.*

Lemma 13.5
Es seien $f, g : \Omega \to \mathbb{R}$ $(\mathfrak{A}, \mathfrak{B})$-meßbare nicht-negative Funktionen und $a \geq 0$. Dann gilt:

$$\begin{aligned}\int (af) \, d\mu &= a \int f \, d\mu, \\ \int (f+g) \, d\mu &= \int f \, d\mu + \int g \, d\mu, \\ f \leq g \implies \int f \, d\mu &\leq \int g \, d\mu.\end{aligned}$$

Beweis

Lemma 12.6 liefert die Existenz von (f_n), $(g_n)_{n\in\mathbb{N}} \subset \mathrm{Prim}_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f$ und $g_n \uparrow g$. Klar:

$$(f_n + g_n)_{n\in\mathbb{N}} \subset \mathrm{Prim}_+(\Omega, \mathfrak{A}), \quad f_n + g_n \uparrow f + g.$$

Verwende nun die Linearität auf $\mathrm{Prim}_+(\Omega, \mathfrak{A})$ und dreimal Lemma 13.4:

$$\begin{aligned}\int (f+g) \, d\mu &= \lim_{n\to\infty} \int (f_n + g_n) \, d\mu \\ &= \lim_{n\to\infty} (\int f_n \, d\mu + \int g_n \, d\mu) \\ &= \lim_{n\to\infty} \int f_n \, d\mu + \lim \int g_n \, d\mu \\ &= \int f \, d\mu + \int g \, d\mu.\end{aligned}$$

Analog:

$$\int (af)\, d\mu = \lim_{n\to\infty} \int (af_n)\, d\mu = a \lim_{n\to\infty} \int f_n\, d\mu = a \int f\, d\mu.$$

Isotonie folgt aus der Definition, denn bei g wird das Supremum über eine größere Menge gebildet. *q.e.d.*

Jede beliebige meßbare Funktion $f : \Omega \to \mathbb{R}$ läßt sich wie folgt als Differenz zweier nicht-negativer meßbarer Funktionen darstellen:

$$\underbrace{f^+ = \max\{f, 0\}}_{\text{Positivteil}}, \quad \underbrace{f^- := -\min\{f, 0\}}_{\text{Negativteil}}, \quad f = f^+ - f^-$$

Beachte: $f^+ + f^- = |f|$.

Definition 13.6 (Stufe 3)
Eine $(\mathfrak{A}, \mathfrak{B})$-meßbare Funktion $f : \Omega \to \mathbb{R}$ heißt , wenn gilt:

$$\int f^+\, d\mu < \infty \quad \{textund \quad \int f^-\, d\mu < \infty.$$

Das μ-Integral einer solchen Funktion wird definiert durch

$$\int f\, d\mu := \int f^+\, d\mu - \int f^-\, d\mu.$$

Andere Schreibweise: $\int f(\omega)\mu(d\omega)$, $\int f(\omega)\, d\mu(\omega)$ oder ganz modern: $\mu(f)$.

Bemerkung 13.7
Gilt mindestens eine der Bedingungen $\int f^+\, d\mu < \infty$, $\int f^-\, d\mu < \infty$, so läßt sich $\int f\, d\mu$ noch sinnvoll als Differenz der beiden definieren; solche Funktionen werden auch „quasi-integrierbar“ oder „integrierbar im erweiterten Sinne“ genannt.

Aufgepaßt: Für $f \equiv 1$ gilt $\int f\, d\ell = \infty$, also f ist quasi-integrierbar, aber nicht integrierbar.

Unmittelbar aus Def. 13.6 und $|f| = f^+ + f^-$ folgt

$$f\mu\text{-integrierbar} \iff \int |f|\, d\mu < \infty; \qquad (*)$$
$$\left| \int f\, d\mu \right| \leq \int |f|\, d\mu.$$

Satz 13.8
Es seien $f, g : \Omega \to \mathbb{R}$ μ-integrierbar und $a \in \mathbb{R}$. Dann sind auch af und $f + g$ μ-integrierbar und es gilt

$$\int (af)\, d\mu = a \int f\, d\mu$$
$$\int (f + g)\, d\mu = \int f\, d\mu + \int g\, d\mu,$$
$$f \leq g \implies \int f\, d\mu \leq \int g\, d\mu.$$

Beweis

Wegen $|f+g| \le |f| + |g|$ folgt Integrierbarkeit von $f+g$ unmittelbar aus $(*)$. $(f+g)^+$ ist im allgemeinen verschieden von $f^+ + g^+$. Es gilt jedoch

$$(f+g)^+ - (f+g)^- = f+g = f^+ - f^- + g^+ - g^-,$$

also

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+.$$

Wende Lemma 13.5 an:

$$\int (f+g)^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu = \int (f+g)^- \, d\mu + \int f^+ \, d\mu + \int g^+ \, d\mu,$$

also

$$\begin{aligned} \int (f+g) \, d\mu &= \int (f+g)^+ \, d\mu - \int (f+g)^- \, d\mu \\ &= \int f^+ \, d\mu + \int g^+ \, d\mu - \int f^- \, d\mu - \int g^- \, d\mu \\ &= \int f \, d\mu - \int g \, d\mu \end{aligned}$$

Der Beweis zu $\int (af) \, d\mu = a \cdot \int f \, d\mu$ dürfte klar sein (betrachte $a \ge 0$, $a < 0$ separat)

Beweis der Monotonieeigenschaft: Übungsaufgabe. *q.e.d.*

3. November 1994

Beispiel 13.9

(i) Ist $(\Omega, \mathfrak{A})$ ein meßbarer Raum, $\omega \in \Omega$, $f : \Omega \to \mathbb{R}$ $(\mathfrak{A}, \mathfrak{B})$-meßbar, so gilt

$$\int f \, d\delta_\omega = f(\omega).$$

$(f = 1_A$: $\int f \, d\delta_\omega = \delta_\omega(A) = 1_A(\omega) = f(\omega))$

(ii) $(\Omega, \mathfrak{A})$ meßbarer Raum mit $\{\omega\} \in \mathfrak{A} \quad \forall \omega \in \Omega$. Ein diskretes Maß[1] μ auf $(\Omega, \mathfrak{A})$ läßt sich schreiben in der Form

$$\mu = \sum_{i=1}^{\infty} \alpha_i \delta_{\omega_i} \quad \text{mit} \quad \alpha_i \ge 0.$$

Eine Funktion $f : \Omega \to \mathbb{R}$ ist genau dann μ-intgrierbar, wenn $\sum \alpha_i |f(\omega_i)| < \infty$ gilt; dann

$$\int f \, d\mu = \sum_{\omega \in A} f(\omega) \mu(\{\omega\}).$$

Im Spezialfall $\Omega = \mathbb{N}$, $\mathfrak{A} = \wp(\mathbb{N})$, $\mu = \sum_{n=1}^{\infty} \delta_n$ ($\mu(A) = \#A$, das abzählende Maß) erhält man $\int f \, d\mu = \sum_{n=1}^{\infty} f(n)$, ergibt sich ein Zusammenhang zu unendlichen Reihen.

[1] d.h. ein auf eine abzählbare Menge $A \in \mathfrak{A}$ konzentriertes Maß

(iii) Es sei $\Omega = [a,b]$, $-\infty < a < b < \infty$, $\mathfrak{A}$ die Spur von $\mathfrak{B}$ auf Ω, $\mu = \ell|_{[a,b]}$. Ist $f : \Omega \to \mathbb{R}$ Borel-meßbar und Riemann-integrierbar, so ist f auch μ-integrierbar und es gilt

$$\int f \, d\mu = \int_a^b f(x) \, dx.$$

Die Funktion $1_{\mathbb{Q} \cap [a,b]}$ ist ein Beispiel für eine μ-integrierbare Funktion, die nicht Riemann-integrierbar ist.

Grob: Die bekannten Formeln für das eigentliche Riemann-Integral lassen sich bei ℓ verwenden (Details in [Ba/Bi]). Wir schreiben oft „dx“ anstelle von „$\ell(dx)$“.

13.2 Die „übliche Maschinerie“ mit Anwendungen

Aussagen über Integrale werden oft zunächst für Indikatorfunktionen, dann für primitive Funktionen, dann für nicht-negative und schließlich für allgemeine Funktionen bewiesen.

Satz 13.10 (Transformationssatz)
Es sei $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum, $(\Omega', \mathfrak{A}')$ ein meßbarer Raum und $T : \Omega \to \Omega'$ $(\mathfrak{A}, \mathfrak{A}')$-meßbar. Dann gilt für alle $(\mathfrak{A}', \mathfrak{B})$-meßbaren ??? $f : \Omega' \to \mathbb{R}$ ist genau dann μ^T-integrierbar, wenn $f \circ T$ μ-integrierbar ist und es gilt

$$\int f \circ T \, d\mu = \int f \, d\mu^T.$$

Bild 1

Beweis

Im Falle $f = 1_A$ gilt:

$$\int f \, d\mu^T = \mu^T(A) = \mu(T^{-1}(A)) = \int \underbrace{1_{T^{-1}(A)}}_{1_A \circ T} \, d\mu = \int f \circ T \, d\mu,$$

die Behauptungen sind also für Indikatorfunktionen richtig.

Mit der Linearität des Integrals folgt nun die behauptete Formel für $f \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$.

Ist $f \geq 0$ (und meßbar), so existiert nach Lemma 12.6 und Lemma 13.4 eine Folge von primitiven Funktionen $(f_n)n \in \mathbb{N} \subset \mathrm{Prim}_+(\Omega', \mathfrak{A}')$ mit $f_n \uparrow f$ und

$$\int f \, d\mu^T = \lim_{n \to \infty} \int f_n \, d\mu^T.$$

Mit $(f_n \circ T)_{n \in \mathbb{N}} \subset \mathrm{Prim}_+(\Omega, \mathfrak{A})$, $f_n \circ T \uparrow f \circ T$ und Lemma 13.4 folgt

$$\int f \, d\mu^T = \lim_{n \to \infty} \int f_n \circ T \, d\mu = \int f \circ T \, d\mu.$$

Ist nun $f : \Omega' \to \mathbb{R}$ $(\mathfrak{A}', \mathfrak{B})$-meßbar, so gilt nach dem bereits gezeigten

$$\int f^{\pm} \, d\mu^T < \infty \quad \Longleftrightarrow \quad \int f^{\pm} \circ T \, d\mu < \infty.$$

Mit $(f \circ T)^+ = f^+ \circ T$, $(f \circ T)^- = f^- \circ T$ folgt

$$f \ \ \mu^T\text{-integrierbar} \quad \Longleftrightarrow \quad f \circ T \ \ \mu\text{-integrierbar},$$

denn

$$\begin{aligned}
\int f \, d\mu^T &= \int f^+ \, d\mu^T - \int f^- \, d\mu^T \\
&= \int f^+ \circ T \, d\mu - \int f^- \circ T \, d\mu \\
&= \int (f \circ T)^+ \, d\mu - \int (f \circ T)^- \, d\mu \\
&= \int f \circ T \, d\mu.
\end{aligned}$$

q.e.d.

Definition 13.11
Es sei $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum, für jedes $\omega \in \Omega$ sei $A(\omega)$ eine Aussage. Man sagt „A gilt μ-fast überall“ (μ-f.ü.), wenn es eine μ-Nullmenge N (d.h. $N \in \mathfrak{A}$, $\mu(N) = 0$) gibt mit $A(\omega)$ wahr $\forall \omega \in N$ (bei Wahrscheinlichkeitsmaßen: P-fast sicher, P-f.s.)

Bild 2

Satz 13.12
Sei $(\Omega, \mathfrak{A}, \mu)$ Maßraum, $f, g : \Omega \to \mathbb{R}$ $(\mathfrak{A}, \mathfrak{B})$-meßbar.

(i) Sei $f \geq 0$. Dann

$$\int f \, d\mu = 0 \quad \Longleftrightarrow \quad f = 0 \ \ \mu\text{-f.ü.}$$

(ii) Ist f integrierbar und gilt $f = g$ μ-f.ü., so ist auch g μ-integrierbar und es gilt

$$\int f \, d\mu = \int g \, d\mu.$$

Beweis

(i) Es sei $N := \{f \neq 0\}$. Annahme: $\int f \, d\mu = 0$. Setze $A_n := \{f \geq \frac{1}{n}\}$. Dann gilt $A_n \uparrow N$, also $\mu(N) = \lim_{n\to\infty} \mu(A_n)$. Wegen

$$0 = \int f \, d\mu \geq \int \frac{1}{n} 1_{A_n} \, d\mu = \frac{1}{n} \mu(A_n) \geq 0$$

gilt: $\mu(A_n) = 0 \quad \forall n$, d.h. $\mu(N) = 0$ und damit $f = 0$ μ-f.ü.
Ist umgekehrt N eine μ-Nullmenge mit $f(\omega) = 0 \quad \forall \omega \notin N$ und $g = \sum_{i=1}^n a_i 1_{A_i} \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$ mit $g \leq f$, $\{a_1, \ldots, a_n\} = g(\Omega)$, so folgt $A_i \subset N$ für alle i mit $a_i > 0$, also

$$\int g \, d\mu = \sum_{\substack{i=1 \\ a_i > 0}}^{n} a_i \mu(A_i) \leq \sum_{\substack{i=1 \\ a_i > 0}}^{n} a_i \mu(N) = 0.$$

Hieraus folgt mit Def. 13.3:

$$\int f \, d\mu = 0.$$

(ii) Es seien zunächst $f, g \geq 0$; $N := \{f \neq g\}$. Dann gilt

$$\int f \, d\mu = \underbrace{\int f 1_N \, d\mu}_{=0 \text{ (siehe (i))}} + \int \underbrace{f 1_{N^c}}_{g 1_{N^c}} \, d\mu = \int g 1_n \, d\mu + \int g 1_{N^c} \, d\mu = \int g \, d\mu.$$

Beachte wieder, daß die Aussage $\int f \, d\mu < \infty \iff \int g \, d\mu < \infty$ einschließt — dies liefert die Aussage über die Integrierbarkeit.

Es seien nun $f, g : \Omega \to \mathbb{R}$ μ-f.ü., gleich und integrierbar. Dann gilt $f^+ = g^+$ μ-f.ü, $f^- = g^-$ μ-f.ü. und damit

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu = \int g^+ \, d\mu - \int g^- \, d\mu = \int g \, d\mu$$

q.e.d.

13.3 Konvergenzsätze

Das Maßintegral hat angenehme Eigenschaften bei der Behandlung von Funktionenfolgen. Es sei $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum, $f, f_1, f_2, \dots : \Omega \to \mathbb{R}$ $(\mathfrak{A}, \mathfrak{B})$-meßbare Funktionen.

Satz 13.13 (Satz von B. Levi, Satz von der monotonen Konvergenz)
Sind $f, f_1, f_2, \dots$ nicht-negativ mit $f_n \uparrow f$, so gilt

$$\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu \quad (= \sup_{n \in \mathbb{N}} \int f_n \, d\mu).$$

Beweis

Zu jedem f_n existiert eine Folge $(g_{nm}) \subset \mathrm{Prim}_+(\Omega, \mathfrak{A})$ mit $g_{nm} \uparrow f_n$ mit $m \to \infty$.

$$\begin{array}{ccccccccc}
g_{11} & \leq & g_{12} & \leq & g_{13} & \leq & \cdots & \to & f_1 \\
 & & & & & & & & \leq \\
g_{21} & \leq & g_{22} & \leq & g_{23} & \leq & \cdots & \to & f_2 \\
 & & & & & & & & \leq \\
g_{31} & \leq & g_{32} & \leq & g_{33} & \leq & \cdots & \to & f_3 \\
\vdots & & \vdots & & \vdots & & & & \downarrow \\
 & & & & & & & & f
\end{array}$$

Setze $h_m := \max\{g_{1m}, \dots, g_{mm}\}$. Klar ist: $h_m \in \mathrm{Prim}_+(\Omega, \mathfrak{A})$ $\forall m$, $h_m \uparrow$???. Für $n < m$ gilt $g_{nm} \leq h_m$, also folgt:

$$f_n = \sup_m g_{nm} = \sup_{m \geq n} g_{nm} \leq \sup_{m \in \mathbb{N}} h_m.$$

Also:

$$f \leq \sup_m h_m.$$

Außerdem: $h_m \leq f_m \leq f$, insgesamt: $h_m \uparrow f$. Lemma 13.4 liefert: $\int f \, d\mu = \lim_{n \to \infty} \int h_m \, d\mu$, also folgt mit

$$\int h_m \, d\mu \leq \int f_m \, d\mu \leq \int f \, d\mu$$

die Behauptung. *q.e.d.*

8. November 1994

$0 \leq f_n \uparrow f$ Levi: $\int \sup = \sup \int$

Lemma 13.14 (von Fatou)
Gilt $f_n \geq 0$ für alle n, $f = \liminf f_n$, so folgt

$$\int f \, d\mu \leq \liminf f_n \, d\mu$$

Beweis

Setze $g_n := \int_{m \geq n} f_m$, dann gilt $g_n \uparrow f$. $g_n \leq f_n$, also folgt mit Satz 13.13:

$$\int f \, d\mu = \lim_{n \to \infty} \int g_n \, d\mu \leq \liminf \int f_n \, d\mu.$$

q.e.d.

Bemerkung 13.15 (siehe auch Bemerkung 12.6 und 13.7)
Die Sätze 13.13 und 13.14 gelten sinngemäß auch für Funktionen mit Werten in $[0, \infty]$ (nach entsprechender Anpassung von Abschnitt 13.1). In dieser Erweiterung gilt Fatou ohne die Einschränkung $\liminf < \infty$.

Satz 13.16 (Satz von der majorisieren Konvergenz, Satz von Lebesgue)
Es gelte: $\lim_{n \to \infty} f_n(\omega) = f(\omega)$ für alle $\omega \in \Omega$. Existiert dann eine μ-integrierbare Funktion $g : \Omega \to \mathbb{R}$ mit $|f_n(\omega)| \leq g(\omega) \quad \forall \omega \in \Omega, n \in \mathbb{N}$, so folgt

$$\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu.$$

In Worten: Existiert eine integrierbare dominierende Funktion, so kann man Limes und Integral vertauschen.

Beweis

Setze $g_n := |f_n - f|$, $h := |f| + g$. Wegen $|h| \leq 2g$ ist h integrierbar. Außerdem:

$$h - g_n = |f| + g - |f_n - f| \geq |f| + g - |f_n| - |f| \geq 0.$$

Wegen $g_n \to 0$ gilt $h - g_n \to h$, also folgt mit Fatou

$$\begin{aligned} \int h \, d\mu &= \int \liminf_{n \to \infty} (h - g_n) \, d\mu \\ &\leq \liminf_{n \to \infty} \int (h - g_n) \, d\mu \\ &= \int h \, d\mu - \limsup_{n \to \infty} \int g_n \, d\mu. \end{aligned}$$

Hier geht Integrierbarkeit von h ein: Aus $\infty - a = \infty$ folgt *nicht* $a = 0$.

Also: $\limsup_{n\to\infty} \int g_n \, d\mu = 0$. Wegen $g_n \geq 0$ bedeutet dies

$$\lim_{n\to\infty} \int |f_n - f| \, d\mu = \lim_{n\to\infty} \int g_n \, d\mu = 0,$$

mit $\int f_n \, d\mu - \int f \, d\mu| = |\int (f_n - f) \, d\mu| \leq \int |f_n - f| \, d\mu$ folgt die Behauptung. *q.e.d.*

Ist μ ein endliches Maß (also z.B. ein Wahrscheinlichkeitsmaß), so kommen konstante Funktionen als Majoranten in Frage.

Bemerkung 13.17
Zu diesen Sätzen existieren „fast-überall"-Versionen. Hat man beispielsweise in Satz 13.16 nur $f_n \to f$ μ-f.ü., so gilt noch stets: $\int f \, d\mu = \lim \int f_n \, d\mu$: gilt $f_n(\omega) \to f(\omega) \quad \forall \omega \in N^c$, $\mu(N) = 0$, so folgt mit Satz 13.12(ii)

$$\begin{aligned}
\int f \, d\mu &= \int f 1_{N^c} \, d\mu \\
&= \int \lim_{n\to\infty} f_n 1_{N^c} \, d\mu \\
&= \lim \int f_n 1_{N^c} \, d\mu \\
&= \lim_{n\to\infty} \inf f_n \, d\mu
\end{aligned}$$

13.4 Maße mit Dichten

Abkürzende Schreibweise für Integrale über meßbaren Teilmengen von Ω:

$$\int_A f \, d\mu = \int f 1_A \, d\mu$$

Lemma 13.18
Es seien $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum und $f : \Omega \to \mathbb{R}$ eine nicht-negative, $(\mathfrak{A}, \mathfrak{B})$-meßbare Funktion. Dann wird durch

$$\nu : \mathfrak{A} \to [0, \infty], \quad \nu(A) := \int_A f \, d\mu$$

ein Maß auf $(\Omega, \mathfrak{A})$ definiert.

Beweis

$\nu(\emptyset) = \int f 1_\emptyset \, d\mu = \int 0 \, d\mu = 0$.

Ist $(A_n)_{n\in\mathbb{N}}$ eine disjunkte Mengenfolge, $A := \bigcup_{n=1}^\infty A_n$, so gilt wegen $f 1_{\bigcup_{i=1}^n A_n} \uparrow f 1_A$:

$$
\begin{aligned}
\nu(\sum_{i=1}^{\infty} a_n) &= \int f 1_A \, d\mu \\
&= \lim_{n\to\infty} \int f \overbrace{1_{\bigcup_{i=1}^n A_i}}^{\sum_{i=1}^n 1_{A_i}} \, d\mu \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \underbrace{\int f 1_{A_i} \, d\mu}_{\nu} \\
&= \sum_{i=1}^{\infty} \nu(A_i).
\end{aligned}
$$

q.e.d.

Definition 13.19
μ, ν seien Maße auf $(\Omega, \mathfrak{A})$. Existiert eine nicht-negative, $(\mathfrak{A}, \mathfrak{B})$-meßbare Funktion $f : \Omega \to \mathbb{R}$ mit

$$\nu(A) = \int_A f \, d\mu \text{ für alle } A \in \mathfrak{A},$$

so heißt f eine μ-Dichte von ν, Schreibweise:

$$f = \frac{d\nu}{d\mu}.$$

Aus Satz 13.12(ii) folgt, daß mit f auch jedes g mit $f = g$ μ-f.ü. μ-Dichte von ν ist: Dichten sind i.a. nicht eindeutig.

Hat ν Dichte f bzgl. μ, so folgt mit Satz 13.12(i)

$$\mu(N) = 0 \implies f 1_N = 0 \ \mu\text{-f.ü.} \implies \nu(N) = 0,$$

d.h. jede μ-Nullmenge ist auch ν-Nullmenge (μ dominiert ν, $\mu \gg \nu$). Der Satz von Radau-Nikodym besagt, daß diese Eigenschaft bei σ-endlichen Maßen bereits die Existenz einer Dichte impliziert.

Beispiel 13.20
$(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}, \mathfrak{B}, \ell)$. Ist $f : \mathbb{R} \to \mathbb{R}$ eine Borel-meßbare nicht-negative Funktion mit $\int f \, d\ell = 1$, so wird durch

$$P : \mathfrak{B} \to [0,1], P(A) = \int_A f(x) \, dx$$

ein Wahrscheinlichkeitsmaß definiert, das Wahrscheinlichkeitsmaß mit (Lebesgue-)Dichte f. Ist X eine Zufallsvariable mit Verteilung P, so heißt f eine Wahrscheinlichkeitsdichte von X. Für die zugehörige Verteilungsfunktion F gilt dann

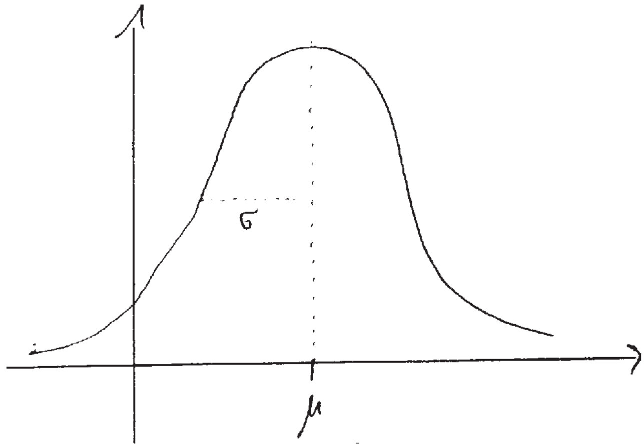
$$F(x) = P((-\infty, x]) = \int_{-\infty}^{x} f(y) \, dy.$$

Die Wahrscheinlichkeitsdiche ist (grob) die Ableitung der Verteilungsfunktion. Für $U(a,b)$ erhält man $f = \frac{1}{(b-a) 1_{(a,b)}}$.

Man kann auch $1_{[a,b]}$ etc. nehmen. Hat F eine stetige Ableitung f, so ist f die eindeutig bestimmte stetige Version der Dichte.

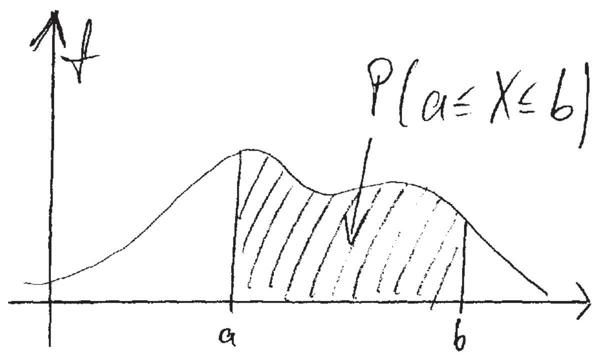
Ein sehr wichtiges Beispiel ist die Normalverteilung mit Parametern μ ($\in \mathbb{R}$) und σ^2 (> 0), kurz $N(\mu\sigma^2)$, mit stetiger Dichte.

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$$



Gaußsche Glockenkurve

Zusammenfassung: Wahrscheinlichkeitsdichten sind nicht-negative Funktionen mit Integral 1; hat X die Wahrscheinlichkeitsdichte f, so ist $\int_a^b f(x)\,dx$ die Wahrscheinlichkeit dafür, daß X in (a,b) (oder $[a,b]$) liegt.



Satz 13.21
μ, ν Maße auf $(\Omega, \mathfrak{A})$, ν habe μ-Dichte f. Dann gilt für alle $(\mathfrak{A}, \mathfrak{B})$-meßbaren $g : \Omega \to \mathbb{R}$: g ist genau dann ν-integrierbar, wenn gf μ-integrierbar ist, und man hat dann

$$\int g\,d\nu = \int gf\,d\mu.$$

Beweis
Mit der „üblichen Maschinerie", daher Übungsaufgabe. *q.e.d.*

Schreibt man für f $\frac{d\nu}{d\mu}$, so wird dies zu

$$\int g\,d\nu = \int g\,\frac{d\nu}{d\mu}\,d\mu.$$

13.5 Erwartungswerte

Einige wichtige Verteilungsparameter: $(\Omega, \mathfrak{A}, P)$ Wahrscheinlichkeitsraum, X, Y Zufallsvariablen.

Definition 13.22

(i) Ist X P-integrierbar, dann heißt $\mathrm{E}X := \int X\, dP$ der *Erwartungswert* (auch: *Mittelwert*) von X. Gilt: $\mathrm{E}|X|^k < \infty$, so heißt $\mathrm{E}X^k$ das k-te *Moment* zu X. (Im Falle $\mathrm{E}|X|^k = \infty$: Das k-te Moment existiert nicht).

(ii) Existiert zu X das zweite Moment, so heißt $\operatorname{var} X := \mathrm{E}(X - \mathrm{E}X)^2$ die Varianz von X, $\sigma(X) := \sqrt{\operatorname{var} X}$ die *Standardabweichung*.

Erwartungswert und Varianz bzw. Standardabweichung können als Lage- bzw. Streuungsparameter der Verteilung von X betrachtet werden.

Kurz: „Erwartungswerte sind Integrale“

Satz 13.23
Wir setzen voraus, daß die beteiligten Erwartungswerte etc. existieren.

(i) $\mathrm{E}(X+Y) = \mathrm{E}X + \mathrm{E}Y, \quad \mathrm{E}(aX) = a\mathrm{E}X \quad (a \in \mathbb{R}).$

(ii) $X \leq Y \implies \mathrm{E}X \leq \mathrm{E}Y.$

(iii) Ist X diskret verteilt, d.h. es gibt eine abzählbare Menge $A \subset \mathbb{R}$ mit $P(X \in A) = 1$, so gilt für alle Borel-meßbaren $g : \mathbb{R} \to \mathbb{R}$:

$$\mathrm{E}g(X) = \sum_{x \in A} g(x) P(X = x).$$

(iv) Ist X absolut-stetig verteilt, d.h. X hat eine Wahrscheinlichkeitsdichte f, so gilt für alle Borel-meßbaren $g : \mathbb{R} \to \mathbb{R}$:

$$\mathrm{E}g(X) = \int g(x) f(x)\, dx.$$

Beweis

(i) Satz 13.8

(ii) ebenfalls Satz 13.8

(iii) $P^X = \sum_{x \in A} P(X = x)\delta_X$, dann Satz 13.10

(iv) Satz 13.10 und Satz 13.21: $\mathrm{E}g(X) = \int g\, dP^X = \int gf\, d\ell$.

q.e.d.

10. November 1994

X diskret verteilt: $\mathrm{E}g(X) = \sum_{x \in \mathbb{R}} g(x) P(X = x)$

X absolut stetig verteilt: $\mathrm{E}g(X) = \int g(x) f(x)\, dx$

Transformationssatz liefert allgemein: $\mathrm{E}g(X) = \int g(x) P^X(dx)$

Insbesondere: $\mathrm{E}X$, $\operatorname{var} X$ hängen nur von der Verteilung von X ab (daher: Verteilungsparameter).

Satz 13.23 beinhaltet eine Verträglichkeitsaussage. ($g(X)$ kann als Zufallsvariable Y aufgefaßt werden.)

Beachte „Kompatibilität“ zu Def. 6.6. Offensichtlich gilt auch weiterhin

- $\operatorname{var}(X) = \mathrm{E}X^2 - (\mathrm{E}X)^2$,
- $\operatorname{var}(a + X) = \operatorname{var}(X)$,
- $\operatorname{var}(aX) = a^2 \operatorname{var}(X)$.

Beispiel 13.24
$X \sim N(0,1)$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$,

$$\mathrm{E}X = \int x f(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}}\, dx = 0 \quad \text{(Antisymmetrie)}$$

$$\begin{aligned}
\mathrm{E}X^2 &= \frac{1}{\sqrt{2\pi}} \int \underset{\downarrow}{x} \cdot \underbrace{(x e^{-\frac{x^2}{2}})}_{\uparrow}\, dx \\
&= \underbrace{-\frac{1}{\sqrt{2\pi}}\, x\, e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty}}_{=0} + \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx}_{\text{Gesamtmasse zu } N(0,1)} \\
&= 1.
\end{aligned}$$

also

$$\operatorname{var}(X) = \mathrm{E}X^2 - (\mathrm{E}X)^2 = 1.$$

Gemäß Aufgabe 20 ist $Y := \mu + \sigma X$ $N(\mu, \sigma^2)$-verteilt, also

$$\mathrm{E}Y = \mu + \sigma \mathrm{E}X = \mu, \quad \operatorname{var}(Y) = \sigma^2 \operatorname{var}(X) = \sigma^2.$$

Die Parameter μ und σ^2 in $N(\mu, \sigma 2)$ sind gerade der zugehörige Erwartungswert und die zugehörige Varianz.

13.6 Ungleichungen

„Eine geeignete Ungleichung ist der halbe Beweis."

Satz 13.25
Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum,

(i) (Markovsche Ungleichung)

$$P(|X| \geq a) \leq \tfrac{1}{a^k} \mathrm{E}|X|^k \quad \forall a > 0$$

(ii) (Ungleichung von Chebychev)
Existiert zu X die Varianz, so gilt

$$P(|X - \mathrm{E}X| \geq a) \leq \tfrac{1}{a^2} \operatorname{var} X \quad \forall a > 0$$

Beweis

(i) Definiere $Y : \Omega \to \mathbb{R}$ durch $Y(\omega) := \begin{cases} 0, & |X(\omega)| < a \\ a, & |X(\omega)| \geq a \end{cases}$. Offensichtlich gilt dann $|Y| \leq |X|$, also $|Y|^k \leq |X|^k$ und damit

$$\begin{aligned} a^k P(|X| \geq a) &= 0^k P(|Y| = 0) + a^k P(Y = a) \\ &= \mathrm{E}|Y|^k \\ &\leq \mathrm{E}|X|^k. \end{aligned}$$

(ii) Wende (i) auf $X - \mathrm{E}X$ mit $k = 2$ an.

q.e.d.

Es sei $I \subset \mathbb{R}$ ein offenes Intervall und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion, d.h.

$$\varphi(\alpha x + (1-\alpha)\, y) \;\leq\; \alpha\, \varphi(x) + (1-\alpha)\, \varphi(y) \quad \forall \alpha \in (0,1),\ \forall x, y \in I$$

Anschaulich: Graph bleibt unterhalb der Sekante.

Bild: Parabel mit Sekante

Lemma 13.26
φ, I wie oben. Dann ex. zu jedem $y \in I$ ein $m \in \mathbb{R}$ mit

$$\varphi(x) \geq \varphi(y) + m(x - y) \quad \forall x \in I$$

Anschaulich: Durch jeden Punkt des Graphen geht eine Gerade, die unterhalb des Graphen bleibt.

Beweis

Übungsaufgabe! *q.e.d.*

Satz 13.27 (Jensensche Ungleichung)
Es sei $I \subset \mathbb{R}$ ein offenes Intervall, $\varphi : I \to \mathbb{R}$ konvex, X eine Zufallsvariable mit $\mathrm{E}|X| < \infty$, $\mathrm{E}|\varphi(X)| < \infty$, $P(X \in I) = 1$. Dann gilt: $\mathrm{E}X \in I$ und

$$\varphi(\mathrm{E}X) \leq \mathrm{E}\varphi(X).$$

Beweis

Ist $I = (-\infty, \infty)$, so gilt automatisch $\mathrm{E}X \in I$.

Ist $X < a$ P-f.s., so gilt $\mathrm{E}X \leq \mathrm{E}a = a$. Wegen $a - X \geq 0$ würde $\mathrm{E}(a - X) = 0$ auf $P(X = a) = 1$ führen (Satz 13.12(i)), also $\mathrm{E}X < a$.

Analog: $X > a$ P-f.s. $\implies$ $\mathrm{E}X > a$, also insgesamt:

$$P(X \in I) = 1 \implies \mathrm{E}X \in I.$$

Nach Lemma 13.26 existiert ein $m \in \mathbb{R}$ mit

$$\varphi(X) \geq \varphi(\mathrm{E}X) + m(X - \mathrm{E}X).$$

Nimm beidseitig den Erwartungswert. *q.e.d.*

Wichtige Spezialfälle:

- $\varphi(x) = |x|$ führt auf $|\mathrm{E}X| \leq \mathrm{E}|X|$,
- $\varphi(x) = x^2$ führt auf $(\mathrm{E}X)^2 \leq \mathrm{E}X^2$, also $\operatorname{var} X \geq 0$.
- Durch Übergang auf $-\varphi$ erhält man die umgekehrte Ungleichung für konkave Funktionen, also beispielsweise

$$\log \mathrm{E}X \geq \mathrm{E} \log X \quad \text{im Falle } X > 0 \ P\text{-f.s.}$$

Sei nun $(\Omega, \mathfrak{A}, \mu)$ ein allgemeiner Maßraum, $\mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) := \{f : \Omega \to \mathbb{R} \mid \int |f|^p \, d\mu < \infty\}$. Dann ist

$$\|f\|_p := (\int |f|^p \, d\mu)^{\frac{1}{p}}.$$

Offensichtlich gilt: $\|f\|_p \geq 0$, $\|af\|_p = |a| \|f\|_p \quad \forall a \in \mathbb{R}$.

Satz 13.28

(i) Höldersche Ungleichung

Hat man $f \in \mathfrak{L}^p$, $g \in \mathfrak{L}^q$ mit $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, so folgt $f \cdot g \in \mathfrak{L}^1$ sowie

$$\|f \cdot g\|_1 \leq \|f\|_p \cdot \|g\|_q$$

(ii)

$$f, g \in \mathfrak{L}^p, \ p \geq 1 \implies f + g \in \mathfrak{L}^p, \ \|f + g\|_p \leq \|f\|_p + \|g\|_p$$

Beweis

(i) Man kann f und g durch $|f|$ und $|g|$ ersetzen, also $f, g \geq 0$ annehmen. Sei $c := \int f^p \, d\mu$. Im Falle $c = 0$ gilt $f = 0$ μ-f.ü., und die Ungleichung läuft auf $0 \leq 0$ hinaus.

Sei also $c > 0$. Es sei P das Wahrscheinlichkeitsmaß mit μ-Dichte $\frac{1}{c} f^p$. Sei weiter $h : \Omega \to \mathbb{R}$, $h(\omega) := \begin{cases} \frac{g(\omega)}{f(\omega)^{p-1}}, & f(\omega) > 0, \\ 0, & f(\omega) = 0 \end{cases}$. Dann:

$$\begin{aligned}
\|f\|_p \cdot \|g\|_q &\geq \|f\|_p \cdot \left\|g 1_{\{f>0\}}\right\|_q \\
&= c^{\frac{1}{p}} \left(\int h^q \underbrace{f^{q(p-1)}}_{f^p} \, d\mu \right)^{\frac{1}{q}} \\
&= c^{\frac{1}{p}} \cdot \left(c \int h^q \, dP \right)^{\frac{1}{q}} \\
&\overset{\text{Jensen}}{\geq} c \int h \underbrace{dP}_{= \frac{1}{c} f^p \, d\mu} \\
&= \int g f \, d\mu \\
&= \|gf\|_1
\end{aligned}$$

(ii) Wegen $|f+g| \leq |f|+|g|$ gilt $\|f+g\|_p \leq \||f|+|g|\|_p$, es reicht also wieder, die Ungleichung für $f, g \geq 0$ zu beweisen.

„$p=1$“:

$$\|f+g\|_1 = \int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu = \|f\|_1 + \|g\|_1 .$$

„$p>1$“: Es gilt

$$(f+g)^p \leq 2 \cdot \max\{f,g\})^p \leq 2^p \leq 2^p(f^p+g^p).$$

Hiermit folgt $f+g \in \mathfrak{L}^p$, also $\|f+g\|_p < \infty$. Mit $q := (1-\frac{1}{p})^{-1}$ folgt die Höldersche Ungleichung:

$$\begin{aligned}
\|f+g\|_p^p &= \int (f+g)^p\, d\mu \\
&= \int f(f+g)^{p-1}\, d\mu + \int g(f+g)^{p-1}\, d\mu \\
&\overset{\text{Hölder}}{\leq} \|f\|_p + \|g\|_p \cdot \left\|(f+g)^{p-1}\right\|_q
\end{aligned}$$

Es gilt

$$\begin{aligned}
\left\|(f+g)^{p-1}\right\|_q &= \left(\int (f+g)^{(p-1)q}\, d\mu\right)^{\frac{1}{q}} \\
&= \left(\left(\int (f+g)^p\, d\mu\right)^{\frac{1}{p}}\right)^{\frac{p}{q}} \\
&= \|f+g\|_p^{p-1}
\end{aligned}$$

Im Falle $\|f+g\|_p = 0$ ist nichts zu zeigen, Im Falle $\|f+g\|_p > 0$ folgt die behauptete Ungleichung nach Division durch $\|f+g\|_p^{p-1} \quad (<\infty)$.

q.e.d.

Als Spezialfall der Hölderschen Ungleichung ergibt sich bei $p=q=2$ die Cauchy-Schwarz-Ungleichung:

$$\left|\int fg\, d\mu\right| \leq \left(\int f^2\, d\mu\right)^{\frac{1}{2}} \cdot \left(\int g^2\, d\mu\right)^{\frac{1}{2}}$$

Bei $\Omega = \{1,\ldots,n\}$, $\mathfrak{A} = \wp(\Omega)$ und $\mu = \sum_{k=1}^n \delta_k$ wird dies zu

$$\left|\sum_{k=1}^n a_k b_k\right| \leq \left(\sum_{k=1}^n a_k^2\right)^{\frac{1}{2}} \cdot \left(\sum_{k=1}^n b_k^2\right)^{\frac{1}{2}}$$

für alle $a_1,\ldots,a_n,b_1,\ldots,b_n \in \mathbb{R}$.

Die Minkowskische Ungleichung zeigt, daß für $\|\cdot\|_p$ die Dreiecksungleichung gilt.

$\|\cdot\|_p$ ist eine Pseudonorm, denn $\|f\|_p = 0$ impliziert nur $f=0$ μ-f.ü.

15. November 1994

14 Produktmaße und Unabhängigkeit

14.1 Das Produkt von zwei Maßräumen

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

Definition 14.1
Das Produkt $\mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2$ der σ-Algebren $\mathfrak{A}_1, \mathfrak{A}_2$ ist die über Ω von den *Recheckmengen* $A_1 \times A_2$, $A_i \in \mathfrak{A}_i$, erzeugte σ-Algebra.

Lemma 14.2
Für alle $A \in \mathfrak{A}, \omega_1 \in \Omega_1, \omega_2 \in \Omega_2$ gilt:

$$\begin{aligned} A_{\omega_1} &:= \{\omega_2 \in \Omega_2 \mid (\omega_1, \omega_2) \in A\} \in \mathfrak{A}_2 \\ A_{\omega_2} &:= \{\omega_1 \in \Omega_1 \mid (\omega_1, \omega_2) \in A\} \in \mathfrak{A}_1. \end{aligned}$$

Man nennt A_{ω_i} den ω_i-Schnitt von A $(i = 1, 2)$.

Beweis

Sei $\omega_1 \in \Omega_1$. Wegen

$$\begin{aligned} \Omega_{\omega_i} &= \Omega_2, \\ (\Omega \setminus A)_{\omega_1} &= \{\omega_2 \mid (\omega_1, \omega_2) \notin A\} \\ &= \{\omega_2 \mid (\omega_1, \omega_2) \in A\}^c \\ &= \omega_2 \setminus A_{\omega_i} \quad \text{und} \\ \left(\bigcup_{n=1}^{\infty} A_n\right)_{\omega_1} &= \bigcup_{n=1}^{\infty} (A_n)_{\omega_i} \end{aligned}$$

folgt: $\mathfrak{A}' := \{A \in \mathfrak{A} \mid A_{\omega_i} \in \mathfrak{A}_2\}$ ist σ-Algebra, die wegen $(A_1 \times A_2)_{\omega_1} = \begin{cases} A_2, & \text{falls } \omega_i \in A \\ \emptyset, & \text{falls } \omega_i \notin A \end{cases}$ das Erzeugendensystem von $\mathfrak{A}$ enthält, d.h. $\mathfrak{A}' = \mathfrak{A}$.

Der zweite Teil folgt analog. *q.e.d.*

Lemma 14.3
Die Maße μ_1 und μ_2 seien σ-endlich. Dann gilt für alle $A \in \mathfrak{A}$:

$$\begin{aligned} \omega_1 \to \mu_2(A_{\omega_1}) &\text{ ist } (\mathfrak{A}_1, \overline{\mathfrak{B}})\text{-meßbar} \\ \omega_2 \to \mu_1(A_{\omega_2}) &\text{ ist } (\mathfrak{A}_2, \overline{\mathfrak{B}})\text{-meßbar.} \end{aligned}$$

Beweis

$\exists (B_n)_{n\in\mathbb{N}} \subset \mathfrak{A}_2$, $\mu_2(B_n) < \infty$, $B_n \uparrow \Omega_2$, da μ_2 σ-endlich. Setze

$$f_A(\omega_i) := \mu_2(A_{\omega_i}), \quad f_{A,n}(\omega_2) := \mu_2(A_\omega \cap B_n).$$

Sei $\mathfrak{D}$ die Menge aller $D \in \mathfrak{A}$ mit $(\mathfrak{A}_1, \mathfrak{B})$-meßbaren $F_{D,n}$. Behauptung: $\mathfrak{D}$ ist ein Dynkin-System. Dazu ist zu zeigen:

(i) $f_{\Omega,n}(\omega_1) = \mu_2(\Omega_\omega \cap B_n) = \mu_2(\Omega_2 \cap B_n) = \mu_2(B_n)$, also, da konstante Abbildungen meßbar sind, $\Omega \in \mathfrak{D}$.

(ii) $f_{D^c,n} = \mu_2(B_n) - f_{D,n}$, also: $D \in \mathfrak{D} \implies D^c \in \mathfrak{D}$ (hier geht $\mu_2(B_n) < \infty$ ein)

(iii) $f_{\sum_{i=1}^\infty D_i,n} = \sum_{i=1}^\infty f_{D_i,n}$, also $(D_i) \in \mathfrak{D}$ disjunkt $\implies \bigcup D_i \in \mathfrak{D}$.

Wegen $f_{A_1\times A_2,n}(\omega_1) = \mu_2(A_2 \cap B_n)1_{A_1}(\omega_1)$ enthält $\mathfrak{D}$ das $\cap$-stabile Erzeugendensystem der Rechteckmengen, d.h. $\mathfrak{D} = \mathfrak{A}$. Also: $f_{A,n}$ ist $(\mathfrak{A}, \mathfrak{B})$-meßbar für alle n. Wegen $f_A = \sup_n f_{A,n}$ liefert Satz 12.8 die $(\mathfrak{A}, \mathfrak{B})$-Meßbarkeit von f_A. *q.e.d.*

Satz & Definition 14.4
Sind μ_1, μ_2 σ-endlich, so existiert genau ein Maß $\mu = \mu_1 \otimes \mu_2$ auf $\mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2$, das Produktmaß von μ_1 und μ_2, mit

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \quad \forall A_i \in \mathfrak{A}.$$

Weiter gilt

$$\mu(A) = \int \mu_2(A_{\omega_1})\mu_1(d\omega_1) = \int \mu_1(A_{\omega_2})\mu_2(d\omega_1) \quad \text{für alle } A \in \mathfrak{A}.$$

Beweis
Es sei wieder $f_A(\omega_i) := \mu_2(A_{\omega_1})$. Ist $(A_n)_{n\in\mathbb{N}} \subset \mathfrak{A}$ eine disjunkte Mengenfolge und $A = \bigcup_{n=1}^\infty A_n$,

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

Außerdem gilt $\int f_\emptyset \, d\mu = 0$, also ist

$$\pi : \mathfrak{A} \to [0, \infty] \quad \pi(A) := \int f_A \, d\mu_1$$

ein Maß auf $\mathfrak{A}$. Für π gilt

$$\pi(A_1 \times A_2) = \int \mu_2(A_2) \cdot 1_{A_1} \, d\mu_1 = \mu_2(A_2)\mu_1(A_1).$$

Analog wird durch $\pi'(A) = \int \mu_1(A_{\omega_2})\mu_2 \, (d\omega_2)$ ein Maß auf $\mathfrak{A}$ definiert, das auf den Rechteckmengen mit π übereinstimmt.

Da μ_1, μ_2 σ-endlich sind, genügt das System der Rechteckmengen den Voraussetzungen von Satz 11.8, also gilt $\pi = \pi' \quad (=: \mu)$. *q.e.d.*

Mit μ_1, μ_2 ist auch $\mu_1 \otimes \mu_2$ σ-endlich.

Wie integriert man bezüglich $\mu_1 \otimes \mu_2$? Ist $f : \Omega \to \mathbb{R}$ eine Abbildung, so schreiben wir f_{ω_1} bzw. f_{ω_2} für die Abbildungen

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

Dies verallgemeinert den Schnittbegriff, setze $f = 1_A$.

Lemma 14.5
Ist f $(\mathfrak{A}, \mathfrak{B})$-meßbar, so ist f_{ω_1} $(\mathfrak{A}_2, \mathfrak{B})$-meßbar für alle $\omega_1 \in \Omega_1$ und f_{ω_2} ist $(\mathfrak{A}_1, \mathfrak{B})$-meßbar für alle $\omega_2 \in \Omega_2$.

Beweis

Es gilt

$$f_{\omega_1}^{-1}(B) = \{\omega_2 \in \Omega_2 \mid f(\omega_1, \omega_2 \in B\} = (\{\omega \in \Omega \mid f(\omega) \in B\})_\omega = (f^{-1}(B))_{\omega_1}.$$

Wende nun Lemma 14.2 an. *q.e.d.*

Satz 14.6 (Fubini I)
Es seien μ_1, μ_2 σ-endlich und $f : \Omega \to \mathbb{R}$ nicht-negativ und $(\mathfrak{A}, \mathfrak{B})$-meßbar. Dann ist $\omega_1 \to \int f_{\omega_1}\, d\mu_2$ $(\mathfrak{A}, \overline{\mathfrak{B}})$-meßbar, $\omega_2 \to \int f_{\omega_2}\, d\mu_1$ $(\mathfrak{A}_2, \overline{\mathfrak{B}})$-meßbar, und es gilt

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

Beweis

(mit Teilen der „üblichen Maschinerie")

Im Falle $f = \sum_{i=1}^n a_i 1_A$ erhält man mit $(1_A)_{\omega_2} = 1_{A_{\omega_2}}$:

$$\int f_{\omega_2}\, d\mu_1 = \sum_{i=1}^n a_i \int 1_{((A_i)_{\omega_2})}\, d\mu_1 = \sum_{i=1}^n a_i \mu_1((A_i)_{\omega_2}),$$

also Meßbarkeit von $\omega_2 \to \int f_{\omega_2}\, d\mu_1$ und damit

$$\begin{aligned} \int \left(\int f_{\omega_2}\, d\mu_1 \right) \mu_2\, (d\omega_1) &= \sum_{i=1}^n a_i \int \mu_1((A_i)_{\omega_2}) \mu_2\, (d\omega_2) \\ &\overset{14.4}{=} \sum_{i=1}^n a_i \mu_1 \otimes \mu_2(A_i) \\ &= \int f\, d(\mu_1 \otimes \mu_2). \end{aligned}$$

Ist $f \geq 0$ $(\mathfrak{A}, \mathfrak{B})$-meßbar, so existiert $(f_n)_{n \in \mathbb{N}} \subset \mathrm{Prim}_+(\Omega, \mathfrak{A})$ mit $f_n \uparrow f$ und es gilt

$$\int f\, d\mu = \lim_{n \to \infty} \int f\, d\mu.$$

Wegen $(f_n)_{\omega_2} \uparrow f_{\omega_2}$, $(f_n)_{\omega_2} \in \mathrm{Prim}_+(\Omega_1, \mathfrak{A}_1)$ folgt

$$g_n(\omega_2) := \int (f_n)_{\omega_2}\, d\mu \uparrow \int f_{\omega_2}\, d\mu_1 \quad \forall \omega_2 \in \Omega_2.$$

Nach dem bereits bewiesenen Teil ist $(g_n)_{n\in\mathbb{N}}$ eine Folge von nichtnegativen $(\mathfrak{A}_2, \overline{\mathfrak{B}})$-meßbaren Abbildungen mit $\int g_n\, d\mu_2 = \int f_n\, d\mu$ (wobei $\mu = \mu_1 \otimes \mu_2$).

Monotone Konvergenz (zusammen mit Bemerkung 13.15) ergibt

$$\begin{aligned}
\int \left(\int f_{\omega_2}\, d\mu_1 \right) \mu_2\,(d\omega_2) &= \lim_{n\to\infty} \int g_n\, d\mu_2 \\
&= \lim_{n\to\infty} \int f_n\, d\mu \\
&= \int f\, d(\mu_1 \otimes \mu_2).
\end{aligned}$$

Wiederholt man diese Schritte mit vertauschten ω_i, erhält man den Rest der Behauptung. *q.e.d.*

Satz 14.7 (Fubini II)
μ_1, μ_2 σ-endlich, $f : \Omega \to \mathbb{R}$ sei $\mu_1 \otimes \mu_2$-integrierbar. Dann sind μ_1-f.a. f_{ω_1} μ_2-integrierbar und μ_2-f.a. f_{ω_2} μ_1-integrierbar, $\omega_1 \to \int f_{\omega_1}\, d\mu_2$ ist μ_1-integrierbar, $\omega_2 \to \int f_{\omega_2}\, d\mu_1$ ist μ_2-integrierbar und es gilt

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

In einprägsamer Form:

$$\int f\, d(\mu_1 \otimes \mu_2) = \iint f(\omega_1, \omega_2) \mu_1 d(\omega_1) \mu_2 d(\omega_2) = \iint f(\omega_1, \omega_2) \mu_2 d(\omega_2) \mu_1 d(\omega_1).$$

Dies zeigt: Man erhält das Integral bezüglich eines Produktmaßes, indem man nacheinander über die Komponenten integriert, wobei die Reihenfolge keine Rolle spielt.

Beweis
Es gilt: $|f|_{\omega_1} = |f_{\omega_1}|$, $(f^{\pm})_{\omega_1} = (f_{\omega_1})^{\pm}$, also nach Fubini I:

$$\int |f|\, d\mu = \iint |f_{\omega_1}|\, d\mu_2\, \mu_1\,(d\mu_1).$$

Nach Voraussetzung ist das Integral endlich, also

$$\mu_1 \left(\left\{ \omega_1 \in \Omega_1 \;\middle|\; \int |f_{\omega_1}|\, d\mu_2 = \infty \right\} \right) = 0,$$

d.h. f_{ω_1} ist μ_1-f.ü. μ_2-integrierbar.

Wendet man Fubini I auf $f^{\pm}_{\omega_1}$ an, so folgt die $(\mathfrak{A}, \mathfrak{B})$-Meßbarkeit von

$$\omega_1 \to \int f\omega_1\, d\mu_2 \quad \left(= \int (f\omega_1\, d\mu_2)^+ - \int (f\omega_1\, d\mu_2)^- \right)$$

sowie

$$\begin{aligned}\int\left(\int f\omega_1\, d\mu_2\right)\mu_1(d\omega_1) &= \int\left(\int f^+_{\omega_1}\, d\mu_2\right)\mu_1(d\omega_2) - \int\left(\int f^-_{\omega_1} d\mu_2\right) d\mu_1 \\ &= \int f^+ d\mu - \int f^{-d}\mu \\ &= \int f d\mu\end{aligned}$$

Aus Symmetriegründen folgt der Rest.

q.e.d.

Hierbei haben wir Funktionen, die nur außerhalb einer Nullmenge N definiert sind, auf N durch den Wert 0 fortgesetzt.

17. November 1994

14.2 Produkte von mehr als zwei Maßräumen

Produkte mit mehr als zwei Faktoren lassen sich auf mehrere Produkte mit zwei Faktoren durch schrittweises Ausführen zurückführen:

Sind $(\Omega_i, \mathfrak{A}_{,}\mu_i)$ σ-endliche Maßräume, $1 \le i \le n$, so ist

$$((\cdots((\mathfrak{A}_1 \otimes \mathfrak{A}_2) \otimes \mathfrak{A}_3) \otimes \cdots) \otimes \mathfrak{A}_{n-1}) \otimes \mathfrak{A}_n$$

eine σ-Algebra auf $(\Omega_1 \times \Omega_2) \times \cdots) \times \Omega_n)$.

Identifiziert man diese Menge mit $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, so erhält man, daß die Produkt-σ-Algebra von den Rechteckmengen $A_1 \times \cdots \times A_n$, $A_i \in \mathfrak{A}_i$ für $1 \le i \le n$ erzeugt und schreibt kurz

$$\mathfrak{A} = \bigotimes_{i=1}^{n} \mathfrak{A}_i.$$

Weiter erhält man (mit Induktion) die Existenz eines Maßes μ auf $\mathfrak{A}$ mit

$$\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i).$$

Die Produktbildung ist assoziativ, Fubini gilt.

Beispiel 14.8
Im Falle $(\Omega_i, \mathfrak{A}_i, \mu_i) = (\mathbb{R}, \mathfrak{B}, \ell)$ erhält man

$$\begin{aligned}\bigotimes_{i=1}^{n} \Omega_i &= \mathbb{R}^n, \\ \mathfrak{B}^{\otimes n} &:= \mathfrak{B}^n,\end{aligned}$$

die σ-Algebra der n-dimensionalen Borelmengen, $\ell^n = \ell^{\otimes n}$ ist das n-dimensionale Lebesgue-Maß.

$\mathfrak{B}^n$ wird erzeugt von $\mathfrak{I}^n := \{(a, b] \mid a, b \in \mathbb{R}^n\}$, wobei

$$(a, b] := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i < x_i < b_i \text{ für } i = 1, \ldots, n\}.$$

Bild 1

Wahrscheinlichkeitsmaße P auf $(\mathbb{R}^n, \mathfrak{B}^n)$ lassen sich durch die zugehörige n-dimensionale Verteilungsfunktion

$$F : \mathbb{R}^n \to [0,1], \quad F(x) := P((-\infty, x])$$

charakterisieren.

Ist $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}^n$ $(\mathfrak{A}, \mathfrak{B}^n)$-meßbar, so heißt X ein n-dimensionaler Zufallsvektor. Hat P^X eine Dichte f bezüglich ℓ^n, so heißt X absolutstetig verteilt, f heißt Wahrscheinlichkeitsdichte von X.

$\mathfrak{B}^n$ wird von den offenen Mengen erzeugt. In der Sprache der elementaren Geometrie: Das Volumen $\ell^n((a,b])$ eines $(a,b]$ ist das Produkt $\prod_{i=1}^n (b_i - a_i)$ der Seitenlängen.

Im Falle $n = 3$ wird das Cavalierische Prinzip zu einem Spezialfall des Satzes von Fubini.

Interessanter (und schwieriger): ∞ viele Faktoren.

Sei $I \neq \emptyset$, $(\Omega_i, \mathfrak{A}_i, P_i)$ für jedes $i \in I$ ein Wahrscheinlichkeitsraum, $\Omega := \times_{i \in I} \Omega_i$, Menge aller Abbildungen ω: $I \to \bigcup_{i \in I} \Omega_i$ mit $\omega_i \in \Omega_i \quad \forall i \in I$ (ω_i kurz für $\omega(i)$).

Für jedes $J \subset I$ wird durch $\pi_J : \Omega \to \bigotimes_{i \in J} \Omega_i$, $\Pi_J(\omega) := \omega|_J$ die Projektion auf die J-Koordinaten definiert. Die Produkt-σ-Algebra $\mathfrak{A} = \bigotimes_{i \in I} \mathfrak{A}_i$ ist die von den Mengen $\Pi_J^{-1}(A_J)$, $A_J \in \bigotimes_{i \in J} \mathfrak{A}_i$, $J \subset I$ endlich(!!!) den Zylindermengen mit endlicher Basis J erzeugte σ-Algebra.

$\mathfrak{A}$ wird auch von den Mengen $\Pi_{\{i\}}^{-1}(A_i)$, $i \in I$, $A_i \in \mathfrak{A}_i$, erzeugt — allerdings ist dieses System nicht $\cap$-stabil. Insbesondere ist für jedes endliche J Π_J $(\mathfrak{A}, \bigotimes_{i \in J} \mathfrak{A}_i)$-meßbar (Produkt-$\sigma$-Algebren werden von den Projektionen erzeugt).

Satz 14.9
Auf $(\Omega, \mathfrak{A})$ existiert genau ein Wahrscheinlichkeitsmaß P $(= \bigotimes_{i \in I} P_i)$ mit

$$P^{\Pi_J} = \bigotimes_{i \in J} P_i$$

für alle endlichen $J \subset I$.

Beweis

hat Parallelen zum Beweis der Existenz von ℓ, siehe [Bauer]. *q.e.d.*

Ein interessanter Spezialfall läßt sich mit Hilfe von ℓ behandeln:

Beispiel 14.10
Modell für den n-fachen Münzwurf einer fairen Münze.

$(\{0,1\}^n, \wp(\{0,1\}^n), Q_n)$, $Q_n(A) = \frac{\# A}{2^n}$

(Laplace-Experiment, wobei 0: Kopf, 1: Zahl). Man sieht leicht, daß dieser Raum das Produkt der Räume $(\Omega_i, \mathfrak{A}_i, P_i)$, $i = 1, \ldots, n$, ist mit

$$(\Omega_i, \mathfrak{A}_i, P_i) = (\{0,1\}, \wp(\{0,1\}), \tfrac{1}{2}(\delta_0 + \delta_1)).$$

Als Modell für eine unendliche Folge von Münzwürfen käme also das unendliche Produkt dieser Räume in Frage, $I = \mathbb{N}$.

Man kann nun $\bigotimes_{i \in \mathbb{N}} P_i$ mit Hilfe von ℓ konstruieren.

Sei $(\Omega', \mathfrak{A}', P') = ([0,1), \mathfrak{B}|_{[0,1)}, \ell|_{[0,1)})$. Zu jedem $\omega' \in \Omega'$ existiert ein eindeutiges $\omega = T(\omega') \in \Omega := \{0,1\}^{\mathbb{N}}$ mit $\omega' = \sum_{k=1}^{\infty} \omega_k 2^{-k}$, $\#\{k \in \mathbb{N} \mid \omega_k = 0\} = \infty$ (Binärdarstellung, siehe Aufgabe 4, 7, 10).

$\mathfrak{A} = \bigotimes_{i=1}^{\infty} \mathfrak{A}_i$ wird von den Mengen $A_{i,\omega_i} := \Pi_{\{i\}}^{-1}(\{\omega_i\}) = \Omega_1 \times \dots \times \Omega_{i-1} \times \{\omega_i\} \times \Omega_{i+1} \times \dots$ erzeugt, (A_{i,ω_i} ist das Ereignis, daß beim i-ten Wurd das Resultat ω_i erscheint). Wegen

$$T^{-1}(A_{i,\omega_i}) = \bigcup_{(\omega_1,\dots,\omega_{i-1}) \in \{0,1\}^{i-1}} B(\omega_1, \dots, \omega_i),$$

wobei $B(\omega_1, \dots, \omega_i) = \left[\sum_{k=1}^{i} \omega_k \cdot 2^{-k}, \sum_{k=1}^{i} \omega_k \cdot 2^{-k} + 2^{-i}\right)$ (Menge aller $x \in [0,1)$, deren Binärdarstellung mit $\omega_1, \dots, \omega_i$ beginnt), ist T $(\mathfrak{A}, \mathfrak{A}')$-meßbar.

Sei P das Bildmaß von P' unter T. Dann gilt

$$\begin{aligned} P(\{\omega_1\} \times \dots \times \{\omega_n\} \times \{0,1\}^{\mathbb{N}}) &= P'(B(\omega_1, \dots, \omega_n)) \\ &= 2^{-n} \\ &= \prod_{i=1}^{n} P_i(\{\omega_i\}) \\ &= (\bigotimes_{i=1}^{n} P_i)(\{\omega_1\}, \dots, \{\omega_n\}) \end{aligned}$$

Da $\#\Omega_i < \infty \quad \forall i$ folgt hieraus bereits

$$P^{\Pi_J} = \bigotimes_{i \in J} P_i \text{ für alle endlichen } J \subset I,$$

d.h. P ist das gewünschte Produktmaß.

$(\bigotimes_{\mathbb{N}} \wp(\{0,1\}) \neq \wp(\{0,1\}^{\mathbb{N}}))$

14.3 Unabhängigkeit

Aus Abschnitt 3 ist bekannt: Unabhängigkeit hat etwas mit dem Produkt von (diskreten) Maßräumen zu tun.

Es sei $(\Omega, \mathfrak{A}, P)$ ein Maßraum.

Definition 14.11
Eine Familie $\{\mathfrak{A}_i \mid i \in I \neq \emptyset\}$ von Unter-σ-Algebren von $\mathfrak{A}$ heißt stochastisch unabhängig, wenn für jede endliche Teilmenge J von I und jede mögliche Wahl von $A_i \in \mathfrak{A}_i$, $i \in J$, gilt

$$P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

Ist weiter für jedes $i \in I$ $(\omega_i', \mathfrak{A}_i')$ ein meßbarer Raum und $X_i : \Omega_i \to \Omega_i'$ eine $(\mathfrak{A}_i, \mathfrak{A}_i')$-meßbare Abbildung, so heißt die Familie $\{X_i \mid i \in J\}$ stochastisch unabhängig, wenn die von den X_i erzeugten σ-Algebren

$$\sigma(X_i) := \sigma(\{X_i^{-1}(A_i) \mid A_i \in \mathfrak{A}_i'\}) = \{X_i^{-1}(A_i) \mid A_i \in \mathfrak{A}_i\}$$

unabhängig sind.

Eine Familie $\{A_i \mid i \in I\} \subset \mathfrak{A}$ von Ereignissen heißt stochastisch unabhängig, wenn die erzeugten σ-Algebren

$$\sigma(\{A_i\}) = \{\emptyset, A_i, A_i^c, \Omega\}$$

unabhängig sind.

Die letzte Bedingung ist äquivalent zu

$$P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i) \quad \forall J \subset I,\ \#J < \infty.$$

Definition 14.11 ist also „abwärtskompatibel" zu Definition 2.5.

Aus der Definition folgt, daß eine Familie von σ-Algebren/Zufallsgrößen/Ereignissen genau dann unabhängig ist, wenn jede endliche Teilfamilie unabhängig ist.

Satz 14.12
Für jedes $i \in I \neq \emptyset$ sei $\mathfrak{A}_i$ eine Unter-σ-Algebra von $\mathfrak{A}$ mit $\cap$-stabilem Erzeugendensystem $\mathfrak{E}_i$. Gilt

$$P(\bigcap_{i\in J} E_i) = \prod P(E_i) \quad \forall J \subset I,\ \#J < \infty,\ E_i \in \mathfrak{E}_i,$$

so ist $\{\mathfrak{A}_i \mid i \in I\}$ unabhängig.

Kurz:

$$\begin{array}{l} P(\bigcap_{i\in J} E_i) = \prod_{i\in J} P(E_i) \quad \forall J \subset I \text{ endlich},\ E_i \in \mathfrak{E}_i \implies \\ P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i) \quad \forall J \subset I \text{ endlich},\ A_i \in \mathfrak{A}_i \end{array}$$

In Worten: Beim Nachweis der Unabhängigkeit kann man sich auf $\cap$-stabile Erzeugensysteme beschränken.

22. November 1994

Beweis

Wir können I endlich voraussetzen, sogar $I = \{1, \ldots, n\}$ annehmen. Sei $\mathfrak{D}_1$ die Menge aller $A_1 \in \mathfrak{A}_1$ mit

$$P(A_1 \cap E_2 \cap \cdots \cap E_n) = P(A_1) \cdot P(E_2) \cdot \cdots \cdot P(E_n) \quad \forall E_2 \in \mathfrak{E}_2, \ldots, E_n \in \mathfrak{E}_n.$$

D_1 ist ein Dynkin-System, das $\mathfrak{E}_1$ enthält, also gilt

$$\mathfrak{D}_1 = \mathfrak{A}_1.$$

Sei $\mathfrak{D}_2$ die Menge aller $A_2 \in \mathfrak{A}_2$ mit

$$P(A_1 \cap A_2 \cap E_3 \cap \cdots \cap E_n) = P(A_1) \cdot P(A_2) \cdot P(E_3) \cdot \cdots \cdot P(E_n) \forall A_1 \in \mathfrak{A}_1, E_3 \in \mathfrak{E}_3, \ldots, E_n \in \mathfrak{E}_n.$$

Wieder: $\mathfrak{D}_2$ ist Dynkin-System, das $\mathfrak{E}_2$ enthält, also $\mathfrak{D}_2 = \mathfrak{A}_2$. Nach insgesamt n solchen Schritten hat man die Behauptung. *q.e.d.*

Beispiel 14.13
(siehe auch Beispiel 14.10)

Es sei $\Omega = \{0,1\}^{\mathbb{N}}$, $\mathfrak{A} = \wp(\{0,1\})^{\otimes\mathbb{N}}$, $P = (\frac{1}{2}(\delta_0 + \delta_1))^{\mathbb{N}}$.

Definiere $X_n : \Omega \to \{0,1\}$ durch $X_n((\omega_i)_{i\in\mathbb{N}}) := \omega_n$. (Projektion auf die n-te Koordinate, $X_n = \Pi_{\{n\}}$ ist das Resultat des n-ten Wurfes.

Nach Konstruktion der Produkt-σ-Algebra ist X_n $(\mathfrak{A}, \mathfrak{B})$-meßbar, also eine Zufallsvariable; es gilt

$$P(X_n = 0) = P(X_n = 1) = 1 \quad \forall n \in \mathbb{N}.$$

Ein $\cap$-stabiles Erzeugendensystem von $\mathfrak{A}_i := \sigma(X_i)$ ist $\mathfrak{E}_i := \{E_i\}$ mit

$$E_i := \{X_i = 1\} = \{0,1\}^{i-1} \times \{1\} \times \{0,1\}^{\mathbb{N}}.$$

Ist $J \subset \mathbb{N}$ endlich, $J = \{i_1, \ldots, i_n\}$, so gilt

$$\begin{aligned} P(X_{i_1} = 1, \ldots, X_{i_n} = 1) = P(\bigcap_{i\in J} E_i) &= P^{\Pi_J}(\{1\} \times \cdots \times \{1\}) = (\tfrac{1}{2})^n \\ &= \prod_{j=1}^{n} P(X_{i_j} = 1). \end{aligned}$$

Nach Satz 14.12 ist $\{X_n \mid n \in \mathbb{N}\}$ unabhängig.

Da die X_n alle dieselbe Verteilung haben, nennt man $X_1, X_2, \ldots$ unabhängig und identisch verteilt (i.i.d., independent, identically distributed). Der Raum $(\Omega, \mathfrak{A}, P)$ kann also als Modell für den unendlich oft „unabhängig" wiederholten Münzwurf dienen.

Es seien nun $X_1, \ldots, X_n$ Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$. Dann ist $X := (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor (Aufgabe 26). Man bezeichnet die Verteilung $P^X = P^{(X_1,\ldots,X_n)}$ dieses Vektors (ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathfrak{B}^n)$) als gemeinsame Verteilung der Zufallsvariable $X_1, \ldots, X_n$, die zugehörige Verteilungsfunktion als gemeinsame Verteilungsfunktion, eine Dichte (ggf.) als gemeinsame Dichte.

In Analogie/Verallgemeinerung von Satz 6.19 gilt:

Satz 14.14
Es seien $X_1, \ldots, X_n$ Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$.

(i) Dann gilt: $X_1, \ldots, X_n$ unabhängig $\iff$ $P^{(X_1,\ldots,X_n)} = \bigotimes_{i=1}^n P^{X_i}$.

(ii) Es sei F_i die Verteilungsfunktion zu X_i, $1 \le i \le n$; F sei die gemeinsame Verteilungsfunktion. Dann gilt:

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

(iii) Es sei f_i eine Dichte von X_i, $f(x) = \prod_{i=1}^n f_i(x_i)$. Dann gilt:

$$X_1, .., X_n \text{ unabhängig} \iff f \text{ ist gemeinsame Dichte.}$$

Beweis

$\mathfrak{I} = \{(-\infty, a] \mid a \in \mathbb{R}\}$ ist ein $\cap$-stabiles Erzeugendensystem zu $\mathfrak{B}$, also ist $\{X_i^{-1}((-\infty, a]) \mid a \in \mathbb{R}\}$ ein $\cap$-stabiles Erzeugensystem von $\sigma(X)$ (denn $X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B)$).

(i) und (ii) folgen also mit Satz 14.12 aus der Definition des Produktmaßes; (iii) folgt mit Fubini aus (i).
q.e.d.

Teil (i) läßt sich auf beliebige Familien von Zufallsgrößen verallgemeinern. Mit $\{X_i \mid i \in I\}$ ist auch $\{Y_i \mid i \in I\}$, wobei $Y_i = g_i(X_i)$ unabhängig (vgl. Satz 6.20), denn $\sigma(Y_i) \subset \sigma(X_i)$, „Funktionen unabhängiger Zufallsgrößen sind unabhängig".

Satz 14.15 (Multiplikationssatz)
Für unabhängige Zufallsvariable X, Y mit $\mathrm{E}|X| < \infty$, $\mathrm{E}|Y| < \infty$ gilt

$$\mathrm{E}XY = \mathrm{E}X\mathrm{E}Y.$$

Beweis

$$\begin{aligned}
\mathrm{E}XY = \int X(\omega)Y(\omega)P(d\omega) &\overset{\text{Transformationssatz}}{=} \int xyP^{(X,Y)}(d(x,y)) \\
&\overset{\text{Satz 14.14(i)}}{=} \int xyP^X \otimes P^Y(d(x,y)) \\
&\overset{\text{Fubini}}{=} \int y \left(\int xP^X(dx) \right) P^Y(dy) \\
&= \mathrm{E}X\mathrm{E}Y
\end{aligned}$$

q.e.d.

Die Kovarianz von X und Y wird definiert durch

$$\mathrm{cov}(X, Y) = \mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y) \quad (= \mathrm{E}XY - \mathrm{E}X\mathrm{E}Y).$$

X und Y heißen unkorelliert, wenn $\mathrm{cov}(X, Y) = 0$, der Korrelationskoeffizient ist

$$\rho(x, y) := \frac{\mathrm{cov}(X, Y)}{\sqrt{\underbrace{\mathrm{var}(X)\,\mathrm{var}(Y)}_{\text{wenn } \neq 0}}}$$

$$\text{unabhängig} \begin{array}{c} \overset{\text{Satz 14.15}}{\Longrightarrow} \\ \not\Longleftarrow \end{array} \text{unkorreliert}$$

Außerdem:

$$\mathrm{var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{var}(X_i) + \sum_{\substack{i,j=1, \\ i \neq j}}^{n} \mathrm{cov}(X_i, X_j),$$

bei paarweise unkorrelierten Zufallsvariable ist die Varianz der Summe also die Summe der Varianzen (Gleichheit von Bienaymé).

Satz 14.16 (Das Borel-Cantelli-Lemma)
Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $(A_n)_{n\in\mathbb{N}}$ eine Folge von Ereignissen;

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

ist das Ereignis, daß unendlich viele der A_n eintreten (Bsp. 1.2(ii)). Dann gilt

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

(ii) $\sum_{n=1}^{\infty} P(A_n) = \infty$ und $\{A_n \mid n \in \mathbb{N}\}$ unabhängig $\implies P(\limsup_{n\to\infty} A_n) = 1$.

Beweis
Teil (i) ist Aufgabe 5. Für den Beweis von (ii) sei $p_n := P(A_n)$. Für alle $n \in \mathbb{N}$ gilt

$$\begin{aligned}
0 &\leq P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \\
&= \lim_{N\to\infty} P\left(\bigcap_{k=n}^{N} A_k^c\right) \\
&= \lim_{N\to\infty} \prod_{k=n}^{N} (1-p_k) \quad (\text{da } A_n^c, \ldots, A_N^c \text{ unabhängig}) \\
&\leq \lim_{N\to\infty} \exp\left(-\sum_{k=n}^{N} p_k\right) \quad (\text{da } 1-p \leq e^{-p} \quad \forall p) \\
&= 0,
\end{aligned}$$

also folgt

$$\begin{aligned}
0 &\leq P((\limsup A_n)^c) \\
&= P\left(\bigcup_{k=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) \\
&\leq \sum_{n=1}^{\infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = 0.
\end{aligned}$$

q.e.d.

15 Gesetze der großen Zahlen

15.1 Konvergenz fast sicher und Konvergenz in Wahrscheinlichkeit

Gesetze der großen Zahlen gibt es in „starker“ und „schwacher“ Form, sie beziehen sich auf unterschiedliche Konvergenzarten.

Definition 15.1
Es seien $X, X_1, X_2, \ldots$ Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$.

(i) X_n konvergiert P-fast-sicher gegen X, Schreibweise $X_n \stackrel{\text{f.s.}}{\to} X$, wenn gilt

$$P(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

(Wegen $\{\lim X_n = X\} = (\liminf(X_n - X))^{-1}(\{0\}) \cap (\limsup(X_n - X))^{-1}(\{0\})$ gilt $\{\lim X_n = X\} \in \mathfrak{A}$, siehe Satz 12.8(iii).)

(ii) X_n konvergiert in Wahrscheinlichkeit gegen X, $X_n \stackrel{\text{P}}{\to} X$, wenn gilt

$$\lim_{n\to\infty} P(\{\omega \mid |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0 \quad \forall \varepsilon > 0$$

Satz 15.2
Konvergent fast sicher impliziert Konvergenz in Wahrscheinlichkeit, die Umkehrung hiervon gilt nicht.

Beweis

„$\Rightarrow$": Sei $N := \{\omega \in \Omega \mid X_n(\omega) \not\to X(\omega)\}$, nach Vor. $P(N) = 0$. Sei weiterhin

$$\varepsilon > 0, \quad A_n := \{\omega \in \Omega \mid \sup_{m \geq n} |X_m(\omega) - X(\omega)| \geq \varepsilon\}.$$

Klar: $A_n > A_{n+1}$. Für reelle Zahlenfolgen $(a_n)_{n\in\mathbb{N}}$ gilt

$$\lim_{n\to\infty} a_n = a \iff \lim_{n\to\infty} \sup_{m\geq n} |a_m - a| = 0.$$

Also hat man $A_n \downarrow N_\varepsilon \subset N$ und es folgt

$$0 \leq P(|X_n - X| \geq \varepsilon) \leq P(A_n) \to P(N_\varepsilon) = 0.$$

„$\Leftarrow$": Gegenbeispiel zu $\stackrel{\text{P}}{\to} \Longrightarrow \stackrel{\text{f.s.}}{\to}$: die „wandernden Buckel": $(\Omega, \mathfrak{A}, P) = ([0,1), \mathfrak{B}|_{[0,1)}, \ell|_{[0,1)})$

Bild 1

Jedes $n \in \mathbb{N}$ läßt sich eindeutig in der Form

$$n = 2^m + k, m \in \mathbb{N}_0, k \in \{0, \ldots, 2^m - 1\}$$

schreiben. Sei

$$X_n := 1_{[k\cdot s^{-m},(k+1)\cdot 2^{-m})}$$

q.e.d.

24. November 1994

15.2 Das schwache Gesetz der großen Zahlen

Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$. Der Titel des Abschnitts ist ein Sammelbegriff für Resultate der Form

$$\frac{1}{a} \cdot \left(\sum_{i=1}^{n} X_i - b_n\right) \xrightarrow{\text{P}} 0, \quad (a_n)_{n\in\mathbb{N}},\ (b_n)_{n\in\mathbb{N}} \subset \mathbb{R}$$

Verallgemeinerung von Satz 6.30:

Satz 15.3
Es sei $\mathrm{E}X_n^2 < \infty\ \forall n$, $\mathrm{cov}(X_i, X_j) = 0$ für $i \neq j$. Dann gilt

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{var}\, X_i = 0 \implies \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathrm{E}X_i) \xrightarrow{\text{P}} 0.$$

Beweis
Sei $Y_n := \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathrm{E}X_i)$. Es gilt $\mathrm{E}Y_n = 0$ sowie $\mathrm{var}(Y_n) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{var}(X_i)$ (Bienaymé), also liefert Chebychev für alle $\varepsilon > 0$:

$$P(|Y_n| > \varepsilon) \leq \frac{1}{\varepsilon^2} \mathrm{var}(Y_n) \to_{n\to\infty} 0.$$

q.e.d.

15.3 Das starke Gesetz der großen Zahlen

Wieder $(X_n)_{n\in\mathbb{N}}$ Zufallsvariable auf $(\Omega, \mathfrak{A}, P)$. Der Titel des Abschnitts ist ein Sammelbegriff für Resultate der Form $\frac{1}{a_n}(\sum_{i=1}^{n} X_i - b_n) \xrightarrow{\text{f.s.}} 0$.

Satz 15.4
Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariable, mit $\mathrm{E}|X_1| < \infty$, so gilt $\overline{X_n} := \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to \mathrm{E}X_1$ P-fast sicher.

Beweis

Zunächst Annahme: $X_k \geq 0 \quad \forall k \in \mathbb{N}$. Sei

$$Y_k := X_k \cdot 1_{\{X_k \leq k\}} \quad (\text{„bei } k \text{ abgeschnitten“})$$
$$S_n^* := \sum_{k=1}^{n} Y_k.$$

Mit dem Satz von der monotonen Konverenz:

$$\mathrm{E}Y_k = \mathrm{E}X_k \cdot 1_{\{X_k \leq k\}} = \mathrm{E}X_1 \cdot 1_{\{X_1 \leq k\}} \to \mathrm{E}X_1.$$

Aus der Analysis ist bekannt:

$$(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}, a_n \to a \implies \frac{1}{n} \sum_{i=1}^{n} a_i \to a$$

(„Césaro-Mittelwerte haben den gleichen Grenzwert").
Also:

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{E} S_n^* = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathrm{E} Y_k = \mathrm{E} X_1 : \tag{1}$$

Die Y-Variablen sind wieder unabhängig, also

$$\begin{aligned} \operatorname{var} S_n^* &= \sum_{i=1}^{n} \operatorname{var} Y_k \\ &\leq \sum_{k=1}^{n} \mathrm{E} Y_k^2 \\ &\leq \sum_{k=1}^{n} \mathrm{E} X_k^2 \cdot 1_{\{X_k \leq n\}} \\ &\leq n \cdot \mathrm{E} X_1^2 \cdot 1_{[0,n]}(X_1) \end{aligned} \tag{2}$$

Wähle nun ein $\alpha > 1$ und setze $m_n := \lfloor \alpha^n \rfloor$. Für $x > 0$ sei definiert:

$$\begin{aligned} \psi(x) &= \sum_{n=1}^{\infty} \frac{1}{m_n} \cdot 1_{[0,m_n]}(x), \\ N(x) &:= \min\{n \mid m_n \geq x\}. \end{aligned}$$

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

$$\begin{aligned} \psi(x) &= \sum_{n=N(x)}^{\infty} \frac{1}{m_n} \\ &\leq 2 \cdot \sum_{n=N(x)}^{\infty} \alpha^{-n} \\ &= 2 \cdot \alpha^{-N(x)} \cdot \frac{1}{1 - \frac{1}{\alpha}} \\ &\leq \frac{k}{x} \quad \text{mit } k := \frac{2\alpha}{\alpha - 1} \end{aligned} \tag{3}$$

Mit Chebychev folgt für alle $\varepsilon > 0$:

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(\tfrac{1}{m_n} \left| S_{m_n}^* - \mathrm{E} S_{m_n}^* \right| > \varepsilon\right) &\leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 M_n^2} \operatorname{var} S_{m_n}^* \\ &\overset{(2)}{\leq} \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 m_n} \mathrm{E} X_1^2 1_{[0,m_n]}(X_1) \\ &= \frac{1}{\varepsilon^2} \mathrm{E} X_1^2 \sum_{n=1}^{\infty} \frac{1}{m_n} 1_{[o,m_n]}(X_1) \text{ (monotone Konvergenz)} \\ &= \frac{1}{\varepsilon^2} \mathrm{E} X_1^2 \psi(X_1) \\ &\overset{(3)}{\leq} k \mathrm{E} X_1 < \infty. \end{aligned}$$

Mit Aufgabe 31(ii) folgt hieraus

$$\frac{1}{m_n}(S^*_{m_n} - \mathrm{E}S^*_{m_n}) \to 0 \ P\text{-f.s.} \tag{4}$$

Wir wollen den „*“ „loswerden“. Mit Aufgabe 29 folgt

$$\begin{aligned}\sum_{n=1}^{\infty} P(X_n \neq Y_n) &= \sum_{n=1}^{\infty} P(X_n > n) = \sum_{n=1}^{\infty} P(X_1 > n)\\ &\leq \int_{[0,\infty)} P(X_1 \geq x)\,dx\\ &= \mathrm{E}X_1 < \infty\end{aligned}$$

Also folgt mit der ersten Hälfte des Borel-Cantelli-Lemmas, daß

$$N_0 := \{\omega \in \Omega \mid X_n(\omega) \neq Y_n(\omega) \text{ für } \infty \text{ viele } n\}$$

eine P-Nullmenge ist. Für alle $\omega \notin N_0$ existiert ein $k(\omega) \in \mathbb{N}$ mit $X_n(\omega) = Y_n(\omega) \forall n \geq k(\omega)$, auf N_0^c gilt also

$$\frac{1}{n}(S_n(\omega) - S^*_n(\omega)) = \frac{1}{n}\left(\sum_{k=1}^{k(\omega)} (X_k(\omega) - Y_k(\omega))\right) \to 0,$$

d.h. $\frac{1}{n}(S_n - S^*_n) \to 0$ P-f.s.
Aus (4) und (1) folgt $\frac{1}{m_n} S^*_{m_n} S^*_{m_n} \to \mathrm{E}X_1$ P-f.s., also

$$\frac{1}{m_n} S_{m_n} \to \mathrm{E}X_1 P - f.s. \tag{5}$$

Als nächstes wollen wir die Einschränkung auf die Teilfolge (m_n) loswerden: Die „Zuwächse“ von S_n sind ≥ 0, also

$$\frac{m_n}{m_{n+1}} \cdot \frac{S_{m_n}}{m_n} \leq \frac{S_k}{k} \leq \frac{m_{n+1}}{m_n} \cdot \frac{S_{m_{n+1}}}{m_{n+1}}$$

für $m_n \leq k \leq m_{n+1}$, also folgt aus (5) und $\frac{m_{n+1}}{m_n} \to \alpha$

$$\begin{aligned}\tfrac{1}{\alpha}\mathrm{E}X_1 &\leq \liminf_{k\to\infty} \tfrac{1}{k} S_k\\ &\leq \limsup_{k\to\infty} \tfrac{1}{k} S_k\\ &\leq \alpha \mathrm{E}X_1 \ P\text{-f.s.}\end{aligned}$$

Setze nun $\alpha = 1 + \frac{1}{j}$ und lasse $j \to \infty$:

$$\frac{1}{n} S_n \to \mathrm{E}X_1 \text{ f.s.} \tag{6}$$

(Die Vereinigung von abzählbar vielen Nullmengen ist wieder eine Nullmenge.)

Als letztes bleibt die Einschränkung auf nicht-negative Zufallsvariable zu beseitigen. Für allgemeine Zufallsvariablen gilt

$$\begin{aligned}\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i{}^+ - \frac{1}{n}\sum_{i=1}^{n} X_i{}^- &\overset{(6)}{\to} \mathrm{E}X_1{}^+ - \mathrm{E}X_1{}^- \\ &= \mathrm{E}X_1 \ P\text{-fast-sicher.}\end{aligned}$$

q.e.d.

Beispiel 15.5
(siehe auch Beispiel 14.13)

Sei $(\Omega, \mathfrak{A}, P) = (\{0,1\}, \wp(\{0,1\}), \frac{1}{2}(\delta_0 + \delta_1))^{\otimes\mathbb{N}})$ das Modell für den ∞ oft wiederholten Münzwurf. X_n (Projektion auf die n-te Koordinate) ist das Resultat des n-ten Wurfes.

Satz 15.4 liefert

$$\frac{1}{n} \cdot \#\{1 \leq i \leq n \mid X_i = 1\} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{\text{f.s.}}{\to} \mathrm{E}X_1 = \frac{1}{2}$$

In maßtheoretischer Sprechweise: Das Produktmaß P ist auf die Menge der $\{0,1\}$-Folgen konzentriert, bei denen die relative Häufigkeiten von 0 und 1 gegen $\frac{1}{2}$ konvergieren.

In wahrscheinlichkeitstheoretischer Sprechweise: Die Wahrscheinlichkeit dafür, daß die relative Häufigkeit des Ereignisses „Kopf" gegen die zugehörige Wahrscheinlichkeit geht, ist 1.

Ersetzt man $\frac{1}{2}(\delta_0 + \delta_1)$ durch $(1-p)\delta_0 + p\delta_1$, so erhält man eine analoge Aussage für Ereignisse mit beliebiger Wahrscheinlichkeit.

Die Gesetze der großen Zahlen bestätigen unseren axiomatischen Aufbau der Stochastik.

Beispiel 15.6
Es sei $p \in \mathbb{N}$, $p > 1$. Zu jedem $x \in [0,1)$ existiert eine eindeutige Entwicklung zur Basis p:

$$x = \sum_{k=1}^{\infty} a_k p^{-k}, (a_k)_{k\in\mathbb{N}} \subset \{0, \ldots, p-1\}, \quad \#\{k \mid a_k \neq p-1\} = \infty$$

(siehe z.B. Aufg. 4 bei p=10)

Sei B die Menge der zur Basis p normalen Zahlen:

$$B := \left\{ x \in [0,1) \mid \lim_{n\to\infty} \frac{1}{n}\#\{1 \leq k \leq n \mid a_k = q\} = \frac{1}{p} \text{ für alle } q \right\}$$

(grob: alle Ziffern sind gleichwahrscheinlich, kommen gleich oft vor)

Wie in Aufgabe 10 bei $p = 10$ erhält man $B \in \mathfrak{B}$. Wir wollen $\ell(B)$ mit Hilfe des starken Gesetzes der großen Zahlen bestimmen:

Es sei $(\Omega, \mathfrak{A}, P) = \{[0,1), \mathfrak{B}|_{[0,1)}, \ell|_{[0,1)})$, $X_n(\omega)$ die n-te Ziffer in der Entwicklung von ω. Dann ist $(X_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge mit $P(X_n = q) = \frac{1}{p} \quad \forall q, n$. (siehe Bsp. 14.13).

Sei nun $q \in \{0, 1, \ldots, p-1\}$ und $Y_n := 1_{\{q\}}(X_n)$.

$(Y_n)_{n\in\mathbb{N}}$ ist wieder eine i.i.d.-Folge, $\mathrm{E}Y_n = P(X_n = q) = \frac{1}{p}$, also liefert Satz 15.4 $P(B_q) = 1$ mit

$$B_q = \left\{ \omega \in \Omega \mid \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k(\omega) = \frac{1}{p} \right\}.$$

Wegen $\sum_{k=1}^n Y_k = \#\{1 \le k \le n \mid a_k = q\}$ gilt

$$B = \bigcap_{q=0}^{p-1} B_q, \qquad \text{also } P(B) = 1,$$

d.h. ℓ-fast-alle Zahlen $x \in [0,1)$ sind normal zur Basis p.

Da p im obigen Argument beliebig war, und der Durchschnitt von abzählbar vielen Mengen der Wahrscheinlichkeit 1 wieder Wahrscheinlichkeit 1 hat, folgt sogar, daß ℓ-fast-alle Zahlen normal zu *allen* Basen sind — dies ist der Satz von Borel über normale Zahlen.

29. November 1994

16 Charakteristische Funktionen

16.1 Grundbegriffe

Ist $(\Omega, \mathfrak{A}, \mu)$ ein Maßraum und $f : \Omega \to \mathbb{C}$, $f = g + i \cdot h$ eine Abbildung mit μ-integrierbaren Realteil g und Imaginärteil h, so setzen wir

$$\int f\, d\mu := \int g\, d\mu + i \cdot \int h\, d\mu.$$

Viele Eigenschaften des Integrals reellwertiger Funktionen gelten auch für diese Erweiterung auf komplexwertige Funktionen. Darüberhinaus hat man $\Re \int f\, d\mu = \int \Re f\, d\mu$ etc.

Lemma 16.1

$$\left| \int f\, d\mu \right| \le \int |f|\, d\mu.$$

Beweis

Wir können annehmen: $\int f\, d\mu \neq 0$ (sonst ist nichts zu zeigen).

Wegen

$$\Re\left(\overline{\int f\, d\mu} \cdot f \right) \le \left| \overline{\int f\, d\mu} \cdot f \right| = \left| \int f\, d\mu \right| \cdot |f|$$

gilt:

$$\begin{aligned}
\left| \int f\, d\mu \right|^2 &= \Re\left(\overline{\int f\, d\mu} \cdot \int f\, d\mu \right) \\
&= \int \Re\left(\overline{\int f\, d\mu} \cdot f \right)\, d\mu \\
&\le \int \left| \int f\, d\mu \right| \cdot |f|\, d\mu \\
&= \left| \int f\, d\mu \right| \cdot \int |f|\, d\mu.
\end{aligned}$$

Dividiere nun durch $\left|\int f\,d\mu\right|$. *q.e.d.*

Definition 16.2
Es sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$. Dann heißt $\varphi_X : \mathbb{R} \to \mathbb{C}$, $\varphi_X(\Theta) = \mathrm{E}e^{i\Theta X}$ die charakteristische Funktion zu X.
Wegen $e^{iy} = \cos y + i \cdot \sin y$ gilt

$$\mathrm{E}e^{i\Theta X} = \mathrm{E}\cos(\Theta X) + i \cdot \mathrm{E}\sin(\Theta X),$$

insbesondere existieren also die Erwartungswerte.

Mit dem Transformationssatz folgt dann

$$\mathrm{E}e^{i\Theta X} = \int_\Omega e^{i\Theta X(\omega)} P(d\omega) = \int_{\mathbb{R}} e^{i\Theta x} p^X(dx),$$

also hängt φ_X von X nur über die Verteilung von X ab: Zufallsvariablen mit derselben Verteilung haben dieselbe charakteristische Funktion. Ist X diskret verteilt, d.h. existiert eine abzählbare Menge $A \subset \mathbb{R}$ mit $P(X \in A) = 1$, so gilt

$$\varphi_X(\Theta) = \sum_{x \in A} e^{i\Theta x} P(X = x).$$

Insbesondere hat man bei einer $\mathbb{N}_0$-wertigen Zufallsvariable mit wahrscheinlichkeitserzeugender Funktion

$$g_x : \varphi_X(\Theta) = g_x(e^{i\Theta}) \quad \text{(siehe Abschnitt 9.2).}$$

Ist X absolutstetig verteilt mit Dichte f, so gilt

$$\varphi_X(\Theta) = \int_{\mathbb{R}} e^{i\Theta x} f(x)\,dx.$$

Man spricht auch von der charakteristischen Funktion (oder (besser): der Fourier-Transformierten) des Wahrscheinlichkeitsmaßes P^X.

Beispiel 16.3
(i) Sei $X \sim B(n,p)$ (siehe Abschnitt 6.2.1). Dann erhält man

$$\begin{aligned}\varphi_X(\Theta) &= \sum_{k=0}^n e^{i\Theta k} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (p \cdot e^{i\Theta})^k (1-p)^{n-k} \\ &= (1 - p + p \cdot e^{i\Theta})^n.\end{aligned}$$

(ii) Im Falle $X \sim U(0,1)$ (siehe Ende Abschnitt 11) erhält man

$$\begin{aligned}\varphi_X(\Theta) &= \int_0^1 e^{i\Theta x}\,dx \\ &= \int_0^1 \cos(\Theta x)\mathfrak{x} + i \cdot \int_0^1 \sim (\Theta x)\,dx \\ &= \frac{1}{T}\sin\Theta - \frac{i}{\Theta}\cos\Theta + \frac{i}{\Theta} \\ &= \frac{1}{i\Theta}(e^i\Theta - 1) \quad \text{für } \Theta \neq 0 \text{ und } \varphi_X(0) = 1.\end{aligned}$$

Charakteristische Funktionen haben die folgenden elementaren Eigenschaften:

Satz 16.4
Ist φ_X die charakteristische Funktion einer Zufallsvariable X, so gilt

(i) $p_X(0) = 1$,

(ii) $|\varphi_X(\Theta)| \leq 1$ für alle $\Theta \in \mathbb{R}$,

(iii) φ_X ist gleichmäßig stetig auf $\mathbb{R}$,

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

(v) $\varphi_{aX+b}(\Theta) = e^{i\Theta b} \cdot \varphi_X(a\Theta)$ für alle $\Theta \in \mathbb{R}$.

Beweis

(i) $\varphi_X(0) = \mathrm{E}e^{i0X} = \mathrm{E}1 = 1$.

(ii) $|\varphi_X(\Theta)| \overset{\text{Lemma 16.1}}{\leq} \mathrm{E}\big|e^{i\Theta X}\big| = \mathrm{E}1 = 1$.

(iii)

$$\begin{aligned} |\varphi_X(\Theta+h) - \varphi_X(\Theta)| &= \Big|\mathrm{E}(e^{i(\Theta+h)X} - e^{i\Theta X})\Big| \\ &\leq \mathrm{E}\big|e^{i\Theta X}(e^{ihX} - 1)\big| \\ &\leq \mathrm{E}\big|e^{ihX} - 1\big| \text{ für alle } \Theta \in \mathbb{R} \end{aligned}$$

Sei nun $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ mit $h_n \to 0$. Dann gilt für alle $\omega \in \Omega$:

$$\cos(h_n X(\omega)) \to 1, \quad sin(h_n X(\omega)) \to 0,$$

Alle diese Funktionen sind betragsmäßig durch die P-integrierbare Funktion $\omega \to 1$ beschränkt, also folgt mit dem Satz von der majorisierten Konvergenz

$$\sup_{\Theta\in\mathbb{R}} |\varphi_X(\Theta+h_n) - \varphi_X(\Theta)| \leq \mathrm{E}|\cos(h_n X) - 1| + \mathrm{E}\sin(h_n X) \to 0 \text{ mit } n \to \infty.$$

(iv)

$$\varphi_{-X}(\Theta) = \mathrm{E}e^{i\Theta(-X)} = \mathrm{E}\cos(-\Theta X) + i\mathrm{E}\sin(-\Theta X) = \mathrm{E}\cos(\Theta X) - i\mathrm{E}\sin(\Theta X) = \overline{\varphi_X(\Theta)}$$

(v)

$$\varphi_{aX+b} = \mathrm{E}e^{i\Theta(aX+b)} = \mathrm{E}e^{i\Theta b} \cdot e^{i(a\Theta)X} = e^{i\Theta b} \cdot \varphi_X(a\Theta).$$

q.e.d.

16.2 Ableitungen und Momente

Wir benötigen zunächst eine Hilfsaussage aus der Analysis.

Lemma 16.5
Für alle $n \in \mathbb{N}_0$ und $y \in \mathbb{R}$ gilt

$$\left| e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!} \right| \leq \min \left\{ \frac{|y|^{n+1}}{(n+1)!}, \frac{2|y|^n}{n!} \right\}$$

Beweis

Wegen $|\overline{z}| = |z|$ reicht es, den Fall $y \geq 0$ zu betrachten. Wir beweisen zunächst

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

mit vollständiger Induktion ((1) kann auch als Taylor-Entwicklung mit einer Integraldarstellung des Restgliedes betrachtet werden.)

Bein ??? $= 0$ erhält man auf der rechten Seite $1 + i \int_0^y e^{is}\ s = e^{iy}$. Partielle Integration ergibt

$$\int_0^y (y-s)^n e^{is}\, ds = \frac{y^{n+1}}{n+1} + \frac{i}{n+1} \int_0^y (y-s)^{n+1} e^{is}\, ds, \tag{2}$$

dies liefert den Schritt $n \to n+1$.

Aus (1) folgt mit

$$\left| \int_0^y (y-s)^n e^{is}\, ds \right| \leq \int_0^y (y-s)^n\, ds = \frac{1}{n+1} y^{n+1}$$

der erste Teil der behaupteten Ungleichung.

Weiter folgt aus (2) mit $n-1$ anstellt von n:

$$\frac{i}{n} \int_0^y (y-s)^n e^{is}\, ds = \int_0^y (y-s)^{n-1} e^{is}\, ds - \frac{y^n}{n},$$

also liefert (1) auch

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

und damit

$$\left| e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!} \right| \leq \frac{1}{(n-1)!} \int_0^y (y-s)^{n-1} \underbrace{|e^{is} - 1|}_{\leq 2}\, ds \leq \frac{2}{n!} y^n$$

q.e.d.

Satz 16.6
Im Falle $\mathrm{E}|X|^k < \infty$ ist φ_X k-mal differenzierbar und es gilt

$$\varphi_X^{(k)}(\Theta) = \mathrm{E}(iX)^k \cdot e^{i\Theta X} \quad \text{für alle } \Theta \in \mathbb{R},$$

insbesondere $\varphi_X^{(k)}(0) = i^k \mathrm{E}X^k$. Außerdem ist $\varphi_X^{(k)}$ gleichmäßig stetig und beschränkt.

Beweis

Wir schreiben kurz φ für φ_X. Für alle $\Theta \in \mathbb{R}$, $h \neq 0$ gilt

$$\frac{\varphi(\Theta + h) - \varphi(\Theta)}{h} = \mathrm{E}e^{i\Theta X} \cdot \frac{e^{ihX} - 1}{h}.$$

Lemma 16.5 ergibt bei $n = 1$

$$\left|e^{ihx} - 1 - ihx\right| \leq \min\{\tfrac{1}{2}h^2x^2, 2|hx|\},$$

also

$$\left|\frac{e^{ihx} - 1}{h}\right| \leq 3|x|,$$
$$\frac{e^{ihx} - 1}{h} \to ix \text{ mit } h \to 0.$$

Ist nun $\mathrm{E}|X| < \infty$, so läßt sich der Satz von der majorisierten Konvergenz anwenden und liefert

$$p'(\Theta) = \lim_{h \to 0} \tfrac{1}{h}(\varphi(\Theta - h)\varphi(\Theta)) = \mathrm{E}(ix) \cdot e^{i\Theta X}.$$

Hat man bereits „$\mathrm{E}|X|^k < \infty \implies \varphi^{(k)}(\Theta) = \mathrm{E}(iX)^k \cdot e^{i\Theta X}$" gezeigt, dann folgt bei $\mathrm{E}|X|^{k+1} < \infty$ mit denselben Argumenten

$$\begin{aligned}
\varphi^{k+1}(\Theta) &= \lim_{h \to 0} \tfrac{1}{h}(\varphi^{(k)}(\Theta + h) - \varphi^{(k)}(\Theta)) \\
&= \lim_{h \to 0} \mathrm{E}(ix)^k \cdot e^{i\Theta X} \cdot (\tfrac{e^{ihX} - 1}{h}) \\
&= \mathrm{E}(iX)^{k+1} \cdot e^{i\Theta X},
\end{aligned}$$

die k-malige Differenzierbarkeit von φ und die Formel für $\varphi^{(k)}$ folgen also mit Induktion. Mit dieser Formel folgt weiter

$$\left|\varphi^{(k)}(\Theta)\right| \leq \mathrm{E}\left|(iX)^k \cdot e^{i\Theta X}\right| \leq \mathrm{E}|X|^k,$$

also ist $\varphi^{(k)}$ beschränkt.

Die gleichmäßige Stetigkeit von $\varphi^{(k)}$ ergibt sich wie im Beweis von Satz 16.4(iii). *q.e.d.*

Man kann also die Momente einer Zufallsvariable über die Ableitugen der charakteristischen Funktion in 0 erhalten (siehe Satz 9.2).

Existieren alle Momente einer Zufallsvariable und fallen diese „schnell genug", so kann man die charakeristische Funktion mit Hilfe der Momente erhalten.

Satz 16.7
Für alle $\Theta \in \mathbb{R}$ mit $\lim_{n\to\infty} \frac{|\Theta|^n}{n!}\mathrm{E}|X|^n = 0$ gilt

$$\varphi_X(\Theta) = \sum_{k=0}^{\infty} \frac{(i\Theta)^k}{k!}\mathrm{E}X^k.$$

Beweis

Lemma 16.5 liefert nach Integration ($y = \Theta x$)

$$\left|\varphi_X(\Theta) - \sum_{k=0}^{n} \frac{(i\Theta)^k}{k!}\mathrm{E}X^k\right| \leq 2\frac{|\Theta|^n}{n!}\mathrm{E}|X|^n.$$

q.e.d.

Zu vielen Standardverteilungen können die charakteristischen Funktionen sehr elegant mit Mitteln der Funktionentheorie berechnet werden. Satz 16.7 bietet gelegentlich eine (umständlichere) Alternative.

Beispiel 16.8
Es sei $X \sim N(0,1)$. Alle Momente zu X existieren („$e^{\frac{-x^2}{2}}$ fällt schneller, als jede Potenz von x wächst“). Es gilt $\mathrm{E}X^{2k+1} = 0$ für alle $k \in \mathbb{N}_0$, mit partieller Integration folgt

$$\mathrm{E}X^{2k} = (2k-1)\cdot \mathrm{E}X^{2k+2},$$

insbesondere

$$\begin{aligned} \frac{\Theta^{2n}}{(2n!} &= \mathrm{E}X^{2n} \\ &= \Theta^{2n}\frac{(2n-1)\cdot(2n-3)\cdot\dots\cdot 5\cdot 3\cdot 1}{2^n n!}?????????????? \\ &= \frac{\Theta^{2n}}{2^n\cdot n!} \end{aligned}$$

Die Bedingung in Satz 16.7 ist für alle $\Theta \in \mathbb{R}$ erfüllt und man erhält

$$\begin{aligned} \varphi_X(\Theta) &= \sum_{k=0}^{\infty} \frac{(i\Theta)^{2k}}{(2k)!}\cdot(2k-1)\cdot(2k-3)\cdot\dots\cdot 5\cdot 3\cdot 1 \\ &= \sum_{k=0}^{\infty} \frac{(-\Theta^2)^k}{2^k\cdot k!} \\ &= e^{-\frac{\Theta^2}{2}} \quad \text{für alle } \Theta \in \mathbb{R} \end{aligned}$$

Im allgemeinen Fall $X \sim N(\mu,\sigma^2)$ verwenden wir Satz 16.4(v). Ist $Y \sim N(0,1)$, so hat $\mu+\sigma Y$ die gleiche Verteilung wie X, also gilt

$$\varphi_X(\Theta) = e^{i\mu\Theta}\cdot\varphi_Y(\sigma\Theta) = \exp(i\mu\Theta - \tfrac{1}{2}\sigma^2\Theta^2) \quad \forall\Theta \in \mathbb{R}.$$

1. Dezember 1994

16.3 Umkehrsätze

$$(\varphi_X(\Theta) = \mathrm{E}e^{i\Theta X} = \int e^{i\Theta x} p^X(dx))$$

Lemma 16.9

$$\lim_{T\to\infty} \int_0^T \frac{\sin x}{x}\, dx = \frac{\pi}{2}$$

Beweis

Analyis/Übungen *q.e.d.*

Man kennt $\int_0^\infty \frac{1}{x}|\sin x|\; dx = \infty$, hat also hier ein Beispiel für eine uneigentlich Riemann-integrierbare Funktion, die nicht Lebesgue-integrierbar ist.

Satz 16.10
Es sei X eine Zufallsvariable mit charakteristischer Funktion φ. Dann gilt für alle $-\infty < a < b < \infty$:

$$\frac{1}{2}P(X=a) + P(a < X < b) + \frac{1}{2}P(X=b) \quad = \quad \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\Theta a} - e^{-i\Theta b}}{i\Theta} \cdot \varphi(\Theta)\; d\Theta$$

Beweis

Seien

$$\begin{aligned}
S(y) &:= \int_0^y \frac{1}{x} \sin x\; dx, \\
I(T) &:= \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\Theta a} - e^{-i\Theta b}}{i\Theta} \cdot \varphi(\Theta)\; d\Theta, \\
\psi &: \mathbb{R} \times [-T, T] \to \mathbb{C}, \\
\psi(\Theta, x) &:= \begin{cases} \frac{1}{i\Theta}(e^{-i\Theta(a-x)} - e^{-i\Theta(b-x)}), & \Theta \neq 0 \\ b-a, & \Theta = 0 \end{cases}.
\end{aligned}$$

Mit Lemma 16.5 folgt, daß ψ stetig ist und daß $|\psi| \leq b-a$ gilt, also ist ψ $P^X \otimes \ell|_{[-T,T]}$-integrierbar. Also läßt sich Fubini II anwenden und liefert

$$\begin{aligned}
I(T) &= \frac{1}{2\pi} \int_{-T}^{T} \frac{1}{i\Theta}(e^{-i\Theta a} - e^{-i\Theta b}) \left(\int e^{-i\Theta x} P^X\; (dx) \right)\; d\Theta \\
&= \frac{1}{2\pi} \int \left(\int_{-T}^{T} \frac{1}{i\Theta}(e^{-i\Theta(a-x)} - e^{-i\Theta(b-x)})\; d\Theta \right) P^X\; (dx)
\end{aligned}$$

Das innere Integral wird zu

$$\psi_{a,b,T}(x) := 2\int_0^T \frac{1}{\Theta} \sin((x-a)\Theta)\; d\Theta - 2\int_0^T \frac{1}{\Theta} \sin((x-b)\Theta)\; d\Theta.$$

Es gilt

$$\int_0^T \frac{1}{\Theta} \sin(c\Theta)\, d\Theta = \mathrm{sgn}(c) \cdot S(T(|c|), \quad \mathrm{sgn}(c) = \begin{cases} 1, & c > 0 \\ 0, & c = 0 \\ -1, & c < 0 \end{cases},$$

also folgt

$$\psi_{a,b,T}(x) = 2 \cdot \mathrm{sgn}(x-a) \cdot S(T|x-a|) - 2 \cdot \mathrm{sgn}(x-b) \cdot S(T|x-b|)$$

und mit Lemma 16.9 folgt

$$\lim_{T\to\infty} \psi_{a,b,T}(x) := \begin{cases} 0, & x < a \\ \pi, & x = a \\ 2\pi, & a < x < b \\ \pi, & x = b \\ 0, & x > b \end{cases} =: \psi_{a,b}$$

Da S als Funktion auf $[0,\infty)$ stetig ist unf mit $x \to \infty$ gegen einen endlichen Wert konvergiert, gilt $\sup_{X\geq 0} |S(X)| < \infty$.

Also existiert zu $\psi(a,b,T)_{T\geq 0}$ eine integrierbare Majorante und der Satz von der majorisierten Konvergenz kann angewendet werden:

$$\begin{aligned} \lim_{T\to\infty} I(T) &= \tfrac{1}{2\pi} \int \psi_{a,b}(x) \cdot P^X\ (dx) \\ &= \tfrac{1}{2} p(x = A) + p(A < x < B) + \tfrac{1}{2} P(X = b). \end{aligned}$$

q.e.d.

Satz 16.11
Sind X, Y Zufallsvariablen mit derselben charakteristischen Funktion, so haben X und Y dieselbe Verteilung (also z.B. $\varphi(\Theta) = e^{-\frac{\Theta^2}{2}} \implies X \sim N(0,1)$)

Beweis

Man sieht gleich, daß für jede abzählbare Menge D das System $I_D := \{(a,b] \mid a,b \notin D\}$ ein $\cap$-stablies Erzeugendensystem von $\mathfrak{B}$ ist.

Setzt man $D := \{a \in \mathbb{R} \mid P(X = a) + P(Y = a) > 0\}$, so folgt aus Satz 16.10, daß P^X und P^Y auf I_D übereinstimmen. Nach Korollar 11.9 bleibt zu zeigen, daß D abzählbar ist. Sei hierzu $A(X) := \{x \mid P(X = x) > 0\}$ die Menge der Atome von X. Wegen

$$\#\{x \in \mathbb{R} \mid P(X = x) \geq \tfrac{1}{n}\} \leq n, \quad A(X) = \textstyle\bigcup_{n=1}^\infty \{w \mid P(X = x) \geq \tfrac{1}{n}\}$$

ist $A(X)$ und damit auch D $(= A(X) \cup A(Y))$ abzählbar. *q.e.d.*

Satz 16.12
Es sei X eine Zufallsvariable mit charakteristischer Funktion φ. Gilt $\int |\varphi|(\Theta)\, d\Theta < \infty$, so hat X eine stetige Dichte f mit

$$f(x) = \frac{1}{2\pi} \int e^{-i\Theta x} \varphi(\Theta)\, d\Theta \quad \forall x \in \mathbb{R}$$

Beweis

Lemma 16.5 liefert

$$\left|\frac{1}{i\Theta}(e^{i\Theta a} - e^{-i\Theta b})\right| \leq |b-a| \tag{1}$$

Ist φ ℓ-integrierbar, so ist $|b-a| \cdot |\varphi|$ eine integrierbare Majorante für den Grenzübergang in Satz 16.10, also

$$\tfrac{1}{2}P(X=a) + P(a < X < b) + \tfrac{1}{2}P(X=b) = \tfrac{1}{2n}\int \tfrac{1}{i\Theta}(e^{-i\Theta a} - e^{-i\Theta b}\varphi(\Theta)\, d\Theta.$$

Mit $(*)$ folgt $P(a < X < b) \leq \frac{1}{2\pi}|b-a|S|\varphi(\Theta)|\, d\Theta$, also

$$P(X=x) = \lim_{n\to\infty} P(x - \tfrac{1}{n} < x < x + \tfrac{1}{n}) = 0 \quad \forall x \in \mathbb{R},$$

d.h X hat keine Atome. Sei F die Verteilungsfunktion zu X:

$$F(b) - F(a) = \frac{1}{2\pi}\int \frac{1}{i\Theta}(e^{-1\Theta a} - e^{-i\Theta b})\varphi(\Theta)\, d\Theta.$$

Wegen $(*)$ läßt sich wieder der Satz von der majorisieren Konvergenz anwenden;

$$\begin{aligned}\lim_{h\to 0}\frac{F(x+h)-F(x)}{h} &= \frac{1}{2\pi}\int e^{-i\Theta x}\cdot(\lim_{h\to 0}\frac{1-e^{-i\Theta h}}{i\Theta h})\cdot\varphi(\Theta)\, d\Theta \\ &= \frac{1}{2\pi}\int e^{-i\Theta x}\varphi(\Theta)\, d\Theta\end{aligned}$$

Wie im Beweis zu Satz 16.4(iii) folgt, daß dieser Limes als Funktion von x stetig ist. *q.e.d.*

16.4 Faltungen

Es seien X und Y unabhängige Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$. Wie hängt die Verteilung P^Z der Summe $Z := X + Y$ von P^X und P^Y ab?

Nach Satz 14.14(i) hat der Zufallsvektor $(X,Y) : \Omega \to \mathbb{R}^2$ die Verteilung $P^X \otimes P^Y$; Z ist eine Funktion von (X,Y):

$$Z = h(X,Y) \quad \text{mit } h : \mathbb{R}'' \to \mathbb{R},\ h(x,y) := x + y.$$

P^Z ist das Bildmaß von $P^X \otimes P^Y$ unter h.

Definition 16.13
Es seien μ_1, μ_2 Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathfrak{B})$. Dann heißt das Bildmaß von $\mu_1 \otimes \mu_2$ unter der Abbildung $(x,y) \mapsto (x+y)$ die *Faltung* $\mu_1 * \mu_2$, von μ_1 und μ_2.

Eine Faltung kann auf mehr als zwei Faktoren ausgedehnt werden und ist kommutativ. Aus dem obigen folgt:

Satz 16.14
Sind X und Y unabhängige Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$, so gilt

$$P^{X+Y} = P^X * P^Y.$$

Nützliche Formeln für die Faltung:

Satz 16.15
(i) Für alle $A \in \mathfrak{B}$ gilt:

$$\mu_1 * \mu_2(A) = \int \mu_1(A-\mu)\mu_2(d\mu) = \int \mu_2(A-x)\mu_1(dx).$$

Hierbei ist $A - x = \{y - x \mid y \in A\}$.

In (ii) und (iii) seien X und Y unabhängige Zufallsvariablen aus $(\Omega, \mathfrak{A}, P)$, $Z := X + Y$.

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

$$P(Z = z) = \sum_{x \in A_X} P(Xx)P(Y = z - x) = \sum_{y \in A_Y} P(X = z - y)P(Y = y).$$

(iii) Sind f_X, f_Y Dichten zu X, Y, so ist auch Z absolutstetig verteilt mit Dichte

$$f_Z(z) = \int f_X(x) * f_Y(z - x)\, dx = \int f_X(z - y) f_Y(y)\, dy$$

Beweis

(i) Sei $B := \{(x, y) \in \mathbb{R}^2 \mid x + y \in A\}$. Dann gilt nach Satz 11.4

$$\mu_1 * \mu_2(A) = \mu_1 \otimes \mu_2(B) = \int \mu_1(B_y)\mu_2\ (dy) = \int \mu_2(B_X)\mu_1\ (dx)$$

mit

$$B_y = \{x \in \mathbb{R} \mid (x, y) \in B\} = \{x \mid x + y \in A\} = A - y,$$

also

$$B_X = \{y \in \mathbb{R} \mid (x, y) \in B\} = A - x.$$

(ii)

$$\begin{aligned} P(Z = z) &= \sum_{x \in A_X} P(Z = z, X = x) \\ &= \sum_{x \in A_X} P(Y = z - x, X = x) \\ &= \sum_{x \in A_X} P(X = x)P(Y = z - x) \quad \text{(Unabhängigkeit)} \end{aligned}$$

(iii) Wir definieren $f : \mathbb{R} \to \mathbb{R}$ durch $f(z) := \int f_Y(z-x) f_X(x)\, dx$. Mit Fubini I erhält man für alle $a \in \mathbb{R}$

$$
\begin{aligned}
\int_{(-\infty,a]} f(z)dz &= \iint 1_{(-\infty,a]}(z) f_Y(z-x) f_X(x)\, dx\, dz \\
&= \int \left(\int 1_{(-\infty,a]} f_Y(z-x)\, dz \right) f_X(x)\, dx \\
&= \int \left(\int 1_{(-\infty,a-x]}(z') f_Y(z')\, dz' \right) f_X(x)\, dx \quad \text{(Substitution: } z' := z-x\text{)} \\
&= \int P^Y(-\infty,a]-x) P^X\,(dx) \\
&\overset{\text{Teil (i)}}{=} P^X * P^Y((-\infty,a]) \\
&= P(Z \le a),
\end{aligned}
$$

wobei wir Teil (i) und Satz 16.14 verwendet haben.

q.e.d.

6. Dezember 1994

Sind X, Y unabhängige Zufallsvariablen, so hat $X+Y$ die Verteilung $P^X * P^Y$.

Formel für die Dichten:

$$f_{X+Y}(z) = \int f_X(z-y) f_Y(y)\, dy$$

Dies verallgemeinert den Faltungsbegriff aus Abschnitt 6.

Beispiel 16.16
Es seien X und Y unabhängig, $X \sim N(0,\sigma^2)$, $Y \sim N(0,\tau^2)$. Als Dichte der Summe ergibt sich mit Satz 16.15(iii)

$$
\begin{aligned}
f_{X+Y}(z) &= \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2\right) \cdot \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2\tau^2}(z-x)^2\right)\, dx \\
&\overset{(*)}{=} \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \cdot \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{1}{2}\left(u - z - \frac{\tau}{\sigma \cdot \sqrt{\sigma^2+\tau^2}}\right)^2 - \frac{z^2}{2(\sigma^2+\tau^2)}\right) du \\
&= \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \exp\left(\frac{1}{2(\sigma^2+\tau^2)} z^2\right),
\end{aligned}
$$

wobei bei $(*)$ $\dfrac{1}{\sigma\tau} \times \sqrt{\sigma^2+\tau^2}$ durch u substituiert wird. Also gilt

$$X + Y \sim N(0, \sigma^2+\tau^2).$$

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

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

in Faltungsschreibweise:

$$N(\mu,\sigma^2) * N(\lambda,\tau^2) = N(\mu+\lambda, \sigma^2+\tau^2).$$

Was hat dies alles nun mit charakteristischen Funktionen zu tun?

Satz 16.17
Sind X, Y unabhängige Zufallsvariablen mit charakteristischen Funktionen φ_X und φ_Y, so gilt für die charakteristische Funktion von $X+Y$:

$$\varphi_{X+Y}(\Theta) = \varphi_X(\Theta)\varphi_Y(\Theta) \quad \text{für alle } \Theta \in \mathbb{R}.$$

Beweis

Sei $\Theta \in \mathbb{R}$. Mit X und Y sind auch $e^{i\Theta X}$ und $e^{i\Theta Y}$ unabhängig, also liefert der Multiplikationssatz

$$\begin{aligned} \varphi_{X+Y}(\Theta) &= \mathrm{E}\exp(i\Theta(X+Y)) \\ &= \mathrm{E}\exp(i\Theta X)\exp(i\Theta Y) \\ &\overset{\text{Multiplikationssatz}}{=} \mathrm{E}e^{i\Theta X}\mathrm{E}e^{i\Theta Y} \\ &= \varphi_X(\Theta)\varphi_Y(\Theta). \end{aligned}$$

q.e.d.

Die „schwierige“ Operation der Faltung von Maßen wird durch die Fourier-Transformation auf die „einfache“ Multiplikation von Funktionen abgebildet.

Beispiel 16.18 (siehe auch Beispiel 16.16)
Sind X, Y unabhängig mit $X \sim N(\mu, \sigma^2)$, $Y \sim (\lambda, \tau^2)$, so gilt nach Beispiel 16.8

$$\begin{aligned} \varphi_X(\Theta) &= \exp(i\mu\Theta - \tfrac{1}{2}\sigma^2\Theta^2), \\ \varphi_Y(\Theta) &= \exp(i\lambda\Theta - \tfrac{1}{2}\tau^2\Theta^2), \end{aligned}$$

also folgt mit Satz 16.17

$$\varphi_{X+Y}(\Theta) = \exp(i(\mu+\lambda)\Theta - \tfrac{1}{2}(\sigma^2+\tau^2)\Theta),$$

also (Korollar 16.11)

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

17 Verteilungskonvergenz

17.1 Definition und Zusammenhang zu anderen Konvergenzbegriffen

Definition 17.1

(i) Es seien $P, P_1, P_2, \ldots$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathfrak{B})$ mit Verteilungsfunktionen $F, F_1, F_2, \ldots$ P_n *konvergiert schwach* gegen P,

$$P_n \xrightarrow{\mathrm{w}} P \qquad \text{(w: weak)},$$

wenn für alle Stetigkeitspunkte x von F gilt:

$$\lim_{n\to\infty} F_n(x) = F(x).$$

(ii) Es seien $X, X_1, X_2, \ldots$ Zufallsvariablen auf (verschiedenen) Wahrscheinlichkeitsräumen $(\Omega, \mathfrak{A}, P)$, $(\Omega_1, \mathfrak{A}_1, P_1), \ldots$.

X_n *konvergiert in Verteilung* gegen X,

$$X_n \xrightarrow{\mathcal{D}} X \qquad (\mathcal{D}\text{: distribution}),$$

wenn $P^{X_n} \xrightarrow{w} P^X$ gilt, d.h. wenn für alle $x \in \mathbb{R}$ mit $P(X = x) = 0$ gilt:

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x).$$

Warum die Einschränkung auf Stetigkeitspunkte? Betrachte dazu das Beispiel

$$X_n \equiv \frac{1}{n}, X \equiv 0;$$
$$F_n = 1_{[\frac{1}{n},\infty)}, F = 1_{[0,\infty)}$$

Bild 1

Klar: $F_n(x) \to F(x) \quad \forall x \neq 0$, $F_n(0) \not\to F(0)$

Beispiel 17.2
Es sei $\lambda > 0$ und $X_1, X_2, \ldots$ eine i.i.d.-Folge mit $P(X_1 \geq x) = e^{-\lambda x} \quad \forall x \geq 0$ (Exponentialverteilung mit Parameter λ, Aufg. 12.30)

Sei $M_n := \max\{X_1, \ldots, X_n\}$, $Y_n := M_n - \frac{1}{\lambda}\log n$. Sei weiter F_n die Verteilungsfunktin zu Y_n. Dann gilt für alle $y \in \mathbb{R}$, $n \in \mathbb{N}$ mit $y + \frac{1}{\lambda}\log n \geq 0$:

$$\begin{aligned}
F_n(y) &= P(M_n \leq y + \tfrac{1}{\lambda}\log n) \\
&= P(X_1 \leq y + \tfrac{1}{\lambda}\log n, \ldots, X_n \leq y + \tfrac{1}{\lambda}\log n) \\
&= (P(X_1 \leq y + \tfrac{1}{\lambda}\log n))^n \quad (\text{da } X_1, \ldots, X_n \text{ i.i.d.}) \\
&= (1 - \exp(-\lambda(y + \tfrac{1}{\lambda}\log n)))^n \\
&= (1 - \tfrac{1}{n}e^{-\lambda y})^n
\end{aligned}$$

also:

$$F_n(y) \to G_\lambda(y) := \exp(-e^{-\lambda y}) \quad \forall y \in \mathbb{R}$$

G_λ ist eine Verteilungsfunktion, die zugehörige Verteilung nennt man *Gumbel*-Verteilung. Für große n hat M_n also ungefähr die Verteilung von $\frac{1}{\lambda}\log n + Z$ mit $Z \sim G_\lambda$.

Im Falle $\lambda = 1, n = 10$:

$$\begin{aligned}
P(M_{10} \leq 3) = (1 - e^{-3})^{10} &\approx 0.6001, \\
G_1(-\log 10 + 3) &\approx 0.6078
\end{aligned}$$

Sind alle Zufallsvariablen auf demselben Wahrscheinlichkeitsraum definiert, so kann $\xrightarrow{\mathcal{D}}$ mit den früher eingeführten Konvergenzbegriffen verglichen werden:

Satz 17.3
Sind $X, X_1, X_2, \dots$ Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$, so gilt

$$\begin{aligned} X_n \stackrel{\text{f.s.}}{\to} X &\implies X_n \stackrel{\mathcal{D}}{\to} X; \\ X_n \stackrel{\text{P}}{\to} X &\implies X_n \stackrel{\mathcal{D}}{\to} X \end{aligned}$$

Beweis
Wegen „$\stackrel{\text{f.s.}}{\to} \implies \stackrel{\text{P}}{\to}$“ reicht es, die zweite Aussage zu beweisen. Es gelte also $X_n \stackrel{\text{P}}{\to} X$. Seien F, F_n $(n \in \mathbb{N})$ die zugehörige Verteilungsfunktionen. Für alle $x \in \mathbb{R}$, $\delta > 0$ gilt:

$$\begin{aligned} P(X_n \le x) &\le P(X \le x + \delta) + P(|X_n - X| > \delta), \\ P(X_n \le x) &\ge P(X \le x - \delta) - P(|X_n - X| > \delta). \end{aligned}$$

Läßt man zunächst n gegen ∞ gehen und dann δ von oben gegen 0, so folgt

$$F(x-) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) = F(x).$$

Ist F in x stetig, so gilt $F(x-) = F(x)$ und damit $F_n(x) \to F(x)$. *q.e.d.*

Schwache Konvergenz ist in der Tat schwächer als Konvergenz in Wahrscheinlichkeit, also ist beispielsweise $X \sim N(0,1)$ und setzt man $X_n = X$ für n gerade und $X_n = -X$ für n ungerade, so haben alle Zufallsvariablen dieselbe Verteilungsfunktion, also gilt $X_n \stackrel{\mathcal{D}}{\to} X$. Andererseits gilt für n ungerade.

$$P(|X_n - X| > \tfrac{1}{2}) = P(|X| > \tfrac{1}{4}) \not\to 0,$$

also $X_n \stackrel{\text{P}}{\not\to} X$. Man hat jedoch den folgenden Darstellungssatz, wir schreiben $X \stackrel{\mathcal{D}}{=} Y$, wenn X und Y dieselbe Verteilung haben.

Satz 17.4 (Skorohod)
Es seien $X, X_1, X_2, \dots$ Zufallsvariablen mit $X_n \stackrel{\mathcal{D}}{\to} X$. Dann existieren ein Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ und hierauf Zufallsvariablen $X', X_1', X_2', \dots$ mit $X' \stackrel{\mathcal{D}}{=} X$, $X_n' \stackrel{\mathcal{D}}{=} X_n \quad \forall n \in \mathbb{N}$ derart, daß gilt:

$$X_n' \stackrel{\text{f.s.}}{\to} X'.$$

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

Sei $(\Omega, \mathfrak{A}, P) := ((0,1), \mathfrak{B}|_{(0,1)}, \ell|_{(0,1)})$. Sei $X' := F^{-1}$, $X_n' := F_n^{-1}$ (siehe Def. 12.14). Dann gilt $X' \stackrel{\mathcal{D}}{=} X$, $X_n' \stackrel{\mathcal{D}}{=} X_n$ (siehe den Beweis zu Satz 12.16). Zu zeigen bleibt: Für P-fast-alle ω:

$$\lim_{n\to\infty} X_n'(\omega) = X'(\omega). \tag{1}$$

In Worten: Wir gehen aus von einer Folge von Verteilungsfunktionen die punktweise in den Stetigkeitspunkten der Grenzfunktion konvergiert und wollen hieraus schließen, daß die zugehörigen Quantilfunktionen („Inversen“) für Lebesgue-fast-alle Punkte punktweise konvergieren.

Bild 2

8. Dezember 1994

Es bleibt zu zeigen:

$F_n(x) \to F(x)$ für alle Stetigkeitsstellen von $F \implies F_n^{-1}(y) \to F^{-1}(y)$ für ℓ-fast-alle y ($F_n^{-1} = X_n'$, $F^{-1} = X'$)

$y \in (0,1) = \Omega$. Sei $\omega \in (0,1)$. Zu gegebenem $\varepsilon > 0$ wähle ein $x \in \mathbb{R}$ mit

$$X'(\omega) - \varepsilon < x < X'(\omega) \text{ und } P(X = x) = 0 \quad (X \text{ hat nur abzählbar viele Atome}).$$

Mit Lemma 12.15 ($y \leq F(x) \iff F^{-1}(y) \leq x$) folgt $F(x) < \omega$, also existiert wegen $F_n(x) \to F(x)$ ein n_0 mit $F_n(x) < \omega$ und damit $X_n'(\omega) > x \quad \forall n \geq n_0$.

Mit $\varepsilon \downarrow 0$ folgt

$$\liminf_{n\to\infty} X_n'(\omega) \geq X'(\omega) \quad \forall \omega \in \Omega.$$

Ist nun $\omega < \omega'$ und $\varepsilon > 0$, so existiert ein x mit

$$X'(\omega') < x < X'(\omega') + \varepsilon, \quad P(X = x) = 0.$$

F ist von rechts stetig, also $F(F^{-1}(y)) \geq y$, also

$$\omega < \omega' \leq F(X'(\omega')) \leq F(x)$$

und damit wegen $F_n(x) \to F(x)$: $\omega \leq F_n(x)$ ab einem n_0. $\varepsilon \downarrow 0$ liefert mit Lemma 12.15

$$\limsup_{n\to\infty} X_n'(\omega) \leq X'(\omega') \text{ für alle } \omega' > \omega. \tag{2}$$

Als monoton nicht-fallende Funktion auf $(0,1)$ kann X' nur abzählbar viele Unstetigkeitsstellen (Sprünge) haben, denn: Jedes übersprungene Intervall enthält eine rationale Zahl, und die Intervalle sind disjunkt. Außerhalb dieser abzählbaren Menge ergibt (2) mit $\omega' \downarrow \omega$

$$\limsup_{n\to\infty} X_n'(\omega) \leq X'(\omega),$$

d.h. (1) gilt außerhalb einer abzählbaren Menge und damit für P-fast-alle $\omega \in \Omega$.

q.e.d.

Satz 17.5
Es sei $C_b(\mathbb{R})$ die Menge aller stetigen und beschränkten Funktionen $h : \mathbb{R} \to \mathbb{R}$. Dann gilt

$$X_n \xrightarrow{\mathcal{D}} X \iff \mathrm{E}h(X_n) \to \mathrm{E}h(X) \quad \forall h \in C_b(\mathbb{R}).$$

Beweis

„$\Rightarrow$“:

Nach Satz 17.4 existieren auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ Zufallsvariablen $X' \stackrel{\mathcal{D}}{\to} X$, $X_n' \stackrel{\mathcal{D}}{\to} X_n$ ($n \in \mathbb{N}$) mit $X_n' \to X'$ P-f.s. Da h stetig ist, gilt dann auch

$$h(X'_n) \to h(X') \quad P\text{-f.s.}$$

Da h beschränkt ist, kann der Satz von der majorisierten Konvergenz angewendet werden und liefert

$$\mathrm{E}h(X_n) = \mathrm{E}h(X'_n) \to \mathrm{E}h(X') = \mathrm{E}h(X).$$

„$\Leftarrow$“:

Für alle $a, b \in \mathbb{R}$ mit $a < b$ definieren wir eine stetige und beschränkte Funktion $h_{a,b} : \mathbb{R} \to \mathbb{R}$ durch

$$h_{a,b}(x) := \begin{cases} 1, & x \leq a, \\ \frac{b-x}{b-a}, & a < x < b, \\ 0, & x \geq b \end{cases}$$

Es seien F, F_n $(n \in \mathbb{N})$ die Verteilungsfunktionen zu X, X_n. Dann gilt für alle $y > x$

$$F_n(x) = \mathrm{E}1_{(-\infty,x]}(X_n) \leq \mathrm{E}h_{x,y}(X_n) \to \mathrm{E}h_{x,y}(X) \leq \mathrm{E}1_{(-\infty,y]}(X) = F(y),$$

also folgt mit $y \downarrow x$

$$\limsup_{n\to\infty} F_n(x) \leq F(x) \quad \forall x \in \mathbb{R} \quad F \text{ stetig von rechts}$$

Analog erhält man für alle $y < x$

$$F_n(x) \geq \mathrm{E}h_{yx}(X_n) \to \mathrm{E}h_{yx}(X) = F(y),$$

also mit $y \uparrow x$

$$\liminf_{n\to\infty} F_n(x) \geq F(x-) \quad \forall x \in \mathbb{R}.$$

Ist F in x stetig, so folgt insgesamt $F_n(x) \to F(x)$.

q.e.d.

Satz 17.6 (Continous Mapping Theorem)
Es seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit $X_n \xrightarrow{\mathcal{D}} X$. Weiter sei $f : \mathbb{R} \to \mathbb{R}$ eine Borel-meßbare Abbildung mit $P(X \in \{x \in \mathbb{R} \mid f \text{ nicht stetig in } x\}) = 0$.

Dann gilt:

$$f(X_n) \xrightarrow{\mathcal{D}} f(X).$$

Beweis

Übungsaufgabe. *q.e.d.*

17.2 Straffheit und charakteristische Funktionen

Wir benötigen einen berühmten Satz aus der Analysis.

Satz 17.7 (Helly)
Zu jeder Folge $(F_n)_{n\in\mathbb{N}}$ von Verteilungsfunktinen existiert eine Teilfolge $(F_{n_k})_{k\in\mathbb{N}}$, sowie eine schwach monoton steigende, rechtsstetige Funktion $G : \mathbb{R} \to [0,1]$ mit $F_{n_k}(x) \to_{k\to\infty} G(x)$ für alle $x \in \mathbb{R}$, in denen G stetig ist.

Beweis

(grobe Beweisskizze)

Für jedes $x \in \mathbb{R}$ hat $(F_n(x))_{n\in\mathbb{N}}$ als beschränkte Folge reeller Zahlen einen Häufungspunkt. Sei $(r_k)_{k\in\mathbb{N}}$ eine Abzählung von $\mathbb{Q}$. Wähle nacheinander Teilfolgen $(F_{n_{k,j}})_{j\in\mathbb{N}}$ mit $F_{n_{k,j}}(r_k) \to_{j\to\infty} G_0(r_k)$.

($(n_{k+1,j})_{j\in\mathbb{N}}$ ist Teilfolge von $(n_{k,j})_{j\in\mathbb{N}}$.)

Betrachte nun die Diagonalfolge $(F_{n_{j,j}})_{j\in\mathbb{N}}$, $F_{n_{j,j}} \to_{j\to\infty} G_0$ auf $\mathbb{Q}$. Setze G_0 auf ganz $\mathbb{R}$ fort durch

$$G(x) := \int \{G_0(r) \mid r \in \mathbb{Q},\ r > x\}.$$

Rest mit „ε-δ“. *q.e.d.*

Beachte: Das G aus Satz 17.7 ist nicht unbedingt eine Verteilungsfunktion:

Im Falle $F_n = 1_{[n,\infty)}$ hat man $G \equiv 0$.

Phänomen: Masse kann im Unendlichen verschwinden.

Definition 17.8
Eine Familie $\mathfrak{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathfrak{B})$ heißt *straff*, wenn zu jedem $\varepsilon > 0$ ein kompaktes Intervall $[a,b]$ existiert mit

$$P([a,b]) \geq 1 - \varepsilon \quad \forall P \in \mathfrak{P}.$$

Man sieht leicht:

$$\begin{gathered}\mathfrak{P}_1 \text{ straff}, \mathfrak{P}_2 \text{ straff} \implies \mathfrak{P}_1 \cup \mathfrak{P}_2 \text{ straff}\\ \#\mathfrak{P} = 1 \implies \mathfrak{P} \text{ straff}\\ \mathfrak{P} \text{ straff}, \mathfrak{P}_0 \subseteq \mathfrak{P} \implies \mathfrak{P}_0 \text{ straff}\end{gathered}$$

Satz 17.9
Ist $\{P_n \mid n \in \mathbb{N}\}$ eine straffe Familie von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathfrak{B})$, so existiert eine Teilfolge $(P_{n_k})_{k\in\mathbb{N}}$ von $(P_n)_{n\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß P mit $P_{n_k} \xrightarrow{w} P$ mit $k \to \infty$.

Beweis

Sei F_n die Verteilungsfunktion zu P_n. Der Satz von Helly liefert ein Folge $(n_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ und eine schwach monton wachsende, rechtsstetige Funktion G mit $F_{n_k}(x) \to G(x)$ in allen Stetigkeitspunkten x von G. Zu zeigen bleibt:

$$\lim_{x\to-\infty} G(x) = 0, \quad \lim_{x\to\infty} G(x) = 1.$$

Sei $\varepsilon > 0$. Da $\{P_n \mid n \in \mathbb{N}\}$ straff, existieren $a, b \in \mathbb{R}$ mit $P_n([a+1, b]) \geq 1 - \varepsilon \quad \forall n$, insbesondere $F_n(a) \leq \varepsilon \quad \forall n \in \mathbb{N}$. Da G höchstens abzählbar viele Unstetigkeitsstellen hat (siehe Beweis zu Satz 17.4), existiert ein $c < a$, in dem G stetig ist. Es folgt: $G(c) = \lim_{n\to\infty} F_{n_k}(c) \leq \varepsilon$, also

$$G(x) \leq \varepsilon \quad \forall x \leq c.$$

Damit ist gezeigt:

$$\forall \varepsilon > 0 \, \exists x \in \mathbb{R} \, \forall y \leq x: \; 0 \leq G(y) \leq \varepsilon.$$

Dies ist $\lim_{x\to\infty} G(x) = 0$ in ε-δ-Schreibweise.

Die andere Aussage wird ganz analog bewiesen *q.e.d.*

Der folgende Satz gibt ein notwendiges und hinreichendes Kriterium für Verteilungskonvergenz über charakteristische Funktionen.

Satz 17.10 (Stetigkeitssatz (für charakteristische Funktionen))
Es seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit charakteristischen Funktionen $\varphi, \varphi_1, \varphi_2, \ldots$ Dann gilt:

$$X_n \overset{\mathcal{D}}{\to} X \quad \Longleftrightarrow \quad \varphi_n(\Theta) \to \varphi(\Theta) \quad \forall \Theta \in \mathbb{R}.$$

Beweis

„$\Rightarrow$“:

Sei $\Theta \in \mathbb{R}$. Die Funktionen $x \to \cos(\Theta x)$, $x \to \sin(\Theta x)$ sind stetig und beschränkt, also folgt mit Satz 17.5:

$$\varphi_n(\Theta) = \mathrm{E}\cos(\Theta X_n) + i\mathrm{E}\sin(\Theta X_n) \to \mathrm{E}\cos(\Theta X) + i\mathrm{E}\sin(\Theta X) = \varphi(\Theta).$$

13. Dezember 1994

„$\Leftarrow$“: Zunächst: $\{P^{X_n} \mid n \in \mathbb{N}\}$ ist straff. Fubini II liefert für alle $\delta > 0$:

$$\begin{aligned} \frac{1}{\delta}\int_{-\delta}^{\delta} (1 - \varphi_n(\Theta))\; d\Theta &= \int \frac{1}{\delta}\left(\int_{-\delta}^{\delta}\left(1 - e^{i\Theta x}\right)\; d\Theta\right) P^{X_n}\,(dx) \\ &= 2\int \left(1 - \frac{\sin(\delta x)}{\delta x}\right) P^{X_n}\,(dx) \\ &\geq 2\int \left(1 - \frac{1}{|\delta x|}\right) P^{X_n}\,(dx), \quad \text{wobei } |x| \geq \frac{2}{\delta} \\ &\geq P^{X_n}\left(\left[-\frac{2}{\delta}, \frac{2}{\delta}\right]^c\right) \end{aligned}$$

Sei nun $\varepsilon > 0$. Da φ in 0 stetig ist (Satz 16.4(iii)), und $j\varphi(0) = 1$, existiert ein $\delta > 0$ mit

$$|1 - \varphi(\Theta)| \leq \frac{\varepsilon}{4} \quad \text{für } |\Theta| \leq \delta,$$

d.h.

$$\begin{aligned}\left|\frac{1}{\delta}\int_{-\delta}^{\delta}(1-\varphi(\Theta))\,d\Theta\right| &\le \frac{1}{\delta}\cdot 2\delta\cdot\frac{\varepsilon}{4}\\ &= \frac{\varepsilon}{2}\end{aligned}$$

Da alle φ_n betragsmäßig durch 1 beschränkt sind und über ein endliches Intervall integriert wird, kann der Satz von der majorisierten Konvergenz angewendet werden:

$$\int_{-\delta}^{\delta}(1-\varphi_n(\Theta))\,d\Theta \to_{n\to\infty} \int_{-\delta}^{\delta}(1-\varphi(\Theta))\,d\Theta,$$

also existiert ein $n_0 \in \mathbb{N}$ derart, daß gilt:

$$\frac{1}{\delta}\int_{-\delta}^{\delta}(1-\varphi_n(\Theta))\,d\Theta \le \varepsilon$$

für alle $n \ge n_0$,

also nach obiger Abschätzung

$$P^{X_n}\left(\left[-\frac{2}{\delta},\frac{2}{\delta}\right]^c\right) \ge 1-\varepsilon \quad \forall n \ge n_0.$$

Zu jedem $n \in \{1.,,,.n_0-1\}$ existiert ein a_n mit

$$P^{X_n}([-a_n,a_n]) \ge 1-\varepsilon,$$

also folgt mit $a := \max\{a_1,\ldots,a_{n_0-1},\frac{2}{\delta}\}$:

$$P^{X_n}([-a,a]) \ge 1-\varepsilon \quad \forall n \in \mathbb{N}.$$

Also: $\{P^{X_n} \mid n \in \mathbb{N}\}$ ist straff. Angenommen, $X_n \xrightarrow{\mathcal{D}} x$ gilt nicht. Dann existiert ein x mit $P(X=x)=0$, $P(X_n \le x) \not\to P(X \le x)$, also existiert ein $\varepsilon > 0$ und eine Teilfolge $(X_{n_k})_{k\in\mathbb{N}}$ mit

$$|P(X_{n_k} \le x) - P(X \le x)| \ge \varepsilon \quad \text{für alle } k \in \mathbb{N} \tag{$*$}$$

Da $\{P^{X_{n_k}} \mid k \in \mathbb{N}\}$ wieder straff ist, existiert nach Satz 17.9 eine Teilteilfolge $\{X_{n_{k_j}} \mid j \in \mathbb{N}\}$ und ein Wahrscheinlichkeitsmaß P_0 mit

$$P^{X_{n_{k_j}}} \xrightarrow{w} P_0.$$

Aus der Voraussetzung $\varphi_n(\Theta) \to \varphi(\Theta)$ und dem „$\Rightarrow$"-Teil (+ Eindeutigkeitssatz) folgt $P_0 = P^X$, also: $X_{n_{k_j}} \xrightarrow{\mathcal{D}} X$, damit $P(X_{n_{k_j}}) \to_{j\to\infty} P(X \le x)$, im Widerspruch zu $(*)$. *q.e.d.*

Noch einmal: Verteilungskonvergenz bezieht sich auf Verteilungen,

$$Y_n \xrightarrow{\mathcal{D}} Y, \quad Y \sim N(0,1) \text{ steht für } P^{Y_n} \xrightarrow{w} N(0,1).$$

Nach Satz 17.10 und Beispiel 16.8 ist hierfür der Nachweis von

$$\mathrm{E}e^{i\Theta Y_n} \to e^{-\frac{1}{2}\Theta^2} \quad \forall \Theta \in \mathbb{R}$$

ausreichend.

18 Der Zentrale Grenzwertsatz

18.1 Identisch verteilte Summenanden

In diesem Absatz wird die aus Abschnitt 8 für bestimmte diskrete Verteilungen bekannte kumulative Form der Normalapproximation verallgemeinert.

Lemma 18.1
Es seien $z_1, \ldots, z_n, w_1, \ldots, w_n \in \{z \in \mathbb{C} \mid |z| \leq 1\}$. Dann gilt

$$\left| \prod_{k=1}^{n} z_k - \prod_{k=1}^{n} w_k \right| \leq \sum_{k=1}^{n} |z_k - w_k|$$

Beweis

$$\begin{aligned}
|z_1 \cdot \cdots \cdot z_n - w_1 \cdot \cdots \cdot w_n| &\leq |z_1 \cdot \cdots \cdot z_n - w_1 \cdot z_2 \cdot \cdots \cdot z_n| \\
&\quad + |w_1 \cdot z_2 \cdot \cdots \cdot z_n - w_1 \cdot w_2 \cdot z_3 \cdot \cdots \cdot z_n| \\
&\quad + \ldots \\
&\quad + |w_1 \cdot \cdots \cdot w_{n-1} \cdot z_n - w_1 \cdot \cdots \cdot w_n| \\
&\leq |z_1 - w_1| \cdot |z_2 \cdot \cdots \cdot z_n| \\
&\quad + |z_2 - w_2| \cdot |w_1 z_3 \cdot \cdots \cdot z_n| \\
&\quad + \ldots \\
&\quad + |z_n - w_n| \cdot |w_1 \cdot \cdots \cdot w_{n-1}|.
\end{aligned}$$

Die jeweils 2. Faktoren sind ≤ 1. *q.e.d.*

Satz 18.2
Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen und identisch verteilten Zufallsvariable mit endlicher, positiver Varianz σ^2 und Erwartungswert μ. Dann gilt für $\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\sqrt{n} \cdot (\overline{X_n} - \mu) \overset{\mathcal{D}}{\to} Z, \quad Z \sim N(0, \sigma^2).$$

Beweis

(mögl. Diplomfrage!!!)

Offensichtlich ist $\overline{X_n} - \mu$ der Mittelwert von $X_1 - \mu, \ldots, X_n - \mu$; wir können also $\mu = 0$ annehmen. Sei φ die charakteristische Funktion zu X_n (hängt *nicht* von n ab, da alle X_n dieselbe Verteilung haben).

Aus $\mathrm{E}X_n = 0$, $\mathrm{E}X_n^2 = \sigma^2$ folgt mit Satz 16.6

$$\varphi(\Theta) = 1 - \frac{1}{2}\Theta^2\sigma^2 + \Theta^2 r(\Theta), \quad \text{mit } \lim_{\Theta \to 0} r(\Theta) = 0.$$

(Taylor zusammen mit φ'' in 0 stetig.)

Sei φ_n die charakteristische Funktion zu $\sqrt{n} \cdot \overline{X_n}$ $\quad (= \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n))$.

Mit Satz 16.4(v) und Satz 16.17 folgt

$$\varphi_n(\Theta) = \varphi\left(\frac{\Theta}{\sqrt{n}}\right)^n = \left(1 - \frac{1}{2n\sigma^2\Theta^2} + \frac{\Theta^2}{n} r\left(\frac{\Theta}{\sqrt{n}}\right)\right)^n$$

Lemma 18.1 liefert

$$\left|\varphi_n(\Theta) - \left(1 - \frac{1}{2n}\sigma^2\Theta^2\right)^n\right| \leq \Theta^2 r\left(\frac{\Theta}{\sqrt{n}}\right),$$

also folgt insgesamt

$$\lim_{n\to\infty} \varphi_n(\Theta) = \lim_{n\to\infty} \left(1 - \frac{1}{2n}\sigma^2\Theta^2\right)^n = \exp\left(-\frac{1}{2}\sigma^2\Theta^2\right) \quad \forall\Theta \in \mathbb{R}$$

Da die rechte Seite die charaktieristischen Funtion zu $N(0, \sigma^2)$ ist, folgt mit Satz 17.10 die Behauptung.
q.e.d.

Bemerkung 18.3
In der Literatur wird der Zentrale Grenzwertsatz häufig in der Form

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} Z, \quad Z \sim N(0,1)$$

angegeben (siehe auch Def. 8.7), es werden also die Partialsummen $Sn = X_1 + \cdots + X_n$ anstelle der Mittelwerte betrachtet. Die linke Seite hat Erwartungswert 0 und Varianz 1, man spricht daher auch von der Standardisierten Summenvarablen. Wegen

$$\frac{1}{\sqrt{n}\sigma}(S_n - n\mu) = \frac{1}{\delta}\sqrt{n}(\overline{X_n} - \mu)$$

und der Eigenschaft $Z_n \xrightarrow{\mathcal{D}} Z,\ c \in \mathbb{Z} \implies cZ_n \xrightarrow{\mathcal{D}} cZ$ (siehe z.B. Aufg. 40(iii)) sind beide Formen äquivalent. Der Zentrale Grenzwertsatz führt auf die folgenden, für die Praxis sehr wichtigen Approximationen:

$$\begin{gathered} N(\mu, \tfrac{\sigma^2}{n}) \text{ für die Verteilung von } \overline{X_n}, \\ N(n\mu, n\sigma^2) \text{ für die Verteilung von } S_n. \end{gathered}$$

Also: Ist eine Zufallsvariable S die Summe einer großen Anzahl unabhängiger, identisch verteilter Zufallsvariablen (mit endlicher Varianz), so ist S in etwa normal verteilt.

Oder: die Verteilung von S kann durch die Normalverteilung mit dem gleichen Erwartungswert und der gleichen Varianz wie S approximiert werden.

In Absatz 18.2: „gleich verteilt“ kann abgeschwächt werden.

Beispiel 18.4
Es seien $X_1, X_2, \ldots$ die zufälligen Lebensdauern (in Stunden) von Glühbirnen eines bestimmten Typs. Annahme: $(X_n)_{n\in\mathbb{N}}$ unabhängig und exponentialverteilt mit Parameter $\frac{1}{100}$ (mittlere Brenndauer 100 Stunden). Wieviele Glühbirnen sollte man kaufen, wenn man eine Gesamtbrenndauer von 1000 Stunden benötigt?

Antwort: unendlich viele!

Modifikation: Wieviele sollte man kaufen, wenn man mit Wahrscheinlichkeit 0,99 sicher sein will, daß sie insgesamt mindestens 1000 Stunden brennen?

Es gilt: $\mathrm{E}X_i = 100$, $\operatorname{var} X_i = 10000$ (Aufg. 21(iii)), also:

$$\begin{aligned} P(S_n \geq 1000) \geq 0,99 \quad &\Longleftrightarrow \\ P(\underbrace{\frac{S_n - \mathrm{E} S_n}{\sqrt{\operatorname{var} S_n}}}_{\approx N(0,1)\text{-verteilt}} \geq \frac{1000 - n \cdot 100}{\sqrt{n} \cdot 100}) &\geq 0,99. \end{aligned}$$

Für eine $N(0,1)$-verteilte Zufallsvariable Z gilt: $P(Z \geq -2,326) \approx 0,99$, der Zentrale Grenzwertsatz führt also auf die Bedingung $\frac{1-n}{\sqrt{n}} \leq -2.326$, d.h. $n \geq 20.54$. Man kauft also 21 Glühbirnen.

(In diesem Beispiel kann man den exakten Wert bestimmen: Ausrechnen liefert, daß 19 Glühbirnen gereicht hätten.)

Wichtiger Spezialfall des Zentralen Grenzwertsatzes:

$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p, \qquad \text{(Satz von de Moivre-Laplace, Satz 8.10)}$$

15. Dezember 1994

18.2 Der Satz von Lindeberg

Satz 18.5 (Lindeberg)
Für jedes $n \in \mathbb{N}$ seien $X_{n1}, \ldots, X_{nr_n}$ unabhängige Zufallsvariablen auf $(\Omega_n, \mathfrak{A}_n, P_n)$ mit endlicher Varianz σ_{nk}^2 und Erwartungswert μ_{nk}, $1 \leq k \leq r_n$, es gelte

$$s_n^2 := \sum_{k=1}^{r_n} \sigma_{nk}^2 > 0.$$

Ist dann die Lindeberg-Bedingung

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

erfüllt, so gilt

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

Bemerkung 18.6

(i) Ist $Y_1, Y_2, \ldots$ eine i.i.d.-Folge auf $(\Omega, \mathfrak{A}, P)$ mit endlicher Varianz $\sigma^2 > 0$, so genügt die Familie $\{X_{nk} \mid n \in \mathbb{N},\ 1 \leq k \leq n\}$ mit $r_n = n$, $X_{nk} = Y_k$, $(\Omega_n, \mathfrak{A}_n, P_n) = (\Omega, \mathfrak{A}, P)$ den Bedingungen des Satzes:

$$s_n^2 = n\sigma^2 \quad \text{und}$$

$$\frac{1}{s_n^2} \sum_{k=1}^{r_n} \int_{|X_{nk}-\mu_{nk}|>\varepsilon s_n} (X_{nk} - \mu_{nk})^2 \, dP_n = \frac{1}{\sigma^2} \int_{|Y_1-\mu_1|>\varepsilon\sqrt{n}\sigma} (Y_1 - \mu_1)^2 \, dP \to_{n\to\infty} 0$$

(majorisierte Konvergenz: $\int (Y_1 - \mu_1)^2 \, dP = \sigma^2 < \infty$)

Also: Satz 18.2 ist ein Spezialfall von Satz 18.5.

(ii) Beim Beweis des Satzes wird deutlich werden, daß die Lindeberg-Bedingung einen dominierenden Einfluß eines einzelnen X_{nk} auf die Summe $X_{n1} + X_{n2} + \cdots + X_{nr_n}$ ausschließt.

Beweis

(zu Satz 18.5)

Wir können $\mu_{nk} = 0$ und $s_n = 1$ annehmen, ersetze X_{nk} sonst durch $\frac{1}{s_n}(X_{nk} - \mu_{nk})$.

Mit $n = 2$ liefert Lemma 16.5

$$\left| e^{i\Theta x} - (1 + i\Theta x - \frac{1}{2}\Theta^2 x^2) \right| \leq \min\left\{ |\Theta x|^2, |\Theta x|^3 \right\}.$$

Für die charakteristische Funktion φ_{nk} zu X_{nk} bedeutet dies:

$$\left| \varphi_{nk}(\Theta) - (1 - \frac{1}{2}\Theta^2 \sigma_{nk}^2) \right| \leq \mathrm{E} \min\left\{ |\Theta X_{nk}|^2, |\Theta X_{nk}|^3 \right\} \quad (< \infty, \text{ da } \mathrm{E}X_{nk}^2 < \infty.)$$

Für jedes $\varepsilon > 0$ läßt sich die rechte Seite nach oben abschätzen durch

$$\int_{|X_{nk}|<\varepsilon} |\Theta X_{nk}|^3 \, dP_n + \int_{|X_{nk}|\geq\varepsilon} |\Theta X_{nk}|^2 \, dP_n \leq \varepsilon \Theta^3 \sigma_{nk}^2 + \Theta^2 \int_{|X_{nk}|\geq\varepsilon} X_{nk}^2 \, dP_n$$

Summiere über k: Der erste Term wird zu $\varepsilon\Theta^3$ (wegen $s_n^2 = 1$), der zweite Term geht wegen der Lindeberg-Bedingung mit $n \to \infty$ gegen 0. Da $\varepsilon > 0$ beliebig war, folgt

$$\lim_{n\to\infty} \sum_{n=1}^{r_n} \left| \varphi_{nk}(\Theta) - \left(1 - \frac{1}{2}\Theta^2 \sigma_{nk}^2\right) \right| = 0 \quad \forall \Theta \in \mathbb{R}. \tag{1}$$

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

$$\lim_{n\to\infty} \left| \prod_{k=1}^{r_n} \varphi_{nk}(\Theta) - \prod_{k=1}^{r_n} \left(1 + \frac{1}{2}\Theta^2 \sigma_{nk}^2\right) \right| = 0 \quad \text{für alle } \Theta \in \mathbb{R}. \tag{2}$$

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

$$\limsup_{n\to\infty} \max\{\sigma_{nk}^2 \mid 1 \leq k \leq r_n\} \leq \varepsilon^2 + \lim_{n\to\infty} \sum_{k=1}^{r_n} \int_{|X_{nk}>\varepsilon|} X_{nk}^2 \, dP_n$$

und hieraus mit $\varepsilon \downarrow 0$:

$$\lim_{n\to\infty} \max\{\sigma_{nk}^2 \mid 1 \leq k \leq \sigma_n\} = 0. \tag{3}$$

Für alle $\Theta \in \mathbb{R}$ existiert also ein $n_0 \in \mathbb{N}$ derart, daß für alle $n \geq n_0$ gilt

$$\left| 1 - \frac{1}{2}\Theta^2 \sigma_{nk}^2 \right| \leq 1 \quad \text{für alle } k \in \{1, \ldots, r_n\}. \tag{4}$$

Für alle solchen n läßt sich die Differenz der Produkte in (2) wegen Lemma 18.1 durch die Summe in (1) abschätzen, also folgt (2) aus (1).

Es bleibt zu zeigen:

$$\lim_{n\to\infty}\left|\underbrace{\prod_{k=1}^{r_n}\exp\left(-\frac{1}{2}\Theta^2\sigma_{nk}^2\right)}_{=e^{-\frac{1}{2}\Theta^2},\text{ charakteristische Funktion zu } N(0,1)} - \prod_{k=1}^{r_n}\left(1-\frac{1}{2}\Theta^2\sigma_{nk}^2\right)\right| = 0, \tag{5}$$

denn aus (5) und (2) folgt wieder mit dem Stetigkeitssatz die Behauptung. Da (4) ab einem n_0 gilt, folgt (5) mit Lemma 18.1, wenn gezeigt werden kann:

$$\lim_{n\to\infty}\sum_{n=1}^{r_n}\left|\exp\left(\frac{1}{2}\Theta^2\sigma_{nk}^2\right)-1+\frac{1}{2}\Theta^2\sigma_{nk}^2\right| = 0. \tag{6}$$

Für reelle x mit $|x|\le\frac{1}{2}$ gilt $|e^x-1-x|\le\frac{1}{2}\sum_{j=2}^{\infty}|x|^j\le x^2$. Dies führt auf

$$\sum_{k=1}^{r_n}\left|\exp(-\frac{1}{2}\Theta^2\sigma_{nk}^2)-1+\frac{1}{2}\Theta^2\sigma_{nk}^2\right|\le\frac{1}{4}\Theta^4\sum_{k=1}^{r_n}\sigma_{nk}^4.$$

Wegen

$$\sum_{k=1}^{r_n}\sigma_{nk}^4\le\underbrace{\max\{\sigma_{nk}^2\mid 1\le k\le r_n\}}_{\to 0\text{ mit } n\to\infty}\underbrace{\sum_{k=1}^{r_n}\sigma_{nk}^2}_{=1}$$

folgt hieraus (6) mit Hilfe von (3). *q.e.d.*

Bemerkung 18.7
Der Zentrale Grenzwertsatz hat eine lange „Verbesserungsgeschichte“ hinter sich, die Lindeberg-Bedingung ist in gewisser Weise das Endresultat. Gelegentlich ist die Lyapunov-Bedingung

$$\lim_{n\to\infty}\frac{1}{s_n}^{2+\delta}\sum_{k=1}^{r_n}\mathrm{E}|X_{nk}-\mu_{nk}|^{2+\delta}=0\quad\text{für ein }\delta>0$$

einfacher zu verwenden.

„Lyapunov $\Longrightarrow$ Lindeberg“ folgt aus

$$\begin{aligned}\int_{|X_{nk}-\mu_{nk}|>\varepsilon s_n}(X_{nk}-\mu_{nk})^2\,dP_n &\le \int_{|X_{nk}-\mu_{nk}|>\varepsilon s_n}\frac{|X_{nk}-\mu_{nk}|^{2+\delta}}{\varepsilon^\delta s_n^\delta}\,dP_n\\ &\le \frac{1}{\varepsilon^\delta}\frac{1}{s_n^\delta}\mathrm{E}|X_{nk}-\mu_{nk}|^{2+\delta}\end{aligned}$$

Die anwendungsorientierte Substanz von Satz 18.5 läßt sich wie folgt zusammenfassen:

Eine Zufallsvariable, die die Summe vieler einzener, unabhängiger, „individuell vernachlässigbarer“ Summanden ist, ist ungefähr normalverteilt.

18.3 Anwendungen

18.3.1 Ein Sammelproblem

Einer Urne mit n Kugeln werden solange Kugeln zufällig und mit Zurücklegen entnommen, bis insgesamt r_n verschiedene Kugeln gezogen worden sind, s_n sei die Anzahl der hierfür notwendigen Züge.

Anwendungsbeispiel: (G. Polya, ZAMM 1930, Eine Wahrscheinlichkeitsaufgabe zur Kundenwerbung) Jedem Paket Haferflocken liegt eines von n verschiedenen Bildern bei. Wie viele Pakete muß man kaufen, bis man r_n verschiedene Bilder hat?

Angenommen, $r = \lceil \rho n \rceil$ mit $0 < \rho < 1$ (man muß mindestens $100\rho\%$ der Bilder gesammelt haben).

Bei n Kugeln werden die wiederholten Entnahmen mit Hilfe von

$$(\Omega_n, \mathfrak{A}_n, P_n) = (\{1, \ldots, n\}, \wp(\{1, \ldots, n\}), \text{Gleichverteilung})^N N$$

beschrieben (siehe Absatz 14.2). Für $k = 1, \ldots, r_n$ sei X_{nk} die Anzahl der Züge, die nötig sind, um eine neue Kugel zu erhalten, wenn bereits $k-1$ verschiedene Kugeln entnommen (und zurückgelegt) worden sind; $X_1 \equiv 1$.

Dann sind $X_{n1}, \ldots, X_{nr_n}$ unabhängig, und X_{nk} ist geometrisch verteilt mit Parameter $\frac{n-k+1}{n}$ (siehe Absatz 6.2.3). Ist $Y \sim \operatorname{geom}(p)$, dann

$$\mathrm{E}Y = \frac{1}{p}, \quad \operatorname{var} Y = \frac{q}{p^2}\ (q := 1-p), \quad \mathrm{E}Y^4 = \frac{24}{p^4}.$$

Damit $\mu_n := \mathrm{E}S_n = \sum_{k=1}^{r_n} \frac{n}{n-k+1}$,

$$\begin{aligned} s_n^2 &:= \sum_{k=1}^{r_n} \frac{1 - \frac{n-k+1}{n}}{(\frac{n-k+1}{n})^2} \\ &= n \sum_{k=1}^{r_n} \frac{k-1}{(n-k+1)^2} \end{aligned}$$

Nachweis Lyapunov mit $\delta = 2$: Ist $Y \sim \operatorname{geom}(p)$, so gilt

$$\begin{aligned} \mathrm{E}(Y - \tfrac{1}{p})^4 \le \mathrm{E}(2\max\{Y, \tfrac{1}{p}\})^4 &= 16\,\mathrm{E}\max\{Y^4, \tfrac{1}{p^4}\} \\ &\le 16(\mathrm{E}Y^4 + \tfrac{1}{p^4}) \\ &\le \tfrac{400}{p^4}, \quad \text{also} \\ \sum_{k=1}^{r_n} \mathrm{E}(X_{nk} - \mathrm{E}X_{nk})^4 &\le 400 \sum_{k=1}^{r_n} \frac{n^4}{(n-k+1)^4} \\ &\le 400\,\lceil \rho n \rceil \frac{1}{(1-\rho)^4} \\ &= O(n) \end{aligned}$$

Wegen $s_n^2 \ge n \cdot \frac{1}{n^2} \sum_{k=1}^{r_n} (k-1) \ge \frac{1}{n}\frac{1}{2}(\rho n - 1)(\rho n - 2) \quad (\approx cn)$
liefert dies

$$\lim_{n\to\infty} \frac{1}{s_n^{2+2}} \sum_{k=1}^{r_n} \mathrm{E}\left(X_{nk} - \mathrm{E}X_{nk}^4\right) = 0$$

20. Dezember 1994

Urne mit n Kugeln, Ziehen mit Zurücklegen bis $r_n = \lceil \rho n \rceil$, $0 < \rho < 1$ verschiedene Kugeln aufgetaucht sind. S_n: Anzahl der nötigen Züge.

S_n genügt dem zentralen Grenzwertsatz: $\frac{S_n - \mathrm{E}S_n}{\sqrt{\operatorname{var} S_n}} \xrightarrow{\mathcal{D}} Z$, $Z \sim N(0,1)$.

$$\begin{aligned}
\mathrm{E}S_n &= \sum_{k=1}^{r_n} \frac{n}{n-k+1} \\
\frac{1}{n} \cdot \mathrm{E}S_n &= \sum_{k=1}^{\lceil \rho n \rceil} \frac{1}{1 - \frac{k-1}{n}} \cdot \frac{1}{n} \\
&= \int_0^\rho \frac{1}{1-x}\, dx + O(\tfrac{1}{n}) \\
&= -\log(1-\rho) + O(\tfrac{1}{n}),
\end{aligned}$$

analog:

$$\frac{1}{n} \operatorname{var} S_n @>> n \to \infty > \int_0^\rho \frac{x}{(1-x)^2}\, dx = \frac{\rho}{1-\rho} + \log(1-\rho).$$

Mit $a(\rho) := -\log(1-\rho)$, $b(\rho) := \sqrt{\frac{\rho}{1-\rho} + \log(1-\rho)}$ folgt

$$\frac{S_n - a(\rho)n}{b(\rho)\sqrt{n}} \xrightarrow{\mathcal{D}} Z, \quad Z \sim N(0,1).$$

Numerisches Beispiel I

Wie groß muß Ihr Bekanntenkreis sein, damit mit Wahrscheinlichkeit ≥ 0.95 an 180 Tagen im Jahr Geburtstag gefeiert werden kann?

$n = 365$, $\rho = \frac{180}{365}$, $P(Z \geq 1.645) \approx 0.05$. Weiterhin gilt

$$S_n \geq k \quad \Longleftrightarrow \quad \frac{S_n - a(\rho)n}{b(\rho)\sqrt{n}} \geq \frac{k - a(\rho)n}{b(\rho)\sqrt{n}}$$

Führt auf $k \geq a(\rho)n + 1.645 \cdot b(\rho)\sqrt{n}$, ich erhalte $k \geq 266$.

Das Verhalten von $a(\rho)$ und $b(\rho)$ mit $\rho \uparrow 1$ läßt vermuten, daß bei $\rho = 1$ (man zieht solange, bis jede Kugel mindestens einmal erschienen ist) ein anderes Verhalten folgt. Man hat

$$\operatorname{var} S_n = n \sum_{k=1}^{n} \frac{k-1}{(n-k+1)^2} = n \sum_{k=1}^{n} \frac{n-k}{k^2} = n^2 \sum_{k=1}^{n} \frac{1}{k^2} - n \sum_{k=1}^{n} \frac{1}{k} = n^2 \cdot \frac{\pi^2}{6} + o(n^2),$$

$$\operatorname{var} X_{n,n} = \frac{1 - \frac{1}{n}}{\left(\frac{1}{n}\right)^2} = n^2 + o(n^2),$$

d.h. die einzelnen Summanden sind nicht mehr vernachlässigbar ($\max_{1 \leq k \leq n} \sigma_{nk}^2 \neq o(s_n^2)$), die Lindeberg-Bedingung ist nicht erfüllt.

In der Tat gilt bei $\rho = 1$:

$$\frac{S_n - n \cdot \log n}{n} \xrightarrow{\mathcal{D}} Z, \quad Z \sim \text{Gumbel (Bsp. 17.2)} \tag{$*$}$$

Numerisches Beispiel II

Soll der Bekanntenkreis so groß sein, daß mit Wahrscheinlichkeit ≥ 0.95 täglich gefeiert werden kann, so erhält man die Bedingung $k \geq 365 \cdot \log(365) + 365 \cdot \underbrace{2.97}_{\text{Quantil der Gumbelverteilung}} \approx 3237.51$, man bräuchte also einen recht umfangreichen Bekanntenkreis.

Heuristisches Argument für $(*)$: Es sei $A_{i,k}$ das Ereignis, daß die i-te Kugel nach k Zügen noch nicht aufgetaucht ist. Dann gilt

$$P(A_{i,k}) = (1 - \frac{1}{n})^k.$$

Wir unterstellen nun, daß die Ereignisse $A_{1,k}, \dots, A_{n,k}$ zumindest näherungsweise unabhängig sind. Dann:

$$P(S_n \leq k) = P(\bigcap_{i=1}^{n} A_{i,k}^c) \approx (1 - (1 - \frac{1}{n})^k)^n.$$

Setze nun $k = k_n = n \cdot \to n + xn$, $x \in \mathbb{R}$ fest. Dann folgt

$$\begin{aligned} P(\frac{S_n - n \cdot \log n}{n} \leq x) &= P(S_n \leq k_n) \\ &\approx (1 - \exp(\underbrace{k_n \cdot \underbrace{\log(1 - \frac{1}{n})}_{\approx \frac{1}{n}}}_{-\log n - x}))^n \\ &\approx (1 - \frac{1}{n} e^{-x})^n \\ &\to e^{-e^{-x}} \end{aligned}$$

18.3.2 Rekorde

Es seien $X_1, X_2, \dots$ eine i.i.d.-Folge auf $(\Omega, \mathfrak{A}, P)$ mit *stetiger* Verteilungsfunktion F. Wir setzen

$$R_n = \begin{cases} 1, & X_n > \max\{X_1, \dots, X_{n-1}\} \\ 0, & \text{sonst} \end{cases}$$

R_n zeigt also an, ob in der n-ten Versuchswiederholung ein Rekord eintritt oder nicht.

Mit Fubini erhält man für $i \neq j$

$$P(X_i = X_j) = P^{X_i} \otimes P^{X_j}(\{(x,x) \mid x \in \mathbb{R}\}) = \int \underbrace{(\int 1_{\{X\}}(y) P^{X_i}\,(dy))}_{F(x) - F(x-) = 0} P^{X_j}\,(dx) = 0,$$

also ist

$$A := \{\omega \in \Omega \mid \exists i \neq j: \; X_i(\omega) = X_j(\omega)\} = \bigcup_{i=2}^{\infty} \bigcup_{j=1}^{i-1} \{X_i = X_j\}$$

eine P-Nullmenge. Die Wahrscheinlichkeit für das Auftreten mehrfacher Werte (engl.: ties) ist 0, A kann also im folgenden vernachlässigt werden.

Sei $\mathbb{P}_n$ die Menge der Permutationen von $\{1, \dots, n\}$. Sei $\Psi_n : \Omega \to \mathbb{P}_n$ diejenige zufällige Permutation, die $X_1, \dots, X_n$ in aufsteigende Reihenfolge bringt, d.h.

$$\Psi_n = \pi \quad \Longleftrightarrow \quad X_{\pi(1)} < X_{\pi(2)} < \cdots < X_{\pi(n)}.$$

Hat man beispielsweise in den ersten 4 Versuchen die Werte 1.1, 2, -4.3 und 0.1 erhalten, so gilt

$$\Psi_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$

$\Psi_n^{-1}(\{\pi\})$ ist Durchschnitt von Mengen $\{X_i < X_j\}$; versieht man $\mathbb{P}_n$ mit der Potenzmenge, so ist Ψ_n meßbar. Für jedes feste $\pi \in \mathbb{P}_n$ haben die Vektoren $(X_{\pi(1)}, \dots, X_{\pi(n)})$ und $(X_1, \dots, X_n)$ dieselbe Verteilung, nämlich $(P^{X_1})^{\otimes n}$. Also folgt mit $B := \{x_1 < x_2 < \cdots < x_n\}$

$$P(\Psi_n = \pi) = P((X_{\pi(1)}, \dots X_{\pi(n)}) \in B) = P((X_1, \dots, X_n) \in B),$$

d.h. $P(\Psi_n = \pi)$ hängt nicht von π ab. Folglich ist Ψ_n auf $\mathbb{P}_n$ gleichverteilt.

$$P(\Psi_n = \pi) = \frac{1}{n!} \quad \text{für alle } \pi \in \mathbb{P}_n. \tag{1}$$

Hieraus:

$$P(R_n = 1) = P(\Psi_n \in \{\pi \in \mathbb{P}_n \mid \pi(n) = n\}) = \frac{1}{n}. \tag{2}$$

Wegen $\{R_{n+1} = 1\} \cap \{\Psi_n = \pi\} = \{\Psi_{n+1} = \tilde{\pi}\}$ mit

$$\tilde{\pi}(i) = \begin{cases} \pi(i), & i \leq n, \\ n+1, & i = n+1 \end{cases}$$

folgt aus (1) und (2)

$$P(\Psi_n = \pi, R_{n+1} = 1) = \frac{1}{(n+1)!} = P(\Psi_n = \pi) P(R_{n+1} = 1).$$

d.h. Ψ_n und R_{n+1} sind unabhängig. Wegen $\sigma(R_1, \dots, R_n) \subset \sigma(\Psi_n)$ sind dann auch $\sigma(R_1, \dots, R_n)$ und $\sigma(R_n + 1)$ unabhängig. Für beliebige $i_1, i_2, \dots, i_n, j_1, \dots, j_n \in \{0, 1\}$ bedeutet dies:

$$\begin{aligned} P(R_{i_1} = j_1, \dots, R_{i_n} = j_n) &= P(Ri_1 = j_1, \dots, R_{i_{n-1}} = j_{n-1}) \cdot P(R_{i_n} = j_n) \\ &= \dots \\ &= P(R_{i_1} = j_1) \cdot \dots \cdot P(R_{i_n} = j_n), \end{aligned}$$

d.h. die Zufallsvariablen $R_1, R_2, \dots$ sind unabhängig.
Wieviele Rekorde gibt es unter den ersten n Versuchen? Sei dazu $S_n = \sum_{i=1}^n R_i$. Es gilt

$$\mathrm{E} S_n = \sum_{i=1}^n P(R_i = 1) = \sum_{i=1}^n \frac{1}{i},$$
$$\operatorname{var} S_n = \sum_{i=1}^n \operatorname{var}(R_i) = \sum_{i=1}^n \frac{1}{1 - \frac{1}{i}} ??? = \sum_{i=1}^n \frac{1}{i} - \sum_{i=1}^n \frac{1}{i^2}$$

insbesondere:

$$\frac{\mathrm{E}S_n}{\log n} \to 1, \quad \frac{\operatorname{var} S_n}{\log N} \to 1.$$

Mit Chebychev:

$$P\left(\left|\frac{S_n}{\mathrm{E}S_n} - 1\right| < \varepsilon\right) \leq \frac{1}{\varepsilon^2} \cdot \frac{1}{(\mathrm{E}Sn)^2} \operatorname{var}(S_n) \to 0,$$

d.h.

$$\frac{S_n}{\log n} \xrightarrow{\mathrm{P}} 1.$$

Der Zentrale Grenzwertsatz erlaubt wieder eine genauere Aussage: Mit

$$s_n := (\operatorname{var} S_n)^{\frac{1}{2}} \sim (\log n)^{\frac{1}{2}},$$
$$\frac{1}{s_n^3} \sum_{i=1}^{n} \mathrm{E}|R_i|^3{}_{R_i} ??? = \frac{1}{s_n^3} \mathrm{E}S_n \to 0,$$

d.h. die Lyapunov-Bedingung ist erfüllt und damit: „Der Zentrale Grenzwertsatz gilt für S_n“, d.h.

$$\frac{S_n - \mathrm{E}S_n}{\sqrt{\operatorname{var} S_n}} \xrightarrow{\mathcal{D}} Z, \quad Z \sim N(0,1).$$

Verwendet man wieder die Eigenschaften der Verteilungskonvergenz (siehe Aufgabe 43), so folgt hieraus

$$\frac{S_n - \log n}{\sqrt{\log n}} \xrightarrow{\mathcal{D}} Z, \quad Z \sim N(0,1).$$

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

22. Dezember 1994

19 Zufallsvektoren

Ziel: Verallgemeinerung von $\mathbb{R}$ auf $\mathbb{R}^d$.

Die Elemente von $\mathbb{R}^d$ sind hier (Spalten-)vektoren, $\langle\cdot\rangle$ ist das Skalarprodukt

$$\langle x, y\rangle = \sum_{k=1}^{d} x_k y_k \quad \text{für} \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}, \; y = \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} \in \mathbb{R}^d.$$

Ist A^t die Transponierte einer $n \times m$-Matrix $A \in \mathbb{R}^{n\times m}$ und faßt man Elemente aus $\mathbb{R}^d$ als $d \times 1$-Matrizen auf, so gilt $\langle x, y\rangle = x^t y$. Ist X ein d-dimensionaler Zufallsvektor auf $(\Omega, \mathfrak{A}, P)$, so wird durch

$$\varphi_x : \begin{cases} \mathbb{R}^d \to \mathbb{C} \\ \theta \mapsto \mathrm{E}\exp(i\langle\theta, x\rangle) \end{cases}$$

die *charakteristische Funktion* zu X defniert. Auch im Mehrdimensionalen gilt

$$\varphi_X = \varphi_Y \implies X \stackrel{\mathcal{D}}{\to} Y.$$

Existieren die Erwartungswerte der Komponenten $X_1, \ldots, X_d$ eines Zufallsvektors X, so nennen wir $\mathrm{E}X := (\mathrm{E}X_1, \ldots, \mathrm{E}X_d)^t$ den *Erwartungswertvektor* zu X. Hat man $\mathrm{E}X_i^2 < \infty$ $(i = 1, \ldots, d)$, so heißt die $d \times d$-Matrix

$$\operatorname{cov} X := \begin{pmatrix} \operatorname{cov}(X_1, X_1) & \cdots & \operatorname{cov}(X_1, X_d) \\ \vdots & & \vdots \\ \operatorname{cov}(X_d, X_1) & \cdots & \operatorname{cov}(X_d, X_d) \end{pmatrix}$$

die *Kovarianzmatrix* zu X. Auf der Diagonalen stehen die Varianzen der einzelnen Komponenten. Dehnt man den Erwartungswertbegriff (komponentenweise) auf zufällige Matrizen aus, so gilt

$$\operatorname{cov}(X) = \mathrm{E}(X - \mathrm{E}X) \cdot (X - \mathrm{E}X)^t.$$

Man erhält leicht die folgenden Rechenregeln: Mit $A \in \mathbb{R}^{l \times d}$, $b \in \mathbb{R}^l$, $Y := AX + b$ gilt:

$$\begin{aligned} \mathrm{E}Y &= A\mathrm{E}X + b \\ \operatorname{cov} Y &= \mathrm{E}((AX + b - (A\mathrm{E}X + b))((AX + b) - (A\mathrm{E}X + b))^t) \\ &= A\mathrm{E}(X - \mathrm{E}X)(X - \mathrm{E}X)^t A^t \\ &= A \operatorname{cov} X A^t \end{aligned}$$

Besonders wichtig sind die höherdimensionalen Verallgemeinerungen der Normalverteilung: X heißt *(d-dimensional) normalverteilt*, wenn $a^t X$ (1-dimensional) normalverteilt ist für $a \in \mathbb{R}^d$. Hierbei betrachten wir eine Zufallsvariable Z mit $P(Z = c) = 1$ für ein $c \in \mathbb{R}$ als $N(c, 0)$-verteilt.

Zu einem solchen X existieren der Erwartungswert μ und die Kovarianzmatrix σ, denn

$$X_i = (0, \ldots, \underset{\text{Position } i}{1}, \ldots, 0)X$$

hat als normale Zufallsvariable ein endliches zweites Moment. Ist nun X normalverteilt mit Erwartungswert μ und Kovarianzmatrix Σ, so gilt nach obigen Rechenregeln

$$\mathrm{E}a^t X = a^t \mu, \ \operatorname{var}(a^t X) = a^t \Sigma a, \quad \text{d.h. } a^t X \sim N(a^t \mu, a^t \Sigma a) \ \forall a \in \mathbb{R}^d$$

und damit

$$\varphi_{a^t X}(\theta) = \exp(i\theta(a^t \mu) - \frac{1}{2}(a^t \Sigma a)\theta^2) \ \forall \theta \in \mathbb{R}$$

für die charakteristische Funktion $\varphi_{a^t X}$ zu $a^t X$.

Für alle $\theta \in \mathbb{R}^d$ gilt

$$\varphi_X(\theta) = \mathrm{E}\exp(i\langle \theta, X \rangle) = \varphi_{\theta^t X}(1) = \exp(i\langle \theta, \mu \rangle - \tfrac{1}{2}\theta^t \Sigma \theta).$$

Eine Konsequenz: Auch im Höherdimensionalen wird eine Normalverteilung durch die „ersten beiden Momente“ festgelegt, wir können als kurz schreiben:

$$X \sim N_d(\mu, \Sigma).$$

Weiter folgt hieraus (oder direkt aus der Definition)

$$\begin{aligned}
\varphi_{AX+b}(\theta) &= \mathrm{E}\exp(i\langle\theta, AX+b\rangle) \\
&= \exp(i\langle\theta, b\rangle)\mathrm{E}\exp(i\langle A^t\theta, X\rangle) \\
&= e^{i\langle\theta,b\rangle}\varphi_X(A^t\theta) \\
&= \exp(i\langle\theta, A\mu+b\rangle - \tfrac{1}{2}\theta^t(A\Sigma A^t)\theta),
\end{aligned}$$

also

$$X \sim N_d(\mu, \Sigma),\ A \in \mathbb{R}^{l\times d},\ b \in \mathbb{R}^l \implies Y := AX + b \sim N(A\mu + b, A\Sigma A^t).$$

Frage nun: Zu welchen $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ existiert eine solche Verteilung? Klar: Σ muß symmetrisch verteilt und positiv semidefinit sein ($a^t\Sigma a$ ist die Varianz von a^tX). Zu jeder solchen Matrix Σ existiert ein $A \in \mathbb{R}^{d\times d}$ mit $AA^t = \Sigma$. Es sei nun $X_1, \ldots, X_d$ unabhängige, $N(0,1)$-verteilte Zufallsgrößen. Für $X := (X_1, \ldots, X_d)$ gilt

$$a^tX = \sum_{k=1}^d a_kx_k \sim N(0, \sum_{k=1}^d a_k^2) = N(0, aI_da^t), \quad \text{d.h.} X \sim N_d(0, I_d).$$

Für $Y := AX + \mu$ folgt

$$Y \sim N(A0 + \mu, AI_dA^t) = N_d(\mu, \Sigma).$$

Mit dieser Konstruktion und einer mehrdimensionalen Variante von Aufgabe 20(i) folgt, daß bei nichtsingulärem Σ

$$f(x \mid \mu, \Sigma) := (2\pi)^{-\frac{d}{2}}(\det\Sigma)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu))$$

eine ℓ^d-Dichte zu $N_d(\mu, \Sigma)$ ist.

Die Verallgemeinerung von Gesetzen der großen Zahlen ist kanonisch.

Zentraler Grenzwertsatz: $X_n \overset{\otimes}{\to} X$ (alles d-dimensionale Zufallsvektoren) bedeute $\lim_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitspunkte x von F (Beispiel 14.8). Die Theorie aus Abschnitt 17 kann also auf $d > 1$ verallgemeinert werden, nur Skorohod braucht einen grundlegend anderen Beweis. Insbesondere gilt der Stetigkeitssatz für charakteristische Funktionen:

$$X_n \overset{\mathcal{D}}{\to} X \iff \varphi_{X_n}(\theta) \to \varphi_X(\theta)\ \forall\theta \in \mathbb{R}^d \tag{$*$}$$

Satz 19.1
Sei $(X_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge von d-dimensionalen Zufallsvektoren mit Erwartungswert μ und Kovarianzmatrix Σ, $\overline{X_n} := \frac{1}{n}\sum_{i=1}^n X_i$. Dann gilt

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

Beweis

Sei $\theta \in \mathbb{R}^d$, $Y_i := \theta^t X_i \ \forall i \in \mathbb{N}$. Mit Satz 18.2 folgt

$$\sqrt{n}(\overline{Y_n} - \theta^t \mu) \xrightarrow{\mathcal{D}} Z_\theta,\ Z_\theta \sim N(0, \theta^t \Sigma \theta),$$

also

$$\mathrm{E}\exp(i(\sqrt{n}(\overline{Y_n} - \theta^t\mu))) \overset{n\to\infty}{\to} \exp(-\frac{1}{2}\theta^t \Sigma \theta).$$

Wegen $\theta^t \overline{X_n} = \overline{Y_n}$ gilt $\varphi_{\sqrt{n}(\overline{X_n}-\mu)}(\theta) = \mathrm{E}\exp(i(\sqrt{n}(\overline{Y_n} - \theta^t\mu)))$ folgt die Behauptung mit $(*)$, da die Grenzfunktion die charakteristische Funktion zu $N_d(0,\Sigma)$ ist. *q.e.d.*

Beispiel 19.2
Gegeben sei ein Experiment mit d möglichen Resultaten $X_{n,k} = 1$, wenn das Resultat k in der n-ten Wiederholung erscheint, 0 sonst.

$$\begin{aligned}
S_n &:= X_1 + \cdots + X_n \quad (= n\overline{X_n}) \text{ ist multinomialverteilt (Abschnitt 6.2.5)}\\
\operatorname{var} X_{1i} &= p_i(1-p_i) \quad (p_i\text{: Wahrscheinlichkeit für Resultat } i)\\
\operatorname{cov}(X_i, X_j) &= -p_i p_j \text{ für } i \neq j\\
&\Longrightarrow (\Sigma)_{i,j} = \begin{cases} p_i(1-p_i) & \text{für } i = j \\ -p_i p_j & \text{sonst} \end{cases}, \quad \mu = \begin{pmatrix} p_1 \\ \vdots \\ p_d \end{pmatrix}.
\end{aligned}$$

Satz 19.1 liefert

$$\frac{1}{\sqrt{n}}(S_n - n\mu) \xrightarrow{\mathcal{D}} Z,\ Z \sim N_d(0,\Sigma).$$

Beim χ^2-Anpassungstest:

$$T(S_n) = \sum_{j=1}^d \frac{(S_{n_j} - np_j)^2}{np_j}.$$

Es gilt:

$$T(S_n) = h(\frac{1}{\sqrt{n}}(S_n - n\mu)) \quad \text{mit} \quad h: \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ (x_1,\ldots,x_d) \mapsto \sum_{j=1}^d \frac{x_j^2}{p_j} \end{cases}.$$

Die mehrdimensionale Variante von Satz 17.6 ergibt

$$h(\frac{1}{\sqrt{n}}(S_n - n\mu)) \xrightarrow{\mathcal{D}} h(Z),$$

und dies liefert (nach einigen Rechnungen, später) einen alternativen (durchsichtigeren) Zugang zu Satz 8.13.

Teil IV

Statistik

10. Januar 1995

20 Ein Blick auf nichtparametrische Verfahren

(Ausführlicher im SoSe95)

Allgemeiner Rahmen (siehe auch Abschnitt 7): Gegen ist eine Beobachtung, die *Daten* X aus einem Stichprobenraum $\mathfrak{X}$, z.B. $x = (x_1, \dots, x_n) \in \mathfrak{X} = \mathbb{R}^n$. Diese werden als Wert (Realisierung) einer Zufallsgröße X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ betrachtet, also $x = X(\omega)$ für ein $\omega \in \Omega$.

Sei $\mathfrak{B}(\mathfrak{X})$ eine geeignete σ-Algebra auf $\mathfrak{X}$. Wir setzen voraus, daß die (unbekannte) Verteilung $\mathbb{P}^X$ Element einer (vorgegebenen) Klasse $\mathfrak{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$ ist. Anhand von x sollen Aussagen über $\mathbb{P}^X$ gemacht werden.

Im Falle $\mathfrak{P} = \{P_\theta \mid \theta \in \Theta\}, \Theta \subset \mathbb{R}^d$ sprechen wir von einer parametrischen Klasse (Abschnitt 21ff.). Hier:

$$(\mathfrak{X}, \mathfrak{B}(\mathfrak{X})) = (\mathbb{R}^n, \mathfrak{B}^n),\ \mathfrak{P} = \{P'^{\otimes n} \mid P' \in \mathfrak{P}\}$$

mit einer Familie $\mathfrak{P}'$ von Wahrscheinlichkeitsmaßen (beispielsweise alls Wahrscheinlichkeitsmaße, oder alle Wahrscheinlichkeitsmaße mit endlichem 1. Moment) auf $(\mathbb{R}, \mathfrak{B})$. Es gilt dann: $X = (X_1, \dots, X_n)$ mit unabhängigen, identisch verteilten Zufallsvariablen $X_1, \dots, X_n$. Man sagt auch: $X_1, \dots, X_n$ ist eine *Stichprobe* vom Umfang n aus der Verteilung P'.

P' wird durch die zugehörige Verteilungsfunktion F festgelegt. $F(x)$ ist die Wahrscheinlichkeit eines bestimmten Ereignisses (ein Resultat $\leq x$ erscheint), kann also durch die zugehörige relative Häufigkeit geschätzt werden (siehe auch Beispiel 7.3).

Definition 20.1
Die Funktion $\hat{F}_n : \mathbb{R} \to [0, 1]$,

$$\hat{F}_n(x) := \frac{1}{n} \cdot \sum_{i=1}^{n} 1_{(-\infty, x]}(X_i),$$

heit die *empirische Verteilungsfunktion* zu $X_1, \dots, X_n$.

Empirische Verteilungsfunktionen sind stückweise konstant und haben in X_i einen Sprung der Höhe $\frac{1}{n} \cdot \#\{j \mid X_j = X_i\}$.

Sind keine Mehrfachwerte („ties“) vorhanden, dann sind alle Sprünge von der Höhe $\frac{1}{n}$.

Bild 1

Beachte: $\hat{F}_n$ ist wieder eine (nunmehr stochastische) Verteilungsfunktion, insbesondere stetig von rechts.

Definition 20.2
Wir nennen $X_{(1:n)}, \dots, X_{(n:n)}$ die Ordnungsstatistiken zu $X_1, \dots, X_n$, wenn gilt

$$X_{(1:n)} \leq X_{(2:n)} \leq \dots \leq X_{(n:n)}, \quad \#\{1 \leq i \leq n \mid X_{(i:n)} = x\} = \#\{1 \leq i \leq n \mid X_i = x\}\ \forall x \in \mathbb{R}$$

Sind keine mehrfachen Werte vorhanden, so ist $(X_{(1:n)}, \dots, X_{(n:n)})$ einfach die in aufsteigender Reiehfolge angeordnete Sichprobe und es gilt

$$\hat{F_n}(X_{(i:n)}) = \frac{i}{n}.$$

Wie gut ist $\hat{F_n}$ als Schätzer von F?

Aus dem starken Gesetz der großen Zahlen folgt unmittelbar

$$\hat{F_n}(x) \stackrel{\text{f.s.}}{\to} F(x)$$

für jedes feste $x \in \mathbb{R}$, wobei wir eine Folge $X_1, X_2, \dots$ voraussetzen. Die folgende stärkere Aussage wird in der älteren Literatur gelegentlich der „Fundamentalsatz der Statistik“ genannt.

Satz 20.3 (Glivenko-Cantelli)
Es sei $X_1, X_2, \dots$ eine i.i.d.-Folge von Zufallsvariablen mit Verteilungsfunktion F, $\hat{F_n}$ sei die empirische Verteilungsfunktion zu $X_1, \dots, X_n$. Dann gilt

$$\sup_{x\in\mathbb{R}} \left| \hat{F_n}(x) - F(x) \right| \to_{n\to\infty} 0 \text{ f.s.}$$

Beweis

Vollständige Schreibweise für $\hat{F_n}$: $\hat{F_n}(\cdot, \omega)$ (macht die Abhängigkeit von ω deutlich).

Sei zunächst $x \in \mathbb{R}$ fest, setze $Y_i := 1_{(-\infty,x]}(X_i)$, $Z_i := 1_{(-\infty,x)}(X_i)$. Auf diese Folgen kann das starke Gesetz der großen Zahlen angewendet werden, wodurch man Mengen $A_x, B_x \in \mathfrak{A}$ mit $\mathbb{P}(A_x) = 0, \mathbb{P}(B_x) = 0$ erhält mit

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} Y_i(\omega) &= \hat{F_n}(x, \omega) \to F(x) \quad \forall \omega \notin A_x, \\
\frac{1}{n}\sum_{i=1}^{n} Z_i(\omega) &= \hat{F_n}(x-, \omega) \to F(x-) \quad \forall w \notin B_x
\end{aligned}$$

Es sei $x_{mk} := F^{-1}(\frac{k}{m})$,

$$A := \bigcup_{m=1}^{\infty} \bigcup_{k=0}^{m} A_{x_{mk}} \cup B_{x_{mk}}.$$

Außerhalb der Nullmenge A gilt

$$\lim_{n\to\infty} D_{mn}(\omega) = 0 \quad \text{für alle } m, \tag{$*$}$$

wobei

$$\begin{aligned}
D_{mn} &:= \max\left\{ D_{mn}^{(1)}, D_{mn}^{(2)} \right\}, \\
D_{mn}^{(1)}(\omega) &:= \max\left\{ \left| \hat{F_n}(x_{mk}, \omega) - F(x_{mk}) \right| \mid 0 \le k \le m \right\}, \\
D_{mn}^{(2)}(\omega) &:= \max\left\{ \left| \hat{F_n}(x_{mk}-, \omega) - F(x_{mk}) \right| \mid 0 \le k \le m \right\}.
\end{aligned}$$

Es seien $x \in \mathbb{R}$, $m \in \mathbb{N}$ beliebig. Dann existiert ein $k \in \{1, \ldots, m\}$ mit $x \in [x_{m,k-1}, x_{mk})$ (genauer: Man kann sich auf $x \in [F^{-1}(0), F^{-1}(1))$ beschränken, denn mit Wahrscheinlichkeit 1 gilt: $\hat{F_n}(x) = 0 = F(x)\ \forall x \le F^{-1}(0)$, $\hat{F_n}(x) = 1 = F(x)\ \forall x \ge F^{-1}(1)$). Also:

$$\begin{aligned} \hat{F_n}(x,\omega) &\le \hat{F_n}(x_{mk}-,\omega) \quad \text{(Monotonie von } \hat{F_n}) \\ &\le F(x_{mk}-) + D^{(2)}_{mn}(\omega) \\ &\le F(x) + \tfrac{1}{m} + D^{(2)}_{mn}(\omega). \end{aligned}$$

Wegen $F(x) \ge F(x_{m,k-1})$ und da $F(x_{mj}-) \le \frac{j}{m} \le F(x_{mj})$ (Definition von x_{mj}, siehe auch Lemma 12.15) impliziert, daß

$$F(x_{m,k-1}) \ge \frac{k-1}{m} = \frac{k}{m} - \frac{1}{m} \ge F(x_{mk}-) - \frac{1}{m}$$

gilt. Analog erhält man

$$\hat{F_n}(x,\omega) \ge F(x) - \frac{1}{m} - D^{(1)}_{mn}(\omega),$$

also

$$\left|\hat{F_n}(x,\omega) - F(x)\right| \le \frac{1}{m} + D_{mn}(\omega).$$

Die rechte Seite hängt nicht von x ab, also folgt mit $(*)$

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\hat{F_n}(x,\omega) - F(x)\right| \le \frac{1}{m} \quad \text{für alle } w \in A.$$

Da $m \in \mathbb{N}$ beliebig war, folgt hieraus mit $m \to \infty$ die Behauptung. *q.e.d.*

Interessiert man sich nicht für das „gesamte P'“, sondern nur für einen von P' (also F) abhängigen Wert (also irgendeinen Parameter der Verteilung, aus der die Stichprobe entnommen wurde), so liegt es nahe, diesen Wert durch den analogen, zu $\hat{F_n}$ gehörenden Wert zu schätzen. Ist beispielsweise $\mathfrak{P}'$ die Menge aller P' mit endlichem Erwartungswert

$$\mu = \int xP'\,(dx) = \int XF\,(dx),$$

so führt dieses *Einsetzprinzip* zum Schätzer

$$\hat{\mu} = \int x\hat{F_n}\,(dx) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X_n},$$

Man schätzt also den unbekannten Erwartungswert durch den Mittelwert der Stichprobe. Analog: Bezeichnet

$$\mathrm{med}(F) := \tfrac{1}{2}(\sup\{x \mid F(x) < \tfrac{1}{2}\} + \inf\{x \mid F(x) > \tfrac{1}{2}\})$$

den Median zu F (häufig wird auch jeder Wert x mit $F(x) \ge \frac{1}{2}$, $1 - F(x-) \ge \frac{1}{2}$, also jeder Wert au dem Intervall $[\sup\{x \mid F(x) < \frac{1}{2}\}, \inf\{x \mid F(x) > \frac{1}{2}]$ als Median bezeichnet), so führt das Einsetzprinzip auf den *Stichprobenmedian:*

$$\operatorname{med}(X_1,\dots,X_n) := \begin{cases} X_{(\frac{n+1}{2}:n)}, & n \text{ ungerade}, \\ \frac{1}{2}(X_{(\frac{n}{2}:n)} + X_{(\frac{n}{2}+1:n)}, & n \text{ gerade} \end{cases}$$

als Schätzer für $\operatorname{med}(F)$.

12. Januar 1995

Glivenko-Cantelli: $\hat{F_n} \to F$ glm. f.s., Einsetzprinzip, führt auf $\overline{X_n}$ als Schätzer für $\mathrm{E}X_1$.

In diesem Fall folgt die Konsistenz direkt aus dem Gesetz der großen Zahlen.

Unter bestimmten Voraussetzungen an F ist der Stichprobenmedian ein konsistenter Schätzer für $\operatorname{med}(F)$. Idee: Glivenko-Cantelli, plus „Stetigkeit des Median-Funktionals“.

Bild 1

Satz 20.4
Es sei $X_1, X_2, \dots$ eine i.i.d.-Folge von Zufallsvariablen mit Verteilungsfunktion F. Für F gelte

$$F(\operatorname{med}(F)+\varepsilon) > \tfrac{1}{2}, \quad F(\operatorname{med}(F)-\varepsilon) < \tfrac{1}{2} \quad \forall \varepsilon > 0.$$

Dann folgt:

$$\operatorname{med}(X_1,\dots,X_n) \to_{n\to\infty} \operatorname{med}(F) \text{ fast sicher.}$$

Beweis
Sei ω aus dem Komplement der Ausnahmenullmenge im Satz von Glivenko-Cantelli fest; $\hat{F_n}$ bezieht sich im folgenden auf dieses ω.

Sei $\varepsilon > 0$. Zu ε existiert ein $\delta > 0$ mit

$$F(m+\varepsilon) > \tfrac{1}{2} + \delta, \quad F(m-\varepsilon) < \tfrac{1}{2} - \delta, \quad m := \operatorname{med}(F). \tag{$*$}$$

Nach Glivenko-Cantelli existiert zu δ ein n_0 mit

$$\sup_{x\in\mathbb{R}} \left| \hat{F_n}(x) - F(x) \right| < \frac{\delta}{2} \; \forall n \geq n_0.$$

Sei $Z_n := \operatorname{med}(X_1,\dots,X_n)$. Dann:

$$F(Z_n) > \hat{F_n}(Z_n) - \frac{\delta}{2} \geq \frac{1}{2} - \delta,$$

also wegen $(*)$ und der Monotonie von F:

$$Z_n > m - \varepsilon.$$

Für alle $x < Z_n$ gilt $\hat{F_n}(x) < \frac{1}{2}$, also

$$F(x) < \hat{F_n}(x) + \delta < \tfrac{1}{2} + \delta,$$

also, wieder wegen $(*)$ und der Monotonie von F,

$$x < m + \varepsilon.$$

Also: $Z_n \leq m + \varepsilon$.

Insgesamt haben wir (für das vorgegebene ω) gezeigt:

$$\forall \varepsilon > 0 \; \exists n_o \in \mathbb{N} \; \forall n \geq n_0 : \; |Z_n - m| < \varepsilon,$$

also $Z_n \to m$. Die Menge der zugelassenen ω hat Wahrscheinlichkeit 1, also: $Z_n \stackrel{\text{f.s.}}{\to} m$. *q.e.d.*

Gesetze der großen Zahlen sind nützlich für Konsistenzaussagen. Ähnlich kann der Zentrale Grenzwertsatz für (asymptotische) Konfidenzintervalle verwendet werden (siehe auch Aufgabe 49).

Beispiel 20.5
Hat X_1 nicht nur endlichen Erwartungswert μ, sondern sogar endliche Varianz $\sigma^2 > 0$, so gilt

$$\sqrt{n}(\overline{X_n} - \mu) \stackrel{\mathcal{D}}{\to} Z, \; Z \sim N(0, \sigma^2) \text{ (Satz 18.2)}$$

Sei q_α das α-Quantil zu $N(0,1)$. Ist σ^2 bekannt, so erhält man aus

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(\sigma q_{\frac{\alpha}{2}} \leq \sqrt{n}(\overline{X_n} - \mu) \leq \sigma q_{1-\frac{\alpha}{2}}) &= \mathbb{P}(q_{\frac{\alpha}{2}} \leq \frac{Z}{\sigma} \leq q_{1-\frac{\alpha}{2}}) \\ &= 1 - \alpha \end{aligned}$$

und

$$\sigma q_{\frac{\alpha}{2}} \leq \sqrt{n}(\overline{X_n} - \mu) \leq \sigma q_{1-\frac{\alpha}{2}} \quad \Longleftrightarrow \quad \overline{X_n} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X_n} - \frac{\sigma}{\sqrt{n}} q_{\frac{\alpha}{2}}$$

folgt, daß

$$\left[\overline{X_n} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}, \overline{X_n} - \frac{\sigma}{\sqrt{n}} q_{\frac{\alpha}{2}} \right]$$

ein asymptotisches $100(1-\alpha)\%$ Konfidenzintervall für μ ist.

Im allgemeinen wird σ^2 nicht bekannt sein. Man kann σ^2 jedoch konsistent durch die Stichprobenvarianz $S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X_n})^2$ schätzen (Aufg. 51), und es folgt mit Aufg ??? 0 und 43:

$$\frac{\sqrt{n}(\overline{X_n} - \mu)}{S_n} \stackrel{\mathcal{D}}{\to} Z, \; Z \sim N(0,1).$$

Dies bedeutet

$$\lim_{n\to\infty} \mathbb{P}(q_{\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X_n} - \mu}{S_n} \leq q_{1-\frac{\alpha}{2}}) = 1 - \alpha,$$

also ist

$$\left[\overline{X_n} - \frac{1}{\sqrt{n}} q_{1-\frac{\alpha}{2}} S_n, \overline{X_n} - \frac{1}{\sqrt{n}} q_\alpha 2 S_n \right]$$

ein im Grenzwert korrektes $100(1-\alpha)\%$ Konfidenzintervall für μ. Aus der Symmetrie von $N(0,1)$ folgt übrigens $q_{\frac{\alpha}{2}} = -q_{1-\frac{\alpha}{2}}$.

Das im 2. Teil verwendete Verfahren, in dem unbekannte Parameter der Grenzverteilung durch konsistende Schätzer ersetzt werden, wird gelegentlich „Studentisierung“ genannt. Insbesondere läßt sich hiermit das in Abschnitt 8.4.1 verwendete Verfahren verbessern.

Die „Bootstrap"-Methode

Im Beispiel 20.5 wurde die unbekannte Verteilung des „pivots" $\overline{X_n} - \mu$ durch $N(0, S_n^2)$ geschätzt. Kann man diese Verteilung direkt, d.h. ohne Anrufung des Zentralen Grenzwertsatzes schätzen?

Es sei R_n die Verteilungsfunktion zur Verteilung von $\overline{X_n} - \mu$. R_n hängt nur von F (und n) ab: R_n ist die Verteilungsfunktion der Differenz des Mittelwertes einer Stichprobe vom Umfang n aus F und dem Erwartungswert zu F. Das Einsetzprinzip legt es nahe, diese Größe durch die entsprechende Größe mit $\hat{F}_n$ anstelle von F zu schätzen, also durch die Größe

$$\text{Mittelwert einer Stichprobe vom Umfang } n \text{ aus } \hat{F}_n\text{-Erwartungswert von } \hat{F}_n \ (= \overline{X_n}),$$

also durch die Verteilung von $\overline{X_n}^* - \overline{X_n}$, $\overline{X_n}^* = \frac{1}{n}\sum_{i=1}^n X_i^*$, $X_1^*, \ldots, X_n^*$ unabhängig mit Verteilungsfunktion $\hat{F}_n$.

Beispiel 20.6
Angenommen, wir haben die Daten 0, -3, 5, 6 und wollen eine 90%-Konfidenzoberschranke für den Erwartungswert konstruieren.

Bild 2

Da $\hat{F}_4$ in den vier Datenwerten eine Sprung von jeweils der Höhe $\frac{1}{4}$ hat, gibt es für die „resamples" $X_1^*, \ldots, X_4^*$ insgesamt $4^4 = 256$ Möglichkeiten, die alle gleich wahrscheinlich sind (resample: „künstliche" Stichprobe aus der Originalstichprobe, ziehen mit zurücklegen).

Jeder liefert einen Wert (nicht notwendigerweise verschiedene) für $\overline{X_4}^* - X_4$:

resample				Mittelwert	$\overline{X_4}^* - \overline{X_4}$
0	0	0	0	0	-2
0	0	0	-3	-0.75	-2.75
...				...	...
5	-3	6	0	2	0
...				...	...
6	6	6	6	6	4

Die empirische Verteilungsfunktion der 256 Werte in der rechten Spalte ist unser Schätzer für die unbekannte Verteilungsfunktion von $\overline{X_n} - \mu$. Sie springt in $x = -0.25$ von $0.089\ldots$ auf $0.1367\ldots$ (modulo Rechenfehler), hat also das 0.1-Quantil -0.25. Ist $q_{0.1}$ das tatsächliche 0.1-Quantil zur Verteilungsfunktion von $\overline{X_n} - \mu$, so gilt $P(\overline{X_n} - \mu \geq q_{0.1}) \geq 0.9$, also wäre $\overline{X_n} - q_{0.1}$ eine 90%-Konfidenzoberschranke für μ.

Setzen wir unseren Schätzwert ein, so erhalten wir $\overline{X_n} - (-0.25) = 2.25$.

Bemerkung 20.7
Auflistung aller n^n resamples ist bei den üblichen n-Werten unmöglich. Man verwendet stattdessen die folgende „Monte-Carlo-Approximation" (Simulationsverfahren):

N resamples werden zufällig ausgewählt (N groß, häufig $N = 1000$), d.h. es wird N-mal eine Stichprobe $X_1^*, \ldots, X_n^*$ vom Umfang n mit Zurücklegen aus den Originaldaten $X_1, .., X_n$ gezogen und jedes Mal $\overline{X_n}^* - \overline{X_n}$ ausgerechnet. Die empirische Verteilungsfunktion zu diesen N Werten wird als Approximation für den tatsächlichen Einsetzschätzer (der alle n^n resamples verwendet) benutzt.

Das starke Gesetz der großen Zahlen stellt sicher, daß diese Approximation bei großem N gut ist.

17. Januar 1995

Bootstrap: Schätzen der Verteilungs eines „pivots" mit dem Einsetzprinzip, numerische Approximation des Schätzers mit der Monte-Carlo-Methode. Funktioniert das? Ja, bei $\overline{X_n} - \mu$, aber nicht immer.

Lemma 20.8 (Teilfolgencharakterisierung der Konvergenz in Wahrscheinlichkeit)
Es seien $X, X_1, X_2, \ldots$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$. Dann sind äquivalent:

1. $X_n \to X$ in Wahrscheinlichkeit ($X_n \xrightarrow{\text{P}} X$).
2. Zu jeder Teilfolge $(X_{n_k})_{k\in\mathbb{N}}$ existiert eine Teilteilfolge $(X_{n_{k_j}})$ mit $X_{n_k} \to X$ P-f.s. mit $j \to \infty$.

Beweis

„$\Rightarrow$“: Sei $(X_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge von $(X_n)_{n\in\mathbb{N}}$. Klar: $X_{n_k} \underset{k\to\infty}{\xrightarrow{\text{P}}} X$. Daher existiert zu jedem $j \in \mathbb{N}$ ein $k_j \in \mathbb{N}$ mit

$$P\left(\left|X_{n_{k_j}} - X\right| > \tfrac{1}{j^2}\right) < \tfrac{1}{j^2}.$$

Sei

$$A_j := \left\{\left|X_{n_{k_j}} - X\right| > \tfrac{1}{j^2}\right\}, \quad A := \bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} A_i,$$

mit Borel-Cantelli folgt $P(A) = 0$. Für alle $\omega \in A^c$ $(= \bigcup_{j=1}^{\infty} \bigcap_{i=j}^{\infty} A_i^C)$ gilt:

$$\exists j_0 \, \forall j \geq j_0 : \left|X_{n_{k_j}}(\omega) - X(\omega)\right| \leq \frac{1}{j^2}$$

d.h.: $X_{n_{k_j}}(\omega) \to X(\omega)$ mit $j \to \infty$.

„$\Leftarrow$“: Angenommen $X_n \xrightarrow{\text{P}} X$ ist falsch. Dann existiert ein $\varepsilon > 0$ mit $P(|X_n - X| > \varepsilon) \geq \varepsilon$ für ∞ viele n, also eine Teilfolge $(X_{n_k})_{k\in\mathbb{N}}$ mit $P(|X_{n_k} - X| > \varepsilon) \geq \varepsilon$ für alle $k \in \mathbb{N}$. Jede fast sicher konvergente Teilteilfolge $(X_{n_{k_j}})_{j\in\mathbb{N}}$ hiervon würde auch in Wahrscheinlichkeit konvergieren (Satz 15.2), d.h. man hätte

$$P\left(\left|X_{n_{k_j}} - X\right| > \varepsilon\right) @ >> n \to \infty > 0.$$

\# *q.e.d.*

Satz 20.9
Es sei $X_1, X_2, \ldots$ eine i.i.d.-Folge von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Verteilungsfunktionen F und endlichem zweiten Moment; μ sei $\mathrm{E}X_1$, $\sigma^2 = \operatorname{var} X_1$, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $\hat{F}_n$ die empirische Verteilungsfunktion zu $X_1, \ldots, X_n$.

Weiter sei $\hat{R}_n$ die (zufällige) Verteilungsfunktion von $\overline{X}_n^* - \overline{X}_n$, $\overline{X}_n^* = \frac{1}{n}\sum_{i=1}^n X_{ni}^*$, wobei $X_{n1}^*, \ldots, X_{nn}^*$ unabhängig und identisch verteilt sind mit Verteilungsfuntion $\hat{F}_n$ und R_n sei die Verteilungsfunktion von $\overline{X}_n - \mu$.

Dann gilt

$$\sup_{x\in\mathbb{R}} \left|\hat{R}_n(x) - R_n(x)\right| \to 0 \text{ in Wahrscheinlichkeit.}$$

Beweis

Sei $S_n := \sqrt{n}(\overline{X}_n - \mu)$. Für jedes $n \in \mathbb{N}$, $\omega \in \Omega$ seien $X^*_{n1}, \ldots, X^*_{nn}$ (genauer: $X^{*,\omega}_{ni}$) unabhängig und identisch verteilte Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega^*, \mathfrak{A}^*, \mathbb{P}^*)$ mit Verteilungsfunktion $\hat{F}_n$ $(= \hat{F}_n(\cdot, \omega))$; $S^*_n := \sqrt{n}(\overline{X}^*_n - \overline{X}_n)$. Dann gilt

$$R_n(x) = \mathbb{P}(S_n \leq \sqrt{n}x), \quad \hat{R}_n(x) = \mathbb{P}^*(S^*_n \leq \sqrt{n}x)$$

(vollständiger:

$$\hat{R}_n(x, \omega) = \mathbb{P}(\{\omega^* \in \Omega^* : \; S^{*,\omega}_n(\omega^*) \leq \sqrt{n}x\}) \quad).$$

Der ZGWS impliziert $S_n \xrightarrow{\mathcal{D}} N(0, \sigma^2)$, d.h.

$$\sup_{x \in \mathbb{R}} \left| R_n\left(\frac{x}{\sqrt{n}}\right) - \Phi\left(\frac{x}{\sigma}\right) \right| \to 0 \quad \text{mit } n \to \infty.$$

(„$\xrightarrow{\mathcal{D}}$" liefert nur punktweise Konvergenz, zur Gleichmäßigkeit siehe Beweis zu Glivenko-Cantellt); Φ ist die Verteilungsfunktion zu $N(0,1)$.

Ziel:

$$\sup_{x \in \mathbb{R}} \left| \hat{R}_n\left(\frac{x}{\sqrt{n}}\right) - \Phi\left(\frac{x}{\sigma}\right) \right| \to 0 \text{ in Wahrscheinlichkeit.} \qquad (*)$$

Mit Dreiecksungleichung und $\sup_{x \in \mathbb{R}} f(\frac{x}{\sqrt{n}}) = \sup_{x \in \mathbb{R}} f(x)$ folgt dann die Behauptung.

Es gilt für jedes $\omega \in \Omega$:

$$\begin{aligned} \mu_{nk} &:= \mathrm{E}_\omega X^*_{nk} = \overline{X}_n(\omega), \\ s^2_n &:= n\sigma^2_{nk} = \sum_{i=1}^n (X_i(\omega) - \overline{X}_n(\omega))^2, \end{aligned}$$

also erhält man

$$\begin{aligned} L_n(\varepsilon) &:= \frac{1}{s^2_n} \sum_{k=1}^n \int_{|X^*_{nk} - \mu_{nk}| > \varepsilon s_n} (X^*_{nk} - \mu_{nk})^2 \, d\mathbb{P}^* \\ &= \frac{n}{s^2_n} \int_{|X^*_{n1} - \overline{X}_n| > \varepsilon s_n} (X^*_{n1} - \overline{X}_n)^2 \, d\mathbb{P}^* \\ &= \frac{1}{s^2_n} \sum_{\substack{1 \leq i \leq n \\ |X_i - \overline{X}_n| > \varepsilon s_n}} (X_i - \overline{X}_n)^2, \end{aligned}$$

denn unter $(\mathbb{P}^*)^{X^*_{n1}}$ hat jeder der Werte $X_1, \ldots, X_n$ die Wahrscheinlichkeit $\frac{1}{n}$.

Klar: $\frac{s^2_n}{n} \to \sigma^2$ (Aufgabe 51). Wegen

$$|X_i - \overline{X}_n| > \varepsilon s_n \implies \left(\frac{1}{\sqrt{n}} s_n < \frac{\sigma}{2}\right) \vee \left(|X_i - \mu| > \frac{\varepsilon}{4}\sqrt{n}\sigma\right) \vee \left(|\overline{X_n} - \mu| > \frac{\varepsilon}{4}\sqrt{n}\sigma\right)$$

(sieht man am leichtesten über die Negierung der Aussage) gilt

$$\begin{aligned}\frac{1}{n}\sum_{1\le i\le n}(X_i-\overline{X}_n)^2 &\le 1_{(0,\frac{\sigma}{2})}\left(\frac{s_n}{\sqrt{n}}\right)\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\\ &+\frac{1}{n}\sum_{\substack{1\le i\le n\\ |X_i-\mu|}}>\frac{\varepsilon}{4}\sqrt{n}\,\sigma(X_i-\overline{X}_n)^2+1_{(\frac{\varepsilon}{4}\sqrt{n}\sigma,\infty)}(\overline{X}_n-\mu)\,\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2.\end{aligned}$$

Wegen $\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\overset{\text{f.s.}}{\to}\sigma^2$ (Aufgabe 51) und $\frac{s_n}{\sqrt{n}}\overset{\text{f.s.}}{\to}\sigma$, $\overline{X}_n-\mu\overset{\text{f.s.}}{\to}0$ konvergieren der erste und der dritte Term $\mathbb{P}$-f.s. gegen 0. Der zweite Term konvergiert wegen

$$\begin{aligned}\mathbb{P}(\exists i\in\{1,\dots,n\}\mid |X_i\mu|>\frac{\varepsilon}{4}\sqrt{n}\sigma) &\le n\mathbb{P}(|X_1-\mu|>\frac{\varepsilon}{4}\sqrt{n}\sigma)\\ &\le n\frac{1}{\frac{\varepsilon^2}{16}n\sigma^2}\int_{|X_1-\mu|>\frac{\varepsilon}{4}\sqrt{n}\sigma}(X_1-\mu)^2\,d\mathbb{P}\end{aligned}$$

und

$$\int(X_1-\mu)^2\,d\mathbb{P}<\infty\quad\text{(majorisierte Konvergenz!)}$$

gegen 0, denn

$$\mathbb{P}(|Z_n|\neq 0)\to 0\implies Z_n\overset{\text{P}}{\to}0.$$

Damit ist $L_n(\varepsilon)\overset{\text{P}}{\to}0\;\forall\varepsilon>0$ gezegt.

Wir wollen nun zeigen: Zu jeder Teilfolge (n_k) existiert eine Teilteilfolge (n_{k_j}) mit

$$L_{n_{k_j}}(\varepsilon)\overset{\text{f.s.}}{\to}0\;\forall\varepsilon>0\tag{$**$}$$

Entscheidend hierbei ist, daß (n_{k_j}) bei allen $\varepsilon>0$ simultan funktioniert: Die Ausnahmenullmenge hängt dann nämlich nicht von ε ab.

Beweis mit dem Diagonalelement:

Aus $L_{n_k}(1)\overset{\text{P}}{\to}0$ folgt mit Lemma 20.8 die Existenz einer Teilfolge $n_{11},n_{12},n_{13},\dots$ mit $L_{n_{1i}}(1)\underset{i\to\infty}{\overset{\text{f.s.}}{\to}}0$. Zu (n_{1i}) existiert wegen $L_{n_{1i}}(\frac{1}{2})\overset{\text{P}}{\to}0$ eine Teilfolge (n_{21}) mit $L_{n_{2i}}(\frac{1}{2})\overset{\text{f.s.}}{\to}0$ etc. Gilt $L_{n_{li}}(\frac{1}{l})@>\text{f.s.}>i\to\infty>0$, so können wir wegen $L_{n_{li}}(\frac{1}{l+1})\overset{\text{P}}{\to}0$ eine Teilfolge $(n_{l+1,i})$ von (n_{li}) finden mit $L_{n_{l+1,i}}(\frac{1}{l+1})\overset{\text{f.s.}}{\to}0$.

Für die Folge $(n_{ll})_{l\in\mathbb{N}}$ (die „Diagonalfolge“) gilt nun $(**)$. Es gibt also ein $A\in\mathfrak{A}$ mit $\mathbb{P}(A)=1$ und

$$L_{n_{k_j}}(\varepsilon,\omega)@>>j\to\infty>0\;\forall\omega\in A,\;\varepsilon>0.$$

Für jedes solche ω liefert nun der Satz von Lindeberg

$$\sqrt{n_{k_j}}\,S^{*,\omega}_{n_{k_j}}\overset{\mathcal{D}}{\to}N(0,\sigma^2),$$

also (siehe oben)

$$\sup_{x\in\mathbb{R}} \left| \hat{R}_{n_{k_j}} \left(\frac{x}{\sqrt{n}}, \omega \right) - \Phi\left(\frac{x}{\sigma}\right) \right| \to 0 \quad \forall \omega \in A.$$

Insgesamt: Zu jeder Teilfolge (n_k) existiert (n_{k_j}) mit

$$\sup_{x\in\mathbb{R}} \left| \hat{R}_{n_{k_j}} \left(\frac{x}{\sqrt{n}} \right) - \Phi\left(\frac{x}{\sigma}\right) \right| \to 0 \text{ } \mathbb{P}\text{-f.s..}$$

Also existiert (geht) ??? nach Lemma 20.8 die Zufallsgröße

$$\sup_{x\in\mathbb{R}} \left| \hat{R}_{n} \left(\frac{x}{\sqrt{n}} \right) - \Phi\left(\frac{x}{\sigma}\right) \right|$$

mit $n \to \infty$ in Wahrscheinlichkeit gegen 0. Dies ist $(*)$. *q.e.d.*

Bemerkung 20.10
Der Beweis zeigt, daß $\hat{R}_n(\frac{\cdot}{\sqrt{n}})$ den gleichen Limes hat wie $R_n(\frac{\cdot}{\sqrt{n}})$, nämlich die Verteilungsfunktion zu $N(0, \sigma^2)$.

Bootstrap-Konfidenzintervalle verwenden die Quantile $\hat{q}_{n\alpha} := \hat{T}_n^{-1}(\alpha)$ von $\hat{R}_n$ anstelle der (unbekannten) Quantile von R_n. Ähnlich wie im Beweis zu Satz 20.4 folgt

$$\sqrt{n}\, \hat{q}_{n\alpha} \to \sigma \Phi^{-1}(\alpha)$$

und damit der asymptotisch korrekte Wert für die Überdeckungswahrscheinlichkeit von Bootstrap-Konfidenzintervallen (für μ, unter den Bedingungen des Satzes).

19. Januar 1995

21 Parametrische Modelle: Maximum-Likelihood-Schätzer

Es sei $\mathfrak{P} = \{P_\theta \mid \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$, eine parametrische Familie von Wahrscheinlichkeitsmaßen auf $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$. Wir setzen voraus, daß alle P_θ, $\theta \in \Theta$ eine Dichte $f(\cdot|\theta)$ bezüglich eines festen, σ-endlichen Maßes μ auf $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$ haben. In Abschnitt 7 wurde dies bei abzählbarem $\mathfrak{X}$, $\mathfrak{B}(\mathfrak{X}) = \wp(\mathfrak{X})$ und abzählendem Maß μ betrachtet.

Der zweite wichtige Spezialfall ist $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X})) = (\mathbb{R}^d, \mathfrak{B}^d)$, $\mu = \ell^d$.

In Verallgemeinerung von Definition 7.1 nennen wir

$$\ell(\cdot|x) : \Theta \to \mathbb{R}, \quad \ell(\theta|x) := f(x|\theta)$$

die *Likelihood-Funktion* (bei Vorliegen der Beobachtung x), den Logarithmus die *Log-Likelihood-Funktion* und eine meßbare Abbildung $\hat{\theta} : \mathfrak{X} \to \Theta$ mit

$$\ell(\hat{\theta}(x)|x) = \sup\{\ell(\theta|x) \mid \theta \in \Theta\}$$

für μ-fast-alle $x \in \mathfrak{X}$ einen *Maximum-Likelihood-Schätzer (MLS)* fü θ. Ist $\hat{\theta}$ ein MLS für θ und $\eta = g(\theta)$, so heißt $\hat{\eta} := g(\hat{\theta})$ MLS für η.

Beispiel 21.1

(i) Es sei $X_1, \ldots, X_n$ eine Stichprobe aus der Exponentialverteilung mit Parameter θ, $\theta > 0$ unbekannt. Mit $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X})) = (\mathbb{R}^n, \mathfrak{B}^n)$, $\mu = \ell^{\otimes n}$ (alternativ: $\mathfrak{X} = \mathbb{R}_+^n$ etc.)

$$f(x_1, \ldots, x_n|\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right) 1_{[0,\infty)}(x_1, \ldots, x_n)$$

ergibt sich

$$\log \ell(\theta|x_1, \ldots, x_n) = n \log \theta - \theta \sum_{i=1}^n x_i,$$

also ist

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

ein MLS für θ. Zur Asymptotik siehe Aufgabe 49.

(ii) Sind $x_1, \ldots, x_n$ Werte einer Stichprobe aus $U(0, \theta)$ (Gleichverteilung auf $[0, \theta]$), $\theta > 0$ unbekannt, so erhält man, wieder mit $\mathfrak{X} = \mathbb{R}^n$, $\mu = \ell^{\otimes n}$,

$$\ell(\theta|x_1, \ldots, x_n) = \begin{cases} \theta^{-n}, & \max\{x_1, \ldots, x_1\} \leq \theta \\ 0, & \text{sonst} \end{cases},$$

also den MLS $\hat{\theta}(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\}$. Zur Asymptotik siehe Aufgabe 54.

(iii) Als mehrdimensionales Beispiel betrachten wir eine Stichprobe aus $N(\mu, \sigma^2)$ mit unbekannten $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Als Log-Likelihood-Funktion ergibt sich mit $\theta = (\mu, \sigma^2) \in \Theta := \mathbb{R} \times (0, \infty)$:

$$\log \ell(\theta|x_1, \ldots, x_n) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Für jedes feste σ^2 wird dies als Funktion von μ durch $\hat{\mu} = \overline{X}_n$ maximiert; die Funktion

$$\sigma^2 \to -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \overline{x}_n)^2$$

wird maximiert in

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x}_n)^2.$$

In vielen Fällen erhält man einen MLS über „Nullsetzen der Ableitung der Log-Likelihood-Funktion“, also als Lösung der Likelihood-Gleichungen (score equations)

$$\frac{\partial}{\partial \theta} \log f(x|\theta) = 0.$$

unter gewissen Regularitätsbedingungen läßt sich in solchen Fällen das asymptotische Verhalten von $\hat{\theta}_n$ mit $n \to \infty$ bestimmen.

Zunächst: informelles Argument.

Ist $X_1, \ldots, X_n$ eine Stichprobe aus einer Verteilung mit μ-Dichte $f(\cdot|\theta)$, $\theta \in \Theta \subset \mathbb{R}$, so gilt

$$s_n(\theta) := \frac{\partial}{\partial\theta} \log \ell(\theta|X_1, \ldots, X_n) = \sum_{i=1}^{n} \underbrace{\frac{\partial}{\partial\theta} \log f(X_i|\theta)}_{Y_i}.$$

Der Mittelwertsatz (sofern anwendbar) liefert

$$s_n(\theta) = s_n(\hat\theta_n) + (\hat\theta_n - \theta) s_n'(\tilde\theta_n),$$

mit $\tilde\theta$ zwischen θ und $\hat\theta_n$. Mit $s_n(\hat\theta_n) = 0$ folgt

$$\sqrt{n}\,(\hat\theta_n - \theta) = \frac{-\frac{1}{\sqrt{n}} s_n(\theta)}{\frac{1}{n} s_n'(\tilde t_n)}. \qquad (*)$$

$$\begin{aligned}
E_\theta Y_i &= \int \left(\frac{\partial}{\partial\theta} f(x|\theta)\right) \cdot \frac{1}{f(x|\theta)} \cdot f(x|\theta)\mu\,(dx) \\
&\overset{\text{heuristisch!!!}}{=} \frac{\partial}{\partial\theta} \int f(x|\theta)\mu\,(dx) \\
&= \frac{\partial}{\partial\theta} 1 \\
&= 0,
\end{aligned}$$

ZWGS:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \xrightarrow{\mathcal{D}} Z, \quad Z \sim N(0, I(\theta))$$

mit

$$I(\theta) := \operatorname{var}_\theta Y_1 = \mathrm{E}_\theta Y_1^2 = \mathrm{E}_\theta(\tfrac{\partial}{\partial\theta} \log f(X_1|\theta))^2.$$

Ist $\hat\theta_n$ konsistent, so muß auch $\hat\theta_n \xrightarrow{\mathrm{P}} \theta$ gelten, „also“ (wieder heuristisch)

$$\tfrac{1}{n} s_n'(\tilde\theta_n) = \tfrac{1}{n} \textstyle\sum_{i=1}^{n} \tfrac{\partial^2}{\partial\theta^2} \cdot \log f(X_i|\theta) \xrightarrow{\mathrm{P}} \mathrm{E}_t heta \tfrac{\partial^2}{\partial\theta^2} \log f(X_1|\theta).$$

Vertauscht man wieder Integral und Ableitung, so folgt

$$\begin{aligned}
\mathrm{E}_\theta \frac{\partial^2}{\partial\theta^2} \log f(X_1|\theta) &= \int \frac{\partial}{\partial\theta} \left(\frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)} \right) f(x|\theta)\mu\,(dx) \\
&= \int \left(\frac{\frac{\partial^2}{\partial\theta^2} f(x|\theta)}{f(x|\theta)} - \frac{(\frac{\partial}{\partial\theta} f(x|\theta))^2}{f(x|\theta)^2} \right) f(x|\theta)\mu\,(dx) \\
&= -\int \left(\frac{\partial}{\partial\theta} \log f(x|\theta) \right)^2 f(x\theta)\mu\,(dx) \\
&= -\mathrm{E}(\frac{\partial}{\partial\theta} \log f(X_1|\theta))^2 \\
&= -I(\theta),
\end{aligned}$$

also folgt insgesamt aus (∗):

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} Z, \quad Z \sim N(0, \tfrac{1}{I(\theta)}).$$

Ziel: Präzisieren und auf höhere Dimensionen verallgemeinern.

Lemma 21.2
Es seien $Z_1, Z_2, \dots$ stochastische $k \times k$-Matrizen, $A \in \mathbb{R}^{n \times n}$ mit $\det A \neq 0$, $X, Y_1, Y_2, \dots$ k-dimensionale stochastische Vektoren. Dann gilt:

$$Z_n \xrightarrow{\mathrm{P}} A,\ Z_n Y_n \xrightarrow{\mathcal{D}} X \implies Y_n \xrightarrow{\mathcal{D}} A^{-1}X.$$

Beweis

(grob)

Wir verwenden eine mehrdimensionale Variante des Darstellungssatzes von Skorohod (Satz 17.4). Es reicht dann zu zeigen:

$$Z_n \xrightarrow{\text{f.s.}} A,\ Z_n Y_n \xrightarrow{\text{f.s.}} X \implies Y_n \xrightarrow{\text{f.s.}} A^{-1}X.$$

Angenommen, man hat $Z_n(\omega) \to A$, $Z_n(\omega)Y_n(\omega) \to X(\omega)$. Wegen $\det Z_n(\omega) \to \det A$ ist $Z - n(\omega)$ ab einem n_0 invertierbar und es folgt

$$Y_n(\omega) = Z_n^{-1}(\omega) \cdot (Z_n(\omega)Y_n(\omega)) \to A^{-1}X(\omega).$$

q.e.d.

Der mehrdimensionale Fall wird mit der im Beweis zu Satz 19.1 verwendeten Technik (Cramér-Wold device) auf den eindimensionalen Fall zurückgeführt:

Sind $X, X_1, X_2, \dots$ k-dimensionale Zufallsverktoren, so gilt

$$X_n \xrightarrow{\mathcal{D}} X \quad \iff \quad \forall a \in \mathbb{R}^k : a'X_n \xrightarrow{\mathcal{D}} a'X.$$

Satz 21.3
Es sei $\Theta \subset \mathbb{R}^k$ offen. Die Familie $\{f(\cdot|\theta) \mid \theta \in \Theta\}$ genüge den folgenden Bedingungen, wobei wir D_i für $\frac{\partial}{\partial \theta_i}$ und D_{ij} für $\frac{\partial}{\partial \theta_i \partial \theta_j}$ schreiben:

(i) $(D_{ij} \log f)(x|\theta)$ existiert für alle $x \in \mathfrak{X}$, $\theta \in \Theta$, $1 \leq i < j \leq k$.

(ii) Zu jedem θ aus Θ existieren Funktionen $h_\theta : \Theta \to \mathbb{R}$, $M_\theta : \mathfrak{X} \to \mathbb{R}$ mit $\lim_{\theta' \to \theta} h_\theta(\theta') = 0$, $\int M_\theta(x) P_\theta\,(dx) < \infty$ und

$$|(D_{ij} \log f)(x|\theta') - (D_{ij} \log f)(x|\theta)| \leq h_\theta(\theta')M_\theta(x)$$

für $1 \leq i, j \leq k$, $x \in \mathfrak{X}$, $\theta' \in \Theta$.

(iii) $\mathrm{E}_\theta(D_i \log f)(X|\theta) = \theta\ \forall \theta \in \Theta,\ 1 \leq i \leq k$.

(iv) Zu $(D_i \log f)(X|\theta)(D_j \log f)(X|\theta)$ existiert der Erwartungswert, und die *Informationsmatrix*, auch *Fisher-Information* genannt, $I(\theta)$ mit

$$(I(\theta))_{ij} := \mathrm{E}(D_i \log f)(X|\theta)(D_j \log f)(X|\theta) \text{ ist positiv definit.}$$

(v) $(I(\theta))_{ij} = -\mathrm{E}_\theta(D_{ij}\log f)(X|\theta)$ für $1 \le i,j \le k$.

Seien nun $\theta \in \Theta$ und $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvektoren mit Verteilung P_θ. Sei weiter für jedes $n \in \mathbb{N}$, $\hat\theta_n$ eine Θ-wertige (meßbare) Zufallsgröße mit

(vi) $\sum_{i=1}^n (D_j \log f)(X_i|\hat\theta_n) = 0$, $1 \le j \le k$.
Gilt dann

(vii) $\hat\theta_n \xrightarrow{\mathrm{P}} \theta$,
so folgt

$$\sqrt{n}(\hat\theta_n - \theta) \xrightarrow{\mathcal{D}} Z, \quad Z \sim N_k(0, I(\theta)^{-1}).$$

24. Januar 1995

Beweis

Sei $a \in \mathbb{R}^k$, $\nabla g = (D_1 g, \ldots, D_k g)'$, $s_n(\theta,\omega) := \sum_{m=1}^n (\nabla \log f)(X_m(\omega)|\theta)$.

Wenden wir den Mittelwertsatz auf $\theta \to a's_n(\theta,\omega)$ an, so erhalten wir mit (i) für ein $\tilde\theta_n$ auf der Verbindungslinie von θ und $\hat\theta_n$:

$$0 = s_n(\hat\theta_n(\omega),\omega) = a's_n(\theta,\omega) + ((\nabla a's_n)(\tilde\theta_n(\omega),\omega))'(\hat\theta_n(\omega) - \theta) \quad \text{(Taylor erster Ordnung)}$$

also

$$(\tfrac{1}{n}(\nabla a's_n)(\tilde\theta_n(\omega),\omega))'(\sqrt{n}(\hat\theta_n(\omega) - \theta)) = -\tfrac{1}{\sqrt{n}} a's_n(\theta,\omega) \tag{1}$$

Setzt man $Y_m := a'(\nabla \log f)(X_m|\theta)$, so wird die rechte Seite zu

$$-\frac{1}{\sqrt{n}}\, a's_n(\theta,\omega) = -\frac{1}{\sqrt{n}} \sum_{m=1}^n Y_m(\omega).$$

Diese Zufallsvariablen sind i.i.d. mit Erwartungswert 0 (wegen (iii)) und Varianz $a'I(\theta)a$ (wegen (iv)), der Zentrale Grenzwertsatz liefert für die rechte Seite von (1) den Verteilungslimes $N(0, a'I(\theta)a)$. Definiere nun zufällige Matrizen $Z_m(\omega|\theta)$ durch

$$(Z_m(\omega|\theta))_{ij} := (D_{ij}\log f)(X_m(\omega)|\theta)$$

Beachtet man (1) und

$$\frac{1}{n}\left(\nabla a's_n\right)(\tilde\theta_n(\omega),\omega) = a'\left(\frac{1}{n}\sum_{m=1}^n Z_m(\omega|\tilde\theta_n(\omega))\right),$$

so kann man das bisher Erreichte wie folgt zusammenfassen:

$$a'\left(\frac{1}{n}\sum_{m=1}^n Z_m(\cdot|\tilde\theta_n)\right)\sqrt{n}(\hat\theta_n - \theta) \xrightarrow{\mathcal{D}} N(0, a'I(\theta)a) \quad \forall a \in \mathbb{R}^d.$$

Mit Cramér-Wold folgt hieraus

$$\left(\frac{1}{n}\sum_{m=1}^{n} Z_m(\cdot|\tilde{\theta}_n)\right)\sqrt{n}(\hat{\theta}_n-\theta) \xrightarrow{\mathcal{D}} N_k(0,I(\theta)).$$

Hieraus folgt die Aussage des Satzes mit Lemma 2.2, wenn noch

$$\frac{1}{n}\sum_{m=1}^{n} Z_m(\cdot|\tilde{\theta}_n) \xrightarrow{\mathrm{P}} -I(\theta) \tag{2}$$

gezeigt werden kann. Aus (ii) folgt

$$\begin{aligned} & \left|\frac{1}{n}\sum_{m=1}^{n} Z_m(\omega,\tilde{\theta}_n(\omega)) - \frac{1}{n}\sum_{m=1}^{n} Z_m(\omega|\theta))_{ij}\right| \\ \leq\ & \frac{1}{n}\sum_{m=1}^{n}\left|(D_{ij}\log f)(X_m(\omega)|\tilde{t}_n(\omega)) - (D_{ij}\log f)(X_m(\omega)|\theta)\right| \\ \leq\ & h_\theta(\tilde{\theta}_n(\omega))\cdot\frac{1}{n}\sum_{m=1}^{n} M_\theta(X_m(\omega)). \end{aligned}$$

Wir haben $\tilde{t}_n \xrightarrow{\mathrm{P}} \theta$ wegen (vii) und der Lage von $\tilde{\theta}_n$ „zwischen" θ und $\hat{\theta}_n$. Mit $h_\theta(\theta') \to 0$ mit $\theta' \to \theta$ folgt also, daß der erste Faktor in Wahrscheinlichkeit gegen 0 geht; nach dem schwachen Gesetz der großen Zahlen und wegen $\mathrm{E}_\theta M_\theta(X) < \infty$ konvergiert der zweite Faktor in Wahrscheinlichkeit gegen eine (endliche) Konstante, d.h. die Obergrenze geht in Wahrscheinlichkeit gegen 0. Man kann in (2) also $\tilde{\theta}_n$ durch θ ersetzen.

Schließlich sind wegen (v) die zufälligen Matrizen $(Z_m(\cdot|\theta))_{m\in\mathbb{N}}$ i.i.d. mit Erwartungswert $-I(\theta)$, d.h.

$$\frac{1}{n}\sum_{m=1}^{n} Z_m(\cdot|\theta) \xrightarrow{\mathrm{P}} -I(\theta),$$

nach dem (komponentenweise angewendeten) schwachen Gesetz. *q.e.d.*

Zwei hauptsächliche Anwendungen:

asymptotische Qualität von MLS

Unter bestimmten Voraussetzungen und bei Einschränkung auf „reguläre" Schätzer ist der Wert $I(\theta)^{-1}$ für die asymptotische Varianz minimal („MLS sind effizient").

Beachte: $I(\theta)^{-1}$ erscheint auch als Untergrenze in der Informationsungleichung (Satz 7.8) für die Varianz eines erwartungstreuen Schätzers.

Konstruktion von (asymptotischen) Konfidenzbereichen

Hierzu benötigen wir zunächst ein Hilfsmittel (auch bei Abschnitt 8.5 interessant):

Satz & Definition 21.4
Sind $Z_1,\ldots,Z_k$ unabhängig und $N(0,1)$-verteilt, so hat $Z_1^2+\cdots+Z_k^2$ Lebesgue-Dichte

$$f_k(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x > 0.$$

Man nennt diese Verteilung die *Chiquadratverteilung* mit k Freiheitsgraden, kurz:

$$Z_1^2 + \cdots + Z_k^2 \sim \chi_k^2.$$

Beweis

Mit den in Aufgabe 21 eingeführten Gammaverteilungen gilt $\chi_k^2 = \Gamma(\frac{k}{2}, \frac{1}{2})$. Da Gammaverteilungen die Eigenschaft $\Gamma(\alpha,\gamma) * \Gamma(\beta,\gamma) = \Gamma(\alpha+\beta,\gamma)$ haben (Aufg. 41), und $Z_1^2, \ldots, Z_k^2$ unabhängig sind, reicht es, die Aussage für $k = 1$ zu beweisen:

$$P(Z_1^2 \leq z) = P(-\sqrt{z} \leq Z_1 \leq \sqrt{z}) = 2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}}\, dx.$$

Jetzt differenzieren nach z. *q.e.d.*

Angenommen, $\hat{I}_n$ ist ein konsistenter Schätzer für $I(\theta)$, $\hat{I}_n$ symmetrisch, so ist für hinreichend große n $\hat{I}_n$ positiv definit, d.h. $\hat{I}_n^{\frac{1}{2}}$ existiert. Satz 21.3 liefert

$$\hat{I}_n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} Z, \quad Z \sim N_k(0, I_k),$$

wobei I_k die $k \times k$-Einheitsmatrix bezeichnet. Da $x \to x^t x$ $(= \|x\|^2)$ stetig ist, folgt mit einer mehrdimensionalen Version des Continous Mapping Theorems (Satz 17.6)

$$(\hat{\theta}_n - \theta)^t \hat{I}_n (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \|Z\|^2, \quad Z \sim N_k(0, I_d).$$

Die Komponenten von Z sind unabhängig und $N(0,1)$-verteilt. Nach Satz 21.4 hat man also χ_k^2 als Grenzverteilung, d.h.

$$\lim_{n\to\infty} P_\theta\left((\hat{\theta} - \theta)^t \hat{I}_n(\hat{\theta}_n - \theta) \leq \chi_{k,1-\alpha}^2\right) = 1 - \alpha,$$

wobei $\chi_{k,\beta}^2$ das β-Quantil zu χ_k^2 bezeichnet (Tabelle). Setzt man

$$\mathfrak{E}(x, A, \lambda) := \{y \mid (y-x)^t A (y-x) \leq \lambda\},$$

so zeigt dies, daß das Ellipsoid $\mathfrak{E}(\hat{\theta}_n, \hat{I}_n, \chi_{k,1-\alpha}^2)$ ein (asymptotisches) $100(1-\alpha)$%-Konfidenzgebiet ist für den unbekannten Parameter θ. Ist die Abbildung $\theta \to I(\theta)$ stetig (und „zugänglich“), so kann man beispielsweise

$$\hat{I}_n := I(\hat{\theta}_n)$$

verwenden.

Bemerkung 21.5
Das folgende heuristische Argument kann gelegentlich zu einem Beweis der Konsistenz des MLS ausgebaut werden: Sei $h_n(\theta) := \frac{1}{n}\sum_{i=1}^n \log f(X_i|\theta)$ eine Folge stochastischer Funktionen auf Θ. Ist θ_0 der wahre Parameter, so liefert das starke Gesetz der großen Zahlen:

$$h_n(\theta) \overset{\text{f.s.}}{\to}_{n\to\infty} h(\theta) := \mathrm{E}_{\theta_0} \log f(X_1|\theta) \quad \forall\theta \in \Theta.$$

Ist diese Konvergenz „hinreichend gleichmäßig" in θ, so wird man erwarten, daß der MLS (der ja h_n maximiert) gegen den Punkt konvergiert, in dem h maximal wird. Es gilt

$$\begin{aligned}
h(\theta) - h(\theta_0) &= \mathrm{E}_{\theta_0} \log \frac{f(X_1|\theta)}{f(X_1|\theta_0)} \\
&\overset{\text{Jensensche Ungleichung, Satz 13.27}}{\leq} \\
&\overset{\text{bei injektiver Parametrisierung}}{<} \log \mathrm{E}_{\theta_0} \frac{f(X_1|\theta)}{f(X_1|\theta_0)} \\
&= \log \int \frac{f(x|\theta)}{f(x|\theta_0)} f(x|\theta_0)\mu\,(dx) \\
&= \log 1 \\
&= 0
\end{aligned}$$

Also wird h in θ_0 maximal, wie gewünscht.

Bild 1

Krümmung der Grenzfunktion im Maximum:

$$\begin{aligned}
\frac{\partial}{\partial\theta^2} h(\theta)|_{\theta=\theta_0} &\overset{\text{nach Vertauschung}}{=} \mathrm{E}_{\theta_0} \frac{\partial}{\partial\theta^2} \log f(X_1|\theta) \\
&= -I(\theta_0)
\end{aligned}$$

26. Januar 1995

22 Parametrische Modelle: Likelihood-Quotienten-Test

Wieder: $\mathfrak{P} = \{P_\theta \mid \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$, parametrische Familie von Wahrscheinlichkeitsmaßen auf $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$, μ ein festes σ-endliches Maß auf $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$, und $f(\cdot,\theta)$ eine μ-Dichte von P_θ.

Bei endlichem (oder abzählbar unendlichem) $\mathfrak{P}$ existiert stets ein solches dominierendes μ, im Falle $\mathfrak{P} = \{P_0, P_1\}$ kann beispielsweise $\mu = P_0 + P_1$ gewählt werden. Sei $\Theta_0 \subset \Theta$, es soll $H_0 : \theta \in \Theta_0$ (gegen die Alternative $H_1 : \theta \in \Theta \setminus \Theta_0$) getestet werden. Ein (randomisierter) Test wird (vgl. Definition 7.11) durch eine $(\mathfrak{B}(\mathfrak{X}), \mathfrak{B})$-meßbare Abbildung $\varphi : \mathfrak{X} \to [0,1]$, die Testfunktion, beschrieben:

Bei Vorliegen der Beobachtung $x \in \mathfrak{X}$ ist $\varphi(x)$ die Wahrscheinlichkeit, mit der H_0 abgelehnt wird. Bei $\Theta = \{\Theta_0, \Theta_1\}$, $\Theta_0 = \{\theta_0\}$ (einfache Hypothese, einfache Alternative) liefert die allgemeine Variante des Neyman-Pearson-Lemmas als besten Test:

$$\varphi(x) = \begin{cases} 1, & f(x|\theta_1) > c \cdot f(x|\theta_0) \\ \gamma, & f(x|\theta_1) = c \cdot f(x|\theta_0) \\ 0, & f(x|\theta_1) < c \cdot f(x|\theta_0) \end{cases}$$

wobei c und $0 < \gamma < 1$ so gewählt werden, daß $\mathrm{E}_{\theta_0}\varphi(X) = \alpha$ gilt. Der Test beruht auf der Testgröße $T(x) := \frac{f(x|\theta_1)}{f(x|\theta_0)}$, dem Likelihood-Quotienten, in dem Sinne, daß abgelehnt wird, wenn T zu groß wird (eigentlich eine *Familie* von Tests). In Verallgemeinerung hiervon bezeichnet man Tests mit der Testgröße

$$T(x) := \frac{\sup_{\theta\in\Theta} f(x|\theta)}{\sup_{\theta\in\Theta_0} f(x|\theta)}.$$

den *Likelihood-Quotienten-Test* für Θ_0 (Meßbarkeit wird unterstellt). Ist $\hat{\theta}$ MLS zu θ, so erscheint im Zähler $f(x|\hat{\theta})$.

Nützliche Überlegung: Jede Testgröße T^*, die eine streng monoton steigende Funktion von T ist, führt auf dieselben Tests (es geht letztlich um $\{x \mid T(x) > c\}$, $\{x \mid T(x) = c\}$).

Beispiel 22.1
Bei einer Stichprobe $X_1, \ldots, X_n$ aus einer Exponentialverteilung mit unbekanntem Parameter $\theta > 0$ soll $H_0 : \theta \leq \theta_0$ gegen $H_1 : \theta > \theta_0$ zum Niveau α getestet werden. Nach Beispiel 21.1 ist $\hat{\theta} = \frac{1}{\overline{X}_n}$ der MLS für θ. Als Funktion von θ ist

$$\begin{aligned} f(x_1, \ldots, x_n|\theta) &= \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right) \\ &= \theta^n \exp\left(-\frac{n\theta}{\hat{\theta}}\right) \end{aligned}$$

streng monoton steigend auf $[0, \hat{\theta}_n]$ und streng monoton fallend auf $[\hat{\theta}_n, \infty)$, d.h.

$$\sup_{\theta\in\Theta_0} f(x_1, \ldots, x_n|\theta) = \begin{cases} f(x|\hat{\theta}_n), & \text{wenn } \hat{\theta}_n \leq \theta_0, \\ f(x|\theta_0) & \text{sonst.} \end{cases}$$

Damit:

$$T(x) = \begin{cases} 1, & \hat{\theta}_n \leq \theta_0, \\ \left(\frac{\hat{\theta}_n}{\theta_0}\right)^n \exp\left(-n + \frac{n\theta_0}{\hat{\theta}_n}\right), & \hat{\theta}_n > \theta_0 \end{cases}.$$

(Hinweis: Die „n“s an $\hat{\theta}$ können auch weggelassen werden.)

Beachte: T hängt nur über $\hat{\theta}$ von den Daten X an, und ist eine monoton steigende Funktion von $\hat{\theta}$. Also: Der LQ-Test lehnt ab, wenn $\hat{\theta}$ zu groß wird. Der kritische Wert muß nun noch über die Bedingung

$$P_\theta(\{x \in \mathfrak{X} | \hat{\theta}_n \geq c(\alpha)\}) \leq \alpha \ \forall \theta \in \Theta_0$$

gefunden werden. Die linke Seite ist maximal für $\theta = \theta_0$, und unter P_{θ_0} hat $\frac{1}{\hat{\theta}_n} = \frac{1}{n}\sum_{i=1}^n X_i$ eine Gamma-Verteilung mit Parametern n und $\frac{\theta_0}{n}$. Hiermit kann $c(\alpha)$ bestimmt werden.

Man erhält den gleichen Test als besten Test (Neyman-Pearson), wenn man $H_0 : \theta = \theta_0$ gegen $H_1 : \theta = \theta_1$ bei beliebigem $\theta_1 > \theta_0$ testet; damit folgt wie in Beispiel 7.14, daß der obige Test unter allen Tests zum Niveau α die Fehlerwahrscheinlichkeit 2. Art gleichmäßig minimiert.

Beispiel 22.2
Wir betrachten ein Zufallsexperiment mit r möglichen Ausgängen der Wahrscheinlichkeiten $p_1, \ldots, p_r$. Der Zufallsvektor $X_i = (X_{i1}, \ldots, X_{ir})$ gäbe das Resultat der i-ten Wiederholung an; $X_{ij} = 1$ oder 0,

je nachdem ob das Resultat j eingetreten ist oder nicht (siehe auch Absatz 6.2.5 und Absatz 8.5). Mit $y_n := x_1 + \cdots + x_n$ gilt

$$f(x_1, \ldots, x_n | p_1, \ldots, p_r) = \prod_{i=1}^{r} p_i^{y_{ni}}$$

mit y_{ni}=Anzahl Versuche mit Resultat i.

Der MLS für $(p_1, \ldots, p_r)$ ist $(\hat{p}_{n1}, \ldots, \hat{p}_{nr})$ mit $\hat{p}_{ni} := \frac{y_{ni}}{n}$:

$$\begin{aligned}
\log f(x_1, \ldots, x_n | p_1, \ldots, p_r) - \log f(x_1, \ldots, x_n | \hat{p}_{n1}, \ldots, \hat{p}_{nr}) &= \sum_{i=1}^{r} y_{ni} \log \left(\frac{p_i}{\hat{p}_{ni}} \right) \\
&\overset{\log x \leq x-1}{\leq} \sum_{i=1}^{r} y_{ni} \left(\frac{np_i}{y_{ni}} - 1 \right) \\
&= n \cdot \underbrace{\sum_{i=1}^{r} p_i}_{=1} - \underbrace{\sum_{i=1}^{r} y_{ni}}_{=n} \\
&= 0
\end{aligned}$$

Will man testen, ob ein bestimmter Vektor $p_1, \ldots, p_r$ vorliegt, so führt der LQ-Test auf die Testgröße

$$T(x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n | \hat{p}_{n1}, \ldots, \hat{p}_{nr})}{f(x_1, \ldots, x_n | p_1, \ldots, p_r)} = \prod_{i=1}^{r} \left(\frac{y_{ni}}{np_i} \right)^{y_{ni}}$$

Was passiert bei großem n? $Z_{ni} := \frac{(Y_{ni} - np_i)}{\sqrt{np_i}}$, dann

$$\begin{aligned}
2 \cdot \log T(X_1, \ldots, X_n) &= 2 \cdot \sum_{i=1}^{r} (np_i + (\underbrace{Y_{ni} - np_i}_{Z_{ni} \cdot \sqrt{np}})) \underbrace{\log \left(1 + \frac{Z_{ni}}{\sqrt{np_i}} \right)}_{\log(1+x) = x + \frac{1}{2}x^2 + \ldots} \\
&= \underbrace{\sum_{i=1}^{r} \frac{(Y_{ni} - np_i)^2}{(np_i)}}_{\text{Testgröße zum } \chi^2\text{-Anpassungstest}} + \underbrace{R_n}_{\xrightarrow{P} 0}
\end{aligned}$$

Man hat hier im Hintergrund einen allgemeineren Sachverhalt:

Satz 22.3
Unter den Voraussetzungen von Satz 21.3 gilt für die Testgröße T_n des LQ-Tests zur Hypothese $H_0 = \theta = \theta_0$, wenn $X_1, X_2, \ldots$ eine unabhängige Folge von Zufallsgrößen mit Verteilung P_{θ_0} ist,

$$2 \cdot \log T_n \xrightarrow{\mathcal{D}} Z, \quad Z \sim \chi_k^2.$$

Beweis
Es sei $L_n(\theta|\omega) = \sum_{m=1}^{n} \log f(X_m(\omega)|\theta)$ die Log-Likelihood-Funktion mit erster Ableitung

$$(s_n(\theta, \omega))_i = \sum_{m=1}^{n} \frac{\partial}{\partial \theta_i} \log f(X_m(\omega)|\theta), \ 1 \leq i \leq k,$$

(Spaltenvektor), und zweiter Ableitung

$$(A_n(\theta|\omega))_{ij} = \sum_{m=1}^{n} \frac{\partial^2}{\partial\theta_i\partial\theta_j} \log f(X_m(\omega)|\theta),\ 1 \le i,j \le k.$$

Dann gilt mit einem $\tilde{\theta}_n(\omega)$ zwischen θ und $\hat{\theta}_n(\omega)$:

$$L_n(\hat{\theta}_n) = L_n(\theta) + s_n(\theta)^t(\hat{\theta}_n - \theta) + \tfrac{1}{2}(\hat{\theta}_n - \theta)^t A_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta)$$

sowie mit einem $\check{\theta}_n$

$$s_n(\theta)^t(\hat{\theta}_n - \theta) = \underbrace{s_n(\hat{\theta}_n)^t}_{=0}(\hat{\theta}_n - \theta) + (\theta - \hat{\theta}_n)^t A_n(\check{\theta}_n)(\hat{\theta}_n - t),$$

also

$$2\log T_n = 2(L_n(\hat{\theta}_n) - L_n(\theta)) = (\hat{\theta}_n - \theta)^t(A_n(\tilde{\theta}_n) - 2A_n(\check{\theta}_n))(\hat{\theta}_n - \theta)$$
$$(\text{Satz 21.3: } \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \frac{1}{I(\theta)})$$

Wie im Beweis zu Satz 21.3:

$$\frac{1}{n}A_n(\tilde{\theta}_n) \xrightarrow{\mathrm{P}} -I(\theta), \quad \frac{1}{n}A_(\check{\theta}_n) \xrightarrow{\mathrm{P}} -I(\theta),$$

d.h. die „Matrix in der Mitte" konvergiert gegen $I(\theta)$.

also von oben fortgesetzt:

$$2\log T_n = \sqrt{n}(\hat{\theta}_n - \theta)I(\theta)\sqrt{n}(\hat{\theta}_n - \theta) \to \|Z\|^2, \quad Z \sim N_k(0, I_k).$$

Das bereits bei Konfidenzellipsoiden verwendete Argument liefert nun die Behauptung. *q.e.d.*

31. Januar 1995

23 Normalverteilte Beobachtungen

Die Verfahren in diesem Abschnitt sind echte Klassiker und werden sehr häufig (vielleicht zu häufig) verwendet.

23.1 Verteilungen

Uns ist bekannt, daß $\sum_{i=1}^n \chi_i^2$ einer unabhängigen $N(0,1)$-verteilten Zufallsvariablen χ_n^2-verteilt ist. Wir benötigen zwei weitere Verteilungsklassen.

Definition 23.1

(i) Die *Student'sche t-Verteilung* mit n Freiheitsgraden ($n \in \mathbb{N}$), Symbol t_n, ist die Verteilung mit der Lebesgue-Dichte f_{t_n},

$$f_{t_n} = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$$

(ii) Die F-Verteilung mit m und n Freiheitsgraden ($m, n \in \mathbb{N}$), Symbol $F_{m,n}$, ist die auf $(0,\infty)$ konzentrierte Verteilung mit der Lebesgue-Dichte $f_{F_{m,n}}$,

$$f_{F_{m,n}} = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \cdot \frac{m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}}, \quad x > 0.$$

Daß diese Funktionen tatsächlich Wahrscheinlichkeitsdichten sind (also isb. das Integral 1 haben), kann man entweder hier bereits nachrechnen (viel Spaß ...) oder aus Lemma 23.3 folgern. Die Verteilungsfunktionen χ^2_n, t_n, $F_{m,n}$ sind vielerorts und ausgiebig vertafelt. Um zeigen zu können, wie diese Verteilungen mit der Normalverteilung zusammenhängen, benötigen wir das folgende Hilfsmittel:

Lemma 23.2
Sind X und Y unabhängige Zufallsvariablen mit Lebesgue-Dichten f_X und f_Y und hilt $Y > 0$, dann hat $Z := \frac{X}{Y}$ die Wahrscheinlichkeitsdichte

$$f_Z(z) = \int_0^\infty y f_Y(y) f_X(yz)\, dy.$$

Beweis

Mit Fubini, Transformationssatz und einer Substitution ergibt sich als Verteilungsfunktion von Z:

$$\begin{aligned}
F_Z(Z) &= P\left(\frac{X}{Y} \le Z\right) \\
&= \int_{\{\omega\in\Omega \mid X(\omega)\le ZY(\omega)\}} 1\ P(d\omega) \\
&= \int_{\{(x,y)\in\mathbb{R}\times(0,\infty) \mid x\le yz\}} f_X(x) f_Y(y) \ell^2\ (dx, dy) \\
&= \int_0^\infty \int_{-\infty}^{yz} f_X(x)\, dx\ f_Y(y)\, dy \\
&= \int_0^\infty \int_{-\infty}^{z} f_X(x,y) y f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{z} \left(\int_0^\infty y \cdot f_X(x,y) f_Y(y)\, dy\right) dx.
\end{aligned}$$

q.e.d.

Lemma 23.3
Es gilt

(i) X, Y unabhängig, $X \sim N(0,1)$, $Y \sim \chi^2_n \implies \frac{\sqrt{n}X}{\sqrt{Y}} \sim t_n$,

(ii) X, Y unabhängig, $X \sim \chi^2_m$, $Y \sim \chi^2_n \implies \frac{\frac{1}{m}X}{\frac{1}{n}Y} \sim F_{m,n}$

Beweis

(ii) Die χ^2_k-Verteilung hat Dichte $f_{\chi^2_k}(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2-1}} e^{-\frac{x}{2}} > 0$. Hat X die Dichte f_X, so hat αX, $\alpha > 0$, die Dichte $x \to \frac{1}{\alpha} f_X(\frac{1}{\alpha x})$, damit erhält man für $\frac{1}{m}X$ die Dichte

$$\begin{aligned} f_{\frac{1}{m}X}(x) &= m \cdot f_X(mx) \\ &= \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} \cdot m^{\frac{m}{2}} x^{\frac{m}{2}-1} e^{-\frac{m}{2}x}, \end{aligned}$$

und analog für $\frac{1}{n}Y$ die Dichte

$$f_{\frac{1}{n}Y}(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \cdot n^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{n}{2}x},$$

Mit Lemma 23.2 ergibt sich für die Dichte des Quotienten $\frac{\frac{1}{m}X}{\frac{1}{n}Y}$

$$\begin{aligned} & \int_0^\infty y f_{\frac{1}{n}Y}(y) f_{\frac{1}{m}X}(yz)\, dy \\ = \quad & \frac{1}{2^{\frac{n+m}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \cdot m^{\frac{m}{2}} n^{\frac{n}{2}} \cdot \int_0^\infty y^{\frac{n}{2}} (yz)^{\frac{m}{2}-1} e^{-\frac{ny}{2}} e^{-\frac{myz}{2}}\, dy \\ \overset{t:=\frac{1}{2}y(n+mz))}{=} \quad & \frac{m^{\frac{m}{2}} n^{\frac{n}{2}}}{2^{\frac{n+m}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \cdot z^{\frac{m}{2}-1} \cdot \frac{2^{\frac{n+m}{2}}}{(n+mz)^{\frac{n+m}{2}}} \cdot \int_0^\infty t^{\frac{n+m}{2}-1} \cdot e^{-t}\, dt \\ = \quad & \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \cdot m^{\frac{m}{2}} \cdot n^{\frac{n}{2}} \cdot \frac{z^{\frac{m}{2}-1}}{(n+mz)^{\frac{n+m}{2}}} \end{aligned}$$

(i) Man kann wieder Lemma 23.2 verwenden oder aber wie folgt verfahren: Die Verteilung von $Z := \frac{\sqrt{n}X}{\sqrt{Y}}$ ist symmetrisch um 0. Wegen

$$\begin{aligned} \frac{d}{dx} P(Z^2 \le x) &= 2\frac{d}{dx} P(0 \le Z \le \sqrt{x}) \\ &= 2\frac{d}{dx} \int_0^{\sqrt{x}} f_Z(z)\, dz \\ &= \frac{1}{\sqrt{x}} f_Z(\sqrt{x}) \end{aligned}$$

läßt sich die (symmetrische) Dichte von Z also aus der Dichte von Z^2 zurückgewinnen. Nach dem bereits bewiesenen Teil (ii) gilt aber $Z^2 \sim F_{1,n}$, d.h. es bleibt zu zeigen, daß

$$\frac{1}{\sqrt{x}} f_{t_n}(\sqrt{x}) = f_{F_{1,n}}(x) \quad \text{für alle } x > 0$$

gilt. Dies ist einfach durchzuführen (beachte: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$).

q.e.d.

23.2 Tests und Konfidenzintervalle für Mittelwert und Varianz

Der folgende Satz ist die Basis für viele klassische Verfahren.

Satz 23.4
Sei $X_1, \ldots, X_n$ eine Stichprobe aus $N(\mu, \sigma^2)$, $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$, $S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$, $T_n = \sqrt{n}\frac{\overline{X_n}-\mu}{S_n}$. Dann gilt

(i) $\overline{X_n}$ und S_n^2 sind unabhängig.

(ii) $\sqrt{n}\frac{\overline{X_n}-\mu}{\sigma} \sim N(0,1)$,

(iii) $(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2_{n-1}$,

(iv) $T_n \sim t_{n-1}$.

Beweis

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

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

Sei nun A eine orthogonale $n \times n$-Matrix mit erster Zeile $(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$. (Das Gram-Schmidt-Verfahren kann zur Konstruktion einer solchen Matrix verwendet werden, für uns ist hier nur die Existenz wesentlich). Sei $Z := A \cdot Y$. Dann gilt

$$Z \sim N_n(0, I_n) \quad \text{(siehe Abschnitt 19)},$$

d.h. die Konponenten von $Z_1, \ldots, Z_n$ von Z bilden wieder eine Stichprobe aus $N(0,1)$. Nach Konstruktion gilt $Z_1 = \sqrt{n}\cdot\overline{Y_n}$, also folgt (ii).

Die Orthogonalität von A liefert außerdem

$$\sum_{i=1}^{n} Z_i^2 = Z^T Z = Z^T A^T A Z = Y^T Y = \sum_{i=1}^{n} Y_i^2,$$

also

$$\begin{aligned}\sum_{i=1}^{n}(Y_i-\overline{Y_n})^2 &= \sum_{i=1}^{n} Y_i^2 - n(\overline{Y_n})^2\\ &= \sum_{i=1}^{n} Z_i^2 - Z_1^2\\ &= \sum_{i=2}^{n} Z_i^2.\end{aligned}$$

Folglich ist S_n^2 eine Funktion von $Z_2, \ldots, Z_n$ und damit unabhängig von $\overline{X_n}$, das ja eine Funktion von Z_1 ist. Damit ist (i) gezeigt.

Außerdem liefert diese Darstellung auch (iii), denn die Beziehung $(n-1)fS_n^2\sigma^2 = \sum_{i=2}^n Z_i^2$ zeigt, daß $(n-1)\frac{S_n^2}{\sigma^2}$ die Summe der Quadrate von $n-1$ unabhängigen, $N(0,1)$-verteilten Zufallsvariablen ist.

Teil (iv) schließlich folgt mit Lemma 23.3(i) aus den bereits bewiesenen Teilen (i), (ii) und (iii). *q.e.d.*

Beispiel 23.5 (Tests und Konfidenzintervalle für den Mittelwert)
Ist $X_1, \ldots, X_n$ eine Stichprobe aus einer Normalverteilung $N(\mu, \sigma^2)$ mit unbekannten $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, so folgt mit Satz 23.4(iv), wenn $t_{n,\beta}$ das β-Quantil zu t_n bezeichnet (Tafel, Computer)

$$P_{\mu,\sigma^2}(t_{n-1,\frac{\alpha}{2}} < \tfrac{\overline{X_n}-\mu}{\frac{S_n}{\sqrt{n}}} \leq t_{n-1,1-\frac{\alpha}{2}}) = 1-\alpha \tag{$*$}$$

für alle $\mu \in \mathbb{R}$ und $\sigma^2 > 0$.

Löst man die Ungleichungen nach μ auf, so sieht man, daß

$$\left[\overline{X_n} - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{S_n}{\sqrt{n}} \quad , \quad \overline{X_n} - t_{n-1,\frac{\alpha}{2}} \cdot \frac{S_n}{\sqrt{n}}\right]$$

ein $100(1-\alpha)$%-Konfidenzintervall für μ ist. Die Symmetrie der t-Verteilungen um Null impliziert $t_{n-1,\frac{\alpha}{2}} = -t_{n-1,1-\frac{\alpha}{2}}$, für große n kann man t_n durch $N(0,1)$ ersetzen (siehe Aufgabe 60) und erhält die in Beispiel 20.5 hergeleiteten aymptotischen Konfidenzintervalle.

Außerdem führt die in Absatz 7.4 besprochene Beziehung zwischen Konfidenzintervallen und Tests auf den berühmten t-Test zur Hypothese $H_0 : \mu = \mu_0$: verwerfe, wenn $\left|\overline{X_n} - \mu_0\right| > \frac{S_n}{\sqrt{n}} \cdot t_{n-1,1-\frac{\alpha}{2}}$.

(Aus $(*)$ folgt direkt, daß dieser Test die Fehlerwahrscheinlichkeit α hat und zwar für alle $\sigma^2 > 0$!)

Die Größe T_n ist eine Funktion der Daten $X_1, \ldots, X_n$ und des unbekannten Parameters μ mit den Eigenschaften

(i) Die Verteilung von T_n hängt nicht vom unbekannten Parameter (μ, σ^2) ab,

(ii) Ungleichungen der Form $T_n \leq t$ können nach μ aufgelöst werden.

2. Februar 1995

$X_1, \ldots, X_n$ Stichprobe aus (N, μ, σ^2).

t-Test beruht auf:

$$\underbrace{\frac{\sqrt{n}(\overline{X_n} - \mu)}{S_n}}_{\substack{\text{Funktion von Daten und Parameter,} \\ \text{kann nach } \mu \text{ aufgelöst werden.}}} \qquad \underbrace{\sim t_{n-1}}_{\text{hängt nicht ab von } \mu \text{ oder } \sigma 2} \qquad \text{unter } P_{\mu,\sigma^2}.$$

Der t-Test ist ein *Pivot*.

Beispiel 23.6 (Tests und Konfidenzintervalle für die Varianz)
Satz 23.4(iii) zeigt, daß $\frac{(n-1)S_n^2}{\sigma^2}$ ein Pivot für σ^2 ist. Bezeichne $\chi^2_{n,\beta}$ das β-Quantil zu χ^2_n, so folgt

$$P_{\mu,\sigma^2}\left(\chi^2_{n-1;\frac{\alpha}{2}} \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{n-1;1-\frac{\alpha}{2}}\right) = 1-\alpha \quad \forall \mu, \sigma^2$$

also ist

$$\left[\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

ein $100(1-\alpha)$%-Konfidanzintervall für σ^2.

Durch die Vorschrift

$$„\text{Verwerfe, wenn } \frac{(n-1)S_n^2}{\sigma_0^2} > \chi^2_{n-1;1-\frac{\alpha}{2}} \text{ oder } \frac{(n-1)S_n^2}{\sigma_0^2} < \chi^2_{n-1;\frac{\alpha}{2}}\text{“},$$

wird ein Test zum Niveau α für $H_0 : \sigma^2 = \sigma_0^2$ definiert.

Mit den obigen Methoden können auch Tests zu einseitigen Hypothesen und Konfidenzober- bzw. -unterschranken konstruiert werden.

Eine „einseitige“ Variante des t-Tests für die Hypothese $H_0 : \mu \le \mu_0$ gegen $H_1 : \mu > \mu_0$ wird gegeben durch

$$„\text{Verwerfe, wenn } \sqrt{n}\cdot\frac{\overline{X_n}-\mu_0}{S_n} \ge t_{n-1;1-\alpha}.\text{“}$$

Ist $\mu \le \mu_0$, so ist $\sqrt{n}\cdot\frac{\overline{X_n}-\mu}{S_n} \sim t_{n-1}$ unter P_{μ,σ^2}, also

$$\begin{aligned}
P_{\mu,\sigma^2}\left(\sqrt{n}\cdot\frac{\overline{X_n}-\mu_0}{S_n} \ge t_{n-1;1-\alpha}\right) &= P_{\mu,\sigma^2}\left(\sqrt{n}\cdot\frac{\overline{X_n}-\mu}{S_n} + \sqrt{n}\cdot\frac{\mu-\mu_0}{S_n} \ge t_{n-1;1-\alpha}\right) \\
&\overset{\sqrt{n}\cdot\frac{\mu-\mu_0}{S_n}\le 0 \text{ auf } H_0}{\le} P_{\mu,\sigma^2}\left(\sqrt{n}\cdot\frac{\overline{X_n}-\mu}{S_n} \ge t_{n-1;1-\alpha}\right) \\
&\le \alpha,
\end{aligned}$$

der Test hält also das vorgegebene Niveau ein.

Wie sieht der Likelihood-Quotienten-Test zu $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \ne \mu_0$ aus? Sei

$$\check{S}_n^2 := \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X_n})^2,$$

dann folgt mit Beispiel 21.1(iii)

$$\begin{aligned}
\sup_{\mu\in\mathbb{R},\sigma^2>0} f(x|\mu,\sigma^2) &= f(x|\overline{x_n},\check{S}_n^2) \\
&= (2\pi\check{S}_n^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\check{S}_n^2}\sum_{i=1}^n (X_i-\overline{X_n})^2\right) \\
&= (2\pi\check{S}_n^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}
\end{aligned}$$

Bei festem μ wird $f(x|\mu,\sigma^2)$ als Funktion von σ^2 maximal in

$$\tilde{S}_n^2 := \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2,$$

d.h. als Testgröße erhält man

$$\begin{aligned}
T_n &= \frac{f(x|\overline{x_n},\check{S}_n^2)}{f(x|\mu_0,\tilde{S}_n^2)} \\
&= \frac{(2\pi\check{S}_n^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}}{(2\pi\tilde{S}_n^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}} \\
&= ???
\end{aligned}$$

ist eine streng monoton steigende Funktion von $\frac{\tilde{S}_n^2}{\check{S}_n^2}$. Beachte nun

$$\begin{aligned}\sum_{i=1}^n (x_i-\mu_0)^2 &= \sum_{i=1}^n (x_i-\overline{x_n}+\overline{x_n}-\mu_0)^2\\ &= \sum_{i=1}^n (x_i-\overline{x_n})^2 + 2\sum_{i=1}^n (x_i-\overline{x_n})(\overline{x_n}-\mu_0) + \sum_{i=1}^n (\overline{x_n}-\mu_0)^2\\ &= n(\overline{x_n}-\mu_0)^2 + n\check{S}_n^2,\end{aligned}$$

also

$$\tilde{S}_n^2 = (\overline{x_n}-\mu_0)^2 + \check{S}_n^2,$$

d.h. T_n ist eine monoton steigende Funktion von $\frac{(\overline{x_n}-\mu_0)^2}{\check{S}_n^2}$, also auch von $\frac{\sqrt{n}\cdot|\overline{x_n}-\mu_0|}{\check{S}_n}$. Der t-Test aus Beispiel 23.??? lehnt ab, wenn dieser Wert zu groß wird, ist also der LQ-Test.

23.3 Modelle mit Hilfsvariablen (Kovariaten)

Bisher: Daten $y_1,\dots,y_n$, Stichprobe aus einer festen, unbekannten Verteilung.

In der Praxis gehören häufig zu jedem y_i Hilfsvariablen (auch „erklärende Variablen“ oder Kovariaten)), man interessiert sich für die Abhängigkeit von diesen.

Klassisches Beispiel: Lineare Regression. Zu jedem y_i sind Werte $x_{i1},\dots,x_{ik}$ gegeben, man setzt einen durch zufällige Fehler gestörten linearen Zusammenhang voraus:

$$y_i = x_{i1}\beta_1,\dots,x_{ik}\beta_k + \varepsilon_i, \tag{$*$}$$

wobei $\beta_1,\dots,\beta_k$ unbekannt sind und $\varepsilon_1,\dots,\varepsilon_n$ als unabhängig und $N(0,\sigma^2)$-verteilt vorausgesetzt werden; $\sigma^2>0$ wieder unbekannt.

In Matrizenschreibweise wird $(*)$ zu

$$y = X\beta + \varepsilon,$$

wobei X, die soganannte *Design-Matrix*, in Zeile i, Spalte j das Element x_{ij} hat, $y=(y_1,\dots,y_n)^T$ ist der Vektor der *abhängigen Variablen* (response) und $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)^T$ ist der Fehlervektor. Der Datenvektor y ist damit Realisierung eines stochastischen Vektors Y, der eine $N_n((X\beta,\sigma^2\cdot I_n)$-Verteilung hat mit unbekanntem $\beta\in\mathbb{R}^k$ und $\sigma^2>0$.

Die auf Gauß zurückgehende Methode der kleinsten Quadrate schätzt β durch den Wert $\hat\beta$, für den die Summe der quadrierten Residuen $r_i := y_i-(x_{i1}\hat\beta_1+\cdots+x_{ik}\hat\beta_k)$ minimal wird, d.h. durch den Vektor (wenn es einen gibt), der die Funktion $\beta\to\|y-X\beta\|^2$ minimiert. Das Maximum-Likelihood-Prinzip führt wegen

$$\log f(y|\beta,\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\underbrace{(y-X\beta)^T(y-X\beta)}_{\|y-X\beta\|^2}$$

auf dasselbe Problem. Beide Methoden liefern auch denselben Schätzer für β; bei invertierbarem X^TX ist dies

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \text{(Aufgabe 58).}$$

Ziel: Optimalitätsaussage für $\hat{\beta}$ von der Form „$\hat{\beta}$ hat kleinste Varianz in der Klasse ... “

Hierzu: Was heißt „Varianz klein“? Die Varianz (=Kovarianzmatrix) eines Zufallsvektors ist bekanntlich eine nicht-negativ definite, symmetrische Matrix. Sind A und B zwei solche Matrizen auf $\mathbb{R}^{k\times k}$, so nennen wir A kleiner als oder gleich B, wenn gilt:

$$x^T A x \leq x^T B x \quad \text{für alle } x \in \mathbb{R}^k.$$

Satz 23.7
Es sei $Y \sim N_n(X\beta, \sigma^2 \cdot I_n)$ mit unbekanntem $\beta \in \mathbb{R}^k, \sigma^2 > 0$. Ist $X^T X$ invertierbar, so ist $\hat{\beta} := (X^T X)^{-1} X^T Y$ ein erwartungstreuer Schätzer für β, der außerdem in der Klasse aller erwartungstreuen Schätzer für β die gleichmäßig kleinste Varianz hat. Kurz: $\hat{\beta}$ ist UMVU, uniformly minimum various unbiased).

7. Februar 1995

Beweis

$$\begin{aligned} \mathrm{E}_{\beta,\sigma^2}\hat{\beta} &= (X^T X)^{-1} \cdot X^T \cdot \mathrm{E}_{\beta,\sigma^2} Y \\ &= (X^T X)^{-1} \cdot X^T \cdot X\beta \\ &= \beta \\ &\Longrightarrow \text{erwartungstreu} \end{aligned}$$

Sei nun $\tilde{\beta}$ irgendein anderer erwartungstreuer Schätzer für β, sei $a \in \mathbb{R}^k$. Dann gilt für $U := \tilde{\beta} - \hat{\beta}$:

$$\mathrm{E}_{\beta,\omega^2} a^T U = 0,$$

d.h.

$$\int a^T U(y)(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T (y - X\beta)\right) dy = 0$$

Multiplikation mit $\exp(\frac{1}{2\omega^2}\beta^T X^T X\beta)$ ergibt:

$$(2\pi\sigma^2)^{-\frac{n}{2}} \int a^T U(y) \exp\left(-\frac{1}{2\sigma^2}(y^T y - 2y^T X\beta)\right) dy = 0.$$

Bilde die partiellen Ableitungen nach β_j, $1 \leq j \leq k$, Differentiation und Integration dürfen vertauscht werden (siehe Aufg. 56):

$$(2\pi\sigma^2)^{-\frac{n}{2}} \cdot \int a^T U(y) \cdot \left(\sum_{i=1}^n x_{ij} y_i\right) \exp\left(-\frac{1}{2\sigma^2} \cdot (y^T y - 2y^T X\beta)\right) dy = 0.$$

Multipliziert man mit $\exp(-\frac{1}{2\sigma^2}\beta^T X^T X\beta)$, und faßt man die k Gleichungen zusammen, so folgt

$$\mathrm{E}_{\beta,\sigma^2} a^T U X^T Y = 0.$$

Multipliziert man von links mit $a^T(X^TX)^{-1}$, so folgt

$$\mathrm{E}_{\beta,\sigma^2}(a^T\tilde{\beta} - a^T\hat{\beta})(a^T\hat{\beta}) = 0,$$

also wegen $\mathrm{E}_{\beta,\sigma^2}(a^T\tilde{\beta} - a^T\hat{\beta}) = 0$

$$\mathrm{cov}_{\beta,\sigma^2}(a^T\tilde{\beta} - a^T\hat{\beta}, a^T\hat{\beta}) = 0.$$

Insgesamt:

$$\begin{aligned}
\mathrm{var}_{\beta,\sigma^2}(a^T\tilde{\beta}) &= \mathrm{E}(a^T\tilde{\beta} - a^T\hat{\beta} + a^T\hat{\beta} - a^T\beta)^2 \\
&= \underbrace{\mathrm{E}(a^T\tilde{\beta} - a^T\hat{\beta})^2}_{\geq 0} + \underbrace{2 \cdot \mathrm{cov}(a^T\tilde{\beta} - a^T\hat{\beta}, a^T\hat{\beta})}_{=0} + \mathrm{var}(a^T\hat{\beta}) \\
&\geq \mathrm{var}_{\beta,\sigma^2}(a^T\hat{\beta})
\end{aligned}$$

q.e.d.

Kovariaten können quantitativ (Ernteertrag in Abhängigkeit von der Dosis eines besitmmten Düngemittels) oder qualitativ (Ernteertrag in Abhängigkeit vom Typ des verwendeten Düngemittels) sein.

Typisches Testverfahren

Beispiel 23.8 (Varianzanalyse bei einem Faktor; ANOVA, analysis of variance)
Situation mit einer qualifizierten Hilfsvariablen (auch „unabhängige" Variable), die k verschiedene Ausprägungen (auch: *Niveaus*) haben kann. In der i-ten Gruppe, $1 \leq i \leq k$, liegen n_i Messungen $y_{i1}, \ldots, y_{in_i}$ vor.

Modell: $y_{ij} = \beta_i + \varepsilon_{ij}$ mit unbekannten Parametern $\beta_1, \ldots, \beta_k$ („Haupteffekt" der Gruppe i) und unabhängigen, $N(0, \sigma^2)$-verteilten ε_{ij}, $1 \leq i \leq k$, $1 \leq j \leq n_i$, auch die Varianz der Fehler ist unbekannt.

Wichtige Hypothese H_0: $\beta_1 = \beta_2 = \cdots = \beta_k$.

In Worten: Das Niveau des Faktors hat keinen Einfluß. Beachte: H_0 ist zusammengesetzt.

Dieses Modell kann in die Form $y = X\beta + \varepsilon$ gebracht werden:

$$y = (y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}, \ldots, y_{k1}, \ldots, y_{kn_k})^T.$$

analog mit ε, und

$$X = \begin{pmatrix} 1 & & & \\ \cdots & & & \\ 1 & & & \\ & 1 & & 0 \\ & \cdots & & \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \\ & & & \cdots \\ & & & 1 \end{pmatrix} \begin{array}{l} \\ n_1 \text{ Zeilen} \\ \\ \\ n_2 \text{ Zeilen} \\ \\ \vdots \\ \\ n_k \text{ Zeilen} \\ \\ \end{array}$$

Offensichtlich gilt

$$X^T X = \text{diag}(n_1, \ldots, n_k),$$

als MLS für β ergibt sich

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{pmatrix} y_1. \\ \ldots \\ y_k. \end{pmatrix}, \quad \text{mit}$$

$$y_i. := \frac{1}{n_i} \cdot \sum_{j=1}^{n_i} y_{ij}.$$

Der Likelihood-Quotienten-Test führt auf die Testgröße $\frac{SSH}{SSE}^{\frac{n}{2}}$, wobei $n = n_1 + \cdots + n_k$ (Gesamtzahl der Beobachtungen), $SSH = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \overline{y})^2$, $SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - y_i.)^2$, $\overline{y} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}$ $(= \frac{1}{n}(n_1 y_1. + \cdots + n_k y_k.))$.

SSH ist die Variabilität der Daten bei Berücksichtigung der Hypothese, SSE die Variabilität ohne die durch H_0 gegebene Einschränkung. SSH läßt sich zerlegen:

$$\begin{aligned} SSH &= \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - y_i. + y_i. - \overline{y})^2 \\ &= SSE + 2 \cdot \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - y_i.)(y_i. - \overline{y}) + \sum_{i=1}^k n_i (y_i - \overline{y})^2. \end{aligned}$$

Die Variabilität unter H_0 ist also die Summe von zwei Termen: der Variabilität *innerhalb* der Gruppen, SSE, und der Variabilität *zwischen* den Gruppen. (Diese Variabilitätszerlegung erklärt den Namen „Varianzanalyse".) Der Test lehnt also ab, wenn die Variabilität zwischen den Gruppen (zu) groß ist gegenüber der Variabilität innerhalb der Gruppen.

Ist H_0 richtig, so gilt

(i) $\frac{1}{\sigma^2} \cdot \sum_{i=1}^k n_i (y_i. - \overline{y})^2 \sim \chi^2_{k-1}$.

(ii) $\frac{1}{\sigma^2} \cdot SSE \sim \chi^2_{n-k}$.

(iii) Die Zufallsvariablen aus (i) und (ii) sind unabhängig.

Hiermit und mit Lemma 23.3(ii) ergibt sich nun der F-Test der Varianzanalyse: Verwerfe H_0 auf dem Niveau α, wenn

$$\frac{\frac{1}{k-1}(SSH - SSE)}{\frac{1}{n-k} SSE} \geq F_{k-1;n-k;1-\alpha}$$

gilt.

(Weiteres in der Vorlesung „Lineare Modelle")

24 Suffizienz

Ziel: Verallgemeinerung von Abschnitt 7.5, hierfür nötig: Bedingte Verteilungen.

24.1 Bedingte Verteilungen

Wir betrachten folgende Situation: $(\Omega, \mathfrak{A}, P)$ ist ein Wahrscheinlichkeitsraum, $(E, \mathfrak{E})$ und $(S, \mathfrak{S})$ meßbare Räume und $X : \Omega \to E$, $T : \Omega \to S$ $(\mathfrak{A}, \mathfrak{E})$- bzw. $(\mathfrak{A}, \mathfrak{S})$-meßbare Abbildungen.

$$\begin{tikzcd} (\Omega, \mathfrak{A}, P) \arrow[r, "X"] \arrow[d, "T"'] & (E, \mathfrak{E}) \\ (S, \mathfrak{S}) & \end{tikzcd}$$

Wie wirkt sich Kenntnis des Wertes von T auf die Verteilung von X aus?
($P(X \in A | T = t) = \frac{P(X \in A, T = t)}{P(T = t)}$ geht nicht, da $P(T = t)$ häufig $\equiv 0$ ist.)

Definition 24.1
Eine Abbildung $Q : S \times \mathfrak{E} \to \mathbb{R}$ heißt *(reguläre Version der) bedingte(n) Verteilung von X unter T*, wenn gilt

(i) $t \to Q(t, b)$ ist $(\mathfrak{S}, \mathfrak{B})$-meßbar für alle $B \in \mathfrak{E}$.

(ii) $B \to Q(t, B)$ ist ein Wahrscheinlichkeitsmaß auf $(E, \mathfrak{E})$ für alle $t \in S$.

(iii) Für alle $A \in S$, $B \in \mathfrak{E}$ gilt

$$P(T \in A, X \in B) = \int_A Q(t, B) P^T\ (dt)$$

9. Februar 1995

$$\begin{tikzcd} (E, \mathfrak{E}) \arrow[d, "\times"', "\overset{\varphi}{=}\mathbb{R}"] \\ (S, \mathfrak{S}) \end{tikzcd}$$

Unter gewissen Bedingungen an $(E, \mathfrak{E})$ (bei $(\mathbb{R}^n, \mathfrak{B}^n)$ erfüllt) existiert Q, außerdem kann $(*)$ verallgemeinert werden zu

$$\mathrm{E}\varphi(X, T) = \int_S \int_E \varphi(t, x) Q(t, dx) P^T\ (dt) \quad \text{(„verallgemeinerter Fubini“)}$$

Wichtiger Spezialfall:

Satz 24.2
Es seien ν ein Maß auf $(E, \mathfrak{E})$, μ ein Maß auf $(S, \mathfrak{S})$ und Q_0 ein beliebiges Wahrscheinlichkeitsmaß auf $(E, \mathfrak{E})$. Wir setzen voraus, daß (X, T) bezüglich $\nu \otimes \mu$ eine Dichte $f_{X,T} : E \times S \to \mathbb{R}$ hat. Dann ist

$$f_T : \begin{cases} S \to \mathbb{R} \\ t \mapsto \int f_{X,T}(x, t) \nu\ (dx) \end{cases}$$

eine μ-Dichte von P^T und $Q : S \times \mathfrak{E} \to \mathbb{R}$,

$$Q(t, B) := \begin{cases} \frac{\int_B f_{X,T}(x,t) \nu\ (dx)}{f_T(t)}, & \text{falls } f_T(t) > 0, \\ Q_0(B), & \text{sonst} \end{cases}$$

ist eine reguläre Version der bedingten Verteilung von X unter T. In einprägsamer, wenn auch unsauberer Form:

$$f_{X|T}(x|t) = \frac{f_{X,T}(x,t)}{f_T(t)}$$

Im diskreten Fall hatten wir $P(X = x|T = t) = \frac{P(X=x,T=t)}{P(T=t)}$.

Beweis

(i) folgt mit den Meßbarkeitsaussagen auf Abschnitt 14.1, (ii) ist trivial. Mit $P(T \in \{f_T = 0\}) = \int_{\{f_T=0\}} f_T \, d\mu = 0$ folgt

$$\begin{aligned}
\int_A Q(t, BP^T\,(dt) &= \int_{A\cap\{f_T>0\}} Q(t,B) f_T(t)\mu\,(dt) \\
&= \int_{A\cap\{f_T>0\}} \int_B f_{X,T}(x,t)\nu\,(dx)\,\mu\,(dt) \\
&= \int_{(A\cap\{f_T>0\})\times B} f_{X,T}(x,t)\nu\otimes\mu\,(dx,dt) \\
&= P(T\in A\cap\{f_T>0\}, X\in B) \\
&= P(T\in A, X\in B).
\end{aligned}$$

q.e.d.

Beispiel 24.3
$X_1, \ldots, X_n$ seien unabhängig, $N(0,1)$-verteilt. Was ist die bedingte Verteilung von X_1 unter dem Mittelwert $\overline{X_n}$? Es gilt

$$(X_1 \setminus \overline{X_n}) = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}}_{A}, \quad \text{also } \begin{pmatrix} X_1 \\ \overline{X_n} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right),$$

mit $\Sigma = AI_nA^T = \begin{pmatrix} 1 & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} \end{pmatrix}$. Mit

$$(E,\mathfrak{E}) = (S,\mathfrak{S}) = (\mathbb{R},\mathfrak{B}),$$

$$f_{X_1,\overline{X_n}}(x,t) = \frac{1}{2\pi\sqrt{\det\Sigma}} \cdot \exp\left(-\frac{1}{2}(x,t)\Sigma^{-1}\begin{pmatrix} x \\ \theta \end{pmatrix}\right) \quad \text{(marginale Diche von } \overline{X_n} \sim N(0,\tfrac{1}{n})\text{)}$$

ist

$$f_{\overline{X_n}}(t) = \frac{1}{\sqrt{2\pi\frac{1}{n}}} \cdot \exp(-\frac{n}{2}t^2).$$

Es gilt $\Sigma^{-1} = \frac{n}{n-1}\begin{pmatrix} 1 & -1 \\ -1 & n \end{pmatrix}$, also erhält man im Exponenten des Quotienten

$$-\frac{1}{2}\left\{\frac{n}{n-1}(x,t)\begin{pmatrix} 1 & -1 \\ -1 & n \end{pmatrix}\begin{pmatrix} x \\ \theta \end{pmatrix} - nt^2\right\} = -\frac{1}{2}\frac{n}{n-1}(x-t)^2,$$

und damit

$$\frac{f_{X_1,\overline{X_n}}(x,t)}{f_{\overline{X_n}}(t)} = \frac{1}{\sqrt{2\pi\frac{n-1}{n}}} \cdot \exp\left(-\frac{1}{2\frac{n-1}{1}}(x-t)^2\right)$$

Dies ist die Dichte zu $N(t, \frac{n-1}{n})$, also insgesamt: Die bedingte Verteilung von X_1 unter $\overline{X_n}$ ist die Normalverteilung mit Mittelwert $\overline{X_n}$ und Varianz $\frac{n-1}{n}$.

Im Falle $(E, \mathfrak{E}) = (\Omega, \mathfrak{A})$, $X = \mathrm{id}$ nennt man Q die bedingte P-Wahrscheinlichkeit unter T.

24.2 Datenreduktion durch Suffizienz

Definition 24.4
Es sei $\mathfrak{P} = \{P_\theta \mid \theta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^n, \mathfrak{B}^n)$, $(S, \mathfrak{S})$ ein meßbarer Raum und $T : \mathbb{R}^n \to S$ $(\mathfrak{B}^n, \mathfrak{S})$-meßbar. Dann heißt T *suffizient* für $\mathfrak{P}$, wenn es eine von θ unabhängige reguläre Version der bedingten P_θ-Wahrscheinlichkeit unter T gibt.

Satz 24.5 (Neyman-Kriterium für Suffizienz)
$\mathfrak{P}$ sei durch ein σ-endliches Maß μ dominiert (d.h. $\mu(N) = 0 \implies P_\theta(N) = 0\ \forall\theta \in \Theta$). Dann sind äquivalent:

(i) T ist suffizient.

(ii) $\exists h : \mathbb{R}^n \to \mathbb{R}$, $\forall\theta\exists g(\cdot|\theta) : S \to \mathbb{R}$ mit

$$f(x|\theta) = g(T(x)|\theta) \cdot h(x) \quad \forall\theta \in \Theta, x \in \mathbb{R}^n$$

Nachweis von (ii) $\implies$ (i) unter der Zusatzbedingung, daß h die μ-Dichte eines Wahrscheinlichkeitsmaßes P ist, d.h.

$$„P_\theta(dx) = g(T(x)|\theta)P(dx)“.$$

Sei Q die bedingte P-Wahrscheinlichkeit unter T. Ist $\varphi : \mathbb{R}^n \times S \to \mathbb{R}_+$ meßbar, so gilt nach dem „verallgemeinerten Fubini“

$$\int \varphi(x, T(x))\, dP = \iint \varphi(x,t) Q(t, dx) P^T\,(dt).$$

Außerdem ist $t \to g(t|\theta)$ eine P^T-Dichte von P_θ^T. Damit:

$$\begin{aligned} P_\theta(A \cap \{T \in B\}) &= \int_{A\cap\{T\in B\}} g(T(x)|\theta)P\,(dx) \\ &= \int_B Q(t,A) g(t|\theta) P^T\,(dt) \\ &= \int_B Q(t,A) P_\theta^T\,(dt) \end{aligned}$$

Also ist Q eine von θ unabhängige Version der bedingten P_θ-Wahrscheinlichkeit unter T.

In Worten: T ist genau dann suffizient, wenn es Dichten gibt, die sich so faktorisieren lassen, daß der parameterabhängige Teil nur über T von den Daten abhängt.

Beispiel 24.6

(i) Ist $X_1, \ldots, X_n$ eine Stichprobe aus $U(0, \theta)$ mit unbekanntem $\theta > 0$, so gilt mit $\mu = \ell^{\otimes n}$

$$\begin{aligned} f(x_1 \ldots, x_n | \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} 1_{[0,\theta]}(x_i) \\ &= \underbrace{\theta^{-n} 1_{(-\infty,\theta]}(\max_{1 \leq i \leq n} x_i)}_{g(T(x)|\theta) \text{ mit } T = \max} \underbrace{1_{[0,\infty)}(\min_{1 \leq i \leq n} x_i)}_{h(x)} \end{aligned}$$

also ist $T(x_1, \ldots, x_n) := \max\{x_1, \ldots, x_n\}$ eine suffiziente Statistik.

(ii) Bei einer Stichprobe aus $N(\mu, \sigma^2)$ erhält man analog

$$\begin{aligned} f(x_1, \ldots, x_n | \mu, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i, \mu)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{n}{2\sigma}^2 \mu^2\right) \end{aligned}$$

also ist $T(x_1, \ldots, x_n) := (\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i)$ eine suffiziente Statistik. Diese Eigenschaft bleibt unter bijektiven Transformationen erhalten, also ist auch $(\overline{X_n}, S_n^2)$ suffizient.

Hat man eine suffiziente Statistik, so kann man sich beim Testen und Schätzen auf Verfahren beschränken, die nur über diese Statistik von den Daten abhängen:

Satz 24.7

Sei T suffizient für $\mathfrak{P}$, φ ein Test mit Gütefunktion β. Dann existiert $g : S \to \mathbb{R}$ derart, daß $\psi : g \circ T$ ein Test mit derselben Gütefunktion ist.

Ist Q eine von θ unabhängige bedingte P_θ-Wahrscheinlichkeit unter T, so läßt sich g wählen als

$$g(t) := \int \varphi(x) Q(t, dx).$$

Satz 24.8 (Rao-Blackwell, vgl. Satz 7.24)

Sei T suffizient für $\mathfrak{P} = \{P_\theta \mid \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, $\hat{\theta}$ ein Schätzer für θ. Dann existiert $g : S \to \mathbb{R}$ derart, daß für $\tilde{\theta} := g \circ T$ gilt:

(i) $\hat{\theta}$ erwartungstreu $\Longrightarrow$ $\tilde{\theta}$ erwartungstreu

(ii) MSE $(\tilde{\theta}, \theta) \leq$ MSE $(\hat{\theta}, \theta) \quad \forall \theta \in \Theta$

Ist Q eine von θ unabhängige bedingte P_θ-Wahrscheinlichkeit unter T, so läßt sich g wählen als

$$g(t) = \int \hat{\theta}(x) Q(t, dx).$$

Anhang

A Literatur

[Bauer] Bauer

[Ba/Bi] Bauer, Billingsley: „Probability & Measure“

B Index