TECHNISCHE
UNIVERSITÄT
DRESDEN

Fakultät Mathematik Institut für Stochastik, Professur für Statistik

Mathematische Statistik

Prof. Dr. Dietmar Ferger

Wintersemester 2020/21

Mitschrift : Eric Kunze
E-Mail : eric.kunze@mailbox.tu-dresden.de

Inhaltsverzeichnis

Kapitel 1
MEDIAN

Viele Schätzer in der Statistik sind definiert als Minimal- oder Maximalstelle von bestimmten **Kriteriumsfunktionen**, z. B. der **Maximum-Likelihood-Schätzer (MLS)** oder **Minimum-Quadrat-Schätzer (MQS, KQS)** oder **Bayes-Schätzer**. Allgemein nennt man solche Schätzer **M-Schätzer**.

Beispiel 1.1 (Maximum-Likelyhood-Schätzer) Gegeben seien $X_1, ..., X_n$ iid. $\sim f_\theta$. Dann ist $\hat{\theta}_n$ die Maximumstelle von der Funktion $l\colon \theta \mapsto \sum_{i=0}^n \log f_\theta(X_i)$ (sogenannte Log-Likelyhood-Funktion). Dabei kommen alle θ aus einer möglichen Menge Θ in Frage.

Ziel: Untersuchung des asymptotischen Verhaltens ($n \to \infty$) von M-Schätzern über einen funktionalen Ansatz. Als Beispiel betrachten wir nun den Median

Sei $X\colon (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ eine reelle Zufallsvariable mit Verteilungsfunktion $F_X\colon \mathbb{R} \to [0,1]$, d. h.

$$F_X(x) := \mathbb{P}(X \le x) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \le x\}) = \mathbb{P}\left(X^{-1}((-\infty, x])\right),$$

also $X \sim F_X$. Definiere

$$\begin{aligned} Y(t) &:= \mathbb{E}(|X - t|) \qquad (1.0) \\ &= \int_\Omega |X(\omega) - t| \, \mathbb{P}(\mathrm{d}\omega) \\ &\overset{??}{=} \int_\mathbb{R} |x - t| \, (\mathbb{P} \circ X^{-1})(\mathrm{d}x) \\ &= \int_\mathbb{R} |x - t| \, F_X(\mathrm{d}x) \qquad \forall t \in \mathbb{R} \end{aligned}$$

und

$$m \in \operatorname*{arg\,min}_{t \in \mathbb{R}} Y(t) \qquad \text{(M)}$$

als (irgendeine) Minimalstelle der Funktion Y.

Definition 1.2 Hierbei heißt $m \in \mathbb{R}$ **Median** von der Zufallsgröße X.

Wir haben den **Korrespondenzsatz** genutzt, der besagt, dass das Bildmaß $\mathbb{P} \circ X^{-1}$ eindeutig von der Verteilungsfunktion F bestimmt ist. Das heißt also wir schreiben $F(\mathrm{d}x)$ und meinen $\mathbb{P} \circ X^{-1}(\mathrm{d}x)$

Notation: $F(m-) := F(m-0) := \lim_{t \uparrow m} F(t)$. Der Ausdruck ist wohldefiniert für Verteilungsfunktionen, denn diese sind rechtsseitig stetig und haben linksseitige Limiten.

In folgendem kleinen Lemma wollen wir die Menge aller Mediane charakterisieren.

Lemma 1.0 Sei $X \sim F_X =: F$ integrierbar und $m \in \mathbb{R}$. Dann sind äquivalent:

(1) $F(m-) \leq \frac{1}{2} \leq F(m)$

(2) $\mathbb{E}\left[|X-t|\right] \geq \mathbb{E}|X-m| \qquad \forall t \in \mathbb{R}$

(3) m ist Median

Beispiel 1.1 Für F, $a < b \in \mathbb{R}$ mit

$$F(x) \begin{cases} < \frac{1}{2}, & \text{für } x < a \\ = \frac{1}{2}, & \text{für } a \leq x < b \\ > \frac{1}{2}, & \text{für } x \geq b \end{cases}$$

ist die Menge der Mediane genau das Intervall $[a, b]$.

Beweis. **(1) ⇒ (2):** Sei $\mathbb{Q} := \mathbb{P} \circ X^{-1}$ das zu F gehörige Bildmaß. Setze

$$h(t) := \mathbb{E}\left[|X-t| - |X-m|\right] \overset{lin.}{=} Y(t) - Y(m)$$

Dann ist (2) äquivalent zu $h(t) \geq 0$ für alle $t \in \mathbb{R}$. Dies wollen wir nun zeigen.

Fall A: Sei $t < m$.

$$\begin{aligned} h(t) &\overset{\text{Trafo}}{=} \int_{\mathbb{R}} |x-t| - |x-m| \; F(\mathrm{d}x) \\ &= \int_{(-\infty,t]} \underbrace{|x-t| - |x-m|}_{=t-x-(m-x)=-(m-t)} \; F(\mathrm{d}x) + \int_{(t,m)} \underbrace{\underbrace{|x-t| - |x-m|}_{\underbrace{x-t}_{\geq 0} - \underbrace{(m-x)}_{\leq m-t}}}_{\geq -(m-t)} \; F(\mathrm{d}x) \\ &\quad + \int_{[m,\infty)} \underbrace{|x-t| - |x-m|}_{x-t-(x-m)=m-t} \; F(\mathrm{d}x) \\ &\geq -(m-t) \cdot \underbrace{Q((-\infty,t])}_{F(t)} + (-(m-t) \cdot F(m-) - F(t)) + (m-t) \cdot \underbrace{\mathbb{Q}([m,\infty))}_{1-\underbrace{\mathbb{Q}((-\infty,m))}_{F(m-)}} \\ &= -\underbrace{(m-t)}_{\geq 0} \cdot \big(\underbrace{1 - 2 \cdot F(m-)}_{\text{wegen 1. Ungl. in (1):}\geq 0} \big) \\ &\geq 0 \end{aligned}$$

Fall B: Sei $t > m$. Wir verfahren ganz ähnlich:

$$
\begin{aligned}
h(t) &= \int_{(-\infty,m]} |x-t| - |x-m|\, F(\mathrm{d}x) + \int_{(m,t]} |x-t| - |x-m|\, F(\mathrm{d}x) \\
&\quad + \int_{(t,\infty)} |x-t| - |x-m|\, F(\mathrm{d}x) \\
&= \int_{(-\infty,m]} t - x - (m-x) F(\mathrm{d}x) + \int_{(m,t]} t - x - (x-m) F(\mathrm{d}x) \\
&\quad + \int_{(t,\infty)} x - t - (x-m) F(\mathrm{d}x) \\
&\geq (t-m)F(m) - (t-m)(F(t) - F(m)) + (m-t)(1-F(t)) \\
&= (t-m)(F(m) - F(t) + F(m) - 1 + F(t)) \\
&= \underbrace{(t-m)}_{>0} \cdot (\underbrace{2 \cdot F(m) - 1}_{\text{wegen 2. Ungl. in (1)} \geq 0}) \\
&\geq 0
\end{aligned}
$$

Fall C: Für $t = m$ ist die Aussage trivial.

(2) $\Rightarrow$ (1): Nach Annahme ist $h(t) \geq 0$ für alle $t \in \mathbb{R}$.

Fall A: Sei $t < m$. Die obige Rechnung im Fall 1 bei $\Rightarrow$ zeigt

$$
\begin{aligned}
0 \leq h(t) &= -(m-t) \cdot F(t) + \int_t^m \underbrace{\underbrace{x}_{<m} - t - (m-x)}_{=2x-t-m \leq m-t} F(\mathrm{d}x) + (m-t) \cdot (1 - F(m-)) \\
&\leq -(m-t) \cdot \Big(F(t) - 1 \underbrace{+ F(m-) - F(m-)}_{=0} + F(t) \Big) \\
&= \underbrace{(m-t)}_{>0} \cdot (1 - 2 \cdot F(t)) \\
&\Rightarrow \forall t < m: \quad 0 \leq (m-t) \cdot (1 - 2 \cdot F(t)) \\
&\Rightarrow \forall t < m: \quad 0 \leq 1 - 2 \cdot F(t) \\
&\Rightarrow \forall t < m: \quad F(t) \leq \frac{1}{2} \\
&\overset{t \uparrow m}{\Rightarrow} F(m-) \leq \frac{1}{2} \qquad \text{(Def. linksseitiger Limes)}
\end{aligned}
$$

Fall B: Sei $t > m$. Siehe 2. Fall, analog:

$$
\begin{aligned}
0 \le h(t) &= (t-m) \cdot F(m) + \int_m^t \underbrace{t - x - (x-m)}_{=t+m-2x \le t-m} F(\mathrm{d}x) + (m-t) \cdot (1 - F(t)) \\
&\le (t-m) \cdot (F(m) + F(t) - F(m) - 1 + F(t)) \\
&= \underbrace{(t-m)}_{>0}(2F(t) - 1) \\
&\Rightarrow \forall t < m: \ \ 0 \le 2F(t) - 1 \\
&\Rightarrow \forall t < m: \ \ F(t) \ge \frac{1}{2} \\
&\overset{t \downarrow m}{\Rightarrow} F(m) \ge \frac{1}{2} \qquad \text{(Rechtsstetigkeit von } F\text{)}
\end{aligned}
$$

(2) $\Rightarrow$ (3): Die Aussage (2) ist offensichtlich äquivalent zur Definition des Medians. $\square$

Bemerkung 1.2

(1) Aussage Punkt (1) in Lemma 1.0 besagt, dass $\{m \in \mathbb{R} : m \text{ erfüllt } (\mathbf{??})\}$ die Menge aller Mediane von F ist. Der Median ist im Allgemeinen nicht eindeutig bestimmt.

(2) Im Allgemeinen gibt es mehrere Mediane. Üblicherweise *wählt* man $m := F^{-1}\left(\frac{1}{2}\right)$, wobei

$$F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \ge u\} \quad \forall u \in (0,1)$$

die **Quantilfuntion** oder auch **verallgemeinerte Inverse** ist. Für weiterführende Literatur siehe [27, Seite 20]. Da

$$F\left(F^{-1}(u)-\right) \le u \le F\left(F^{-1}(u)\right) \qquad \forall u \in (0,1),$$

erfüllt $m = F^{-1}\left(\frac{1}{2}\right)$ die Bedingung (1) in Lemma 1.0 und ist somit ein Median, nämlich der kleinste.

(3) Die obige Funktion (1.0)

$$Y : \mathbb{R} \to \mathbb{R} \text{ mit } Y(t) = \int_{\mathbb{R}} |x - t| \ \ F(\mathrm{d}x) \qquad \forall t \in \mathbb{R}$$

ist stetig[1], aber im Allgemeinen nicht differenzierbar, z. B. falls $F \sim X$ eine diskrete Zufallsvariable ist. In diesem Fall ist somit die Minimierung über Differentiation nicht möglich.

(4) Sei $\mu = \mathbb{E}(X)$ der Erwartungswert von X. Dann gilt (Übung):

$$\mu = \arg\min_{t \in \mathbb{R}} \mathbb{E}\left[(X-t)^2\right] = \arg\min_{t \in \mathbb{R}} \mathbb{E}\left[(X-t)^2 - X^2\right].$$

(Das zweite X^2 wird abgezogen, da das $\arg\min$ nicht davon betroffen ist und so die Bedin-

[1] zur Stetigkeit von Y: nutze Folgenkriterium + dominierte Konvergenz mit Majorante $|X| + |t|$

gung $\mathbb{E}\left[X^2\right] < \infty$ entfällt.) Begründung:

$$\begin{aligned}\mathbb{E}\left[(X-t)^2) - X^2\right] &= \mathbb{E}\left[X^2 - 2tX + t^2 - X^2\right] \\ &= \mathbb{E}\left[X^2\right] - 2t\mu + t^2 - \mathbb{E}\left[X^2\right] = t^2 - 2t\mu\end{aligned}$$

Minimiere nun diese quadratische Funktion und erhalte die gewünschte Resultat.

In der Statistik identifizieren wir oft stillschweigend Zufallsgrößen mit ihren Realisationen, also $X \leftrightsquigarrow x$ was sich formal $X(\omega_0) = x$ schreiben lässt.

Zur Schätzung von m seien $X_1, \dots, X_n \sim F$ iid Zufallsvariablen mit zugehöriger **empirischer Verteilungsfunktion**

$$F_n\colon \mathbb{R} \to [0,1] \quad \text{mit} \quad F_n(x) := \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{1}_{\{x_i \le x\}}(x) \qquad \forall x \in \mathbb{R}.$$

Tatsächlich ist F_n die Verteilungsfunktion zum **empirischen Maß**

$$Q_n\colon \mathcal{B}(\mathbb{R}) \to [0,1] \quad \text{mit} \quad Q_n(B) := \frac{1}{n} \cdot \sum_{i=1}^{n} \delta_{x_i}(B)$$

wobei das **Dirac-Maß** in t definiert ist als

$$\delta_t\colon \mathcal{B}(\mathbb{R}) \to [0,1] \quad \text{mit} \quad \delta_t(B) := \begin{cases} 0, & \text{falls } t \notin B \\ 1, & \text{falls } t \in B \end{cases} \qquad \forall B \in \mathcal{B}(\mathbb{R}) \ \ \forall t \in \mathbb{R}$$

Gemäß dem Satz von Gliwenko-Cantelli gilt:

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{n \to \infty} 0 \qquad \mathbb{P}\text{-fast sicher für alle Verteilungsfunktionen } F$$

Die empirische Verteilungsfunktion konvergiert also gegen die wahre Verteilungsfunktion.

Erinnerung. Für das Dirac-Maß $\delta_x\colon \mathcal{A} \to \mathbb{R}_+$ mit $\delta_x(A) := \mathbb{1}_A(x)$ gilt:

$$\int_{\mathbb{R}} f(t)\, \delta_x(\mathrm{d}t) = f(x) \qquad \forall x \in \mathbb{R} \tag{Dir}$$

Notation. Wir identifizieren eine Verteilung $\mathbb{P} \circ X^{-1}$ mit der zugehörigen Verteilungsfunktion F_X, also $\mathbb{P} \circ X^{-1} \leftrightarrow F_X$ und $F_X(\mathrm{d}x) := (\mathbb{P} \circ X^{-1}\, \mathrm{d}x)$.

Erinnerung. Das Lebesgue-Maß ist linear im Maß, d. h.:

$$\int_{\omega} f \mathrm{d}(a \cdot \mu + b \cdot \nu) = a \cdot \int_{\omega} f \mathrm{d}\mu + b \cdot \int_{\omega} f\, \mathrm{d}\nu \tag{Lin}$$

Vorlesung „Mathematische Statistik" bei Prof. Dr. Ferger im Wintersemester 2020/21

Kapitel 2
METRISCHE RÄUME

Sei $(\mathcal{S}, d)$ ein metrischer Raum.

Beispiel 2.1 (Supremums-Metrik)

$$\begin{aligned} \mathcal{S} = C([0,1]) &:= \{f\colon [0,1] \to \mathbb{R} : f \text{ stetig}\} \\ d(f,g) &:= \sup_{t\in[0,1]} |f(t) - g(t)| \qquad \forall f,g \in C([0,1]) \end{aligned}$$

Definition 2.2 (1) Für $x \in \mathcal{S}$ und $r > 0$ ist

$$B(x,r) := B_d(x,r) := \{y \in \mathcal{S} : d(x,y) < r\}$$

die **offene Kugel um Mittelpunkt x und Radius r.**

(2) Sei $A \subseteq \mathcal{S}$. Dann ist:

- $A^\circ \ldots$ das **Innere** von A
- $\overline{A} \ldots$ die **abgeschlossene Hülle** von A
- $\partial A := \overline{A} \cap \overline{A^C} = \overline{A} \setminus A^\circ \ldots$ der **Rand** von A
- $A^\mathsf{C} := \mathcal{S} \setminus A \ldots$ das **Komplement** von A

(3) Die durch d induzierte **Topologie** ist

$$\begin{aligned} \mathcal{G} := \mathcal{G}(\mathcal{S}) &:= \{G \subseteq \mathcal{S} : G \text{ ist offen bzgl. } d\} \\ &= \{G \subseteq \mathcal{S} : \forall x \in G : \exists r > 0 : B_d(x,r) \subseteq G\} \end{aligned}$$

und

$$\mathcal{F} := \mathcal{F}(\mathcal{S}) := \{F \subseteq \mathcal{S} : F \text{ ist abgeschlossen}\} = \left\{F : F = G^\mathsf{C}, G \in \mathcal{G}\right\}$$

ist die **Menge der abgeschlossenen Mengen.**

(4) Sei $\emptyset \neq A \subseteq \mathcal{S}$ und $x \in \mathcal{S}$. Dann ist $d(x,A) := \inf\{d(x,a) : a \in A\} \geq 0$ der **Abstand von x zu A.**

(5) Es ist

$$\begin{aligned} C(\mathcal{S}) &:= \{f\colon S \to \mathbb{R} : f \text{ stetig}\} \\ C^b(\mathcal{S}) &:= \{f \in C(\mathcal{S}) : f \text{ beschränkt}\} \\ \|f\| := \|f\|_\infty &:= \sup_{x\in\mathcal{S}} |f(x)| \end{aligned}$$

Lemma 2.3 Sei $A \neq \emptyset$.

(1) Es ist $x \in \overline{A} \quad \Leftrightarrow \quad d(x, A) = 0$.

(2) Für alle $x, y \in \mathcal{S}$ gilt $|d(x, A) - d(y, A)| \leq d(x, y)$.

(3) Die Abbildung $d(\cdot, A)\colon \mathcal{S} \to [0, \infty)$, $x \mapsto d(x, A)$ ist gleichmäßig stetig.

Beweis. (zu 1) **(⇒)** Sei $x \in \overline{A}$. Dann existiert für alle $\varepsilon > 0$ ein $a \in A$ mit $d(x, a) < \varepsilon$. Somit

$$0 \leq d(x, A) \leq d(x, a) < \varepsilon \quad \forall \varepsilon > 0 \overset{\varepsilon \to 0}{\Rightarrow} d(x, A) = 0$$

(⇐) Sei $d(x, A) = 0$. Dann folgt aus der Infimumseigenschaft:

$$\forall \varepsilon > 0 \ \exists a \in A: \quad 0 \leq d(x, a) \leq 0 + \varepsilon = \varepsilon \quad \Rightarrow \quad x \in \overline{A}$$

(zu 2) Seien $x, y \in \mathcal{S}$. Dann gilt $d(x, a) \leq d(x, y) + d(y, a)$ mit der Dreiecksungleichung für alle $a \in A$ und somit

$$d(x, A) \leq d(x, y) + d(y, A) \quad \Rightarrow \quad d(x, A) - d(y, A) \leq d(x, y)$$

Vertauschen von x und y liefert $d(y, A) - d(x, A) \leq d(y, x) = d(x, y)$ und daraus folgt die Behauptung.

(zu 3) Folgt aus Punkt (2) da die Funktion $d(\cdot, A)$ Lipschitz-stetig und damit gleichmäßig stetig ist. □

Satz 2.4 (Stetige Approximation von Indikatorfunktionen) Zu $A \subseteq \mathcal{S}$ und $\varepsilon > 0$ existiert eine gleichmäßig stetige Funktion

$$f\colon \mathcal{S} \to [0, 1] \quad \text{mit der Eigenschaft} \quad f(x) = \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{falls } d(x, A) \geq \varepsilon \end{cases}$$

Beweis. Setze

$$\varphi\colon \mathbb{R} \to [0, 1] \quad \text{mit} \quad \varphi(t) := \begin{cases} 1 & \text{falls } t \leq 0 \\ 1 - t & \text{falls } 0 < t < 1 \\ 0 & \text{falls } t \geq 1 \end{cases}$$

Dann ist φ gleichmäßig stetig auf $\mathbb{R}$. Sei

$$f(x) := \varphi\left(\frac{1}{\varepsilon} \cdot d(x, A)\right) \qquad \forall x \in \mathcal{S}$$

Dann hat dieses f die gewünschte Eigenschaft wegen Lemma 2.3. □

Definition 2.5 Ein metrischer Raum $(\mathcal{S}, d)$ heißt **separabel**

$$
\begin{aligned}
&:\Leftrightarrow \exists \text{ abzählbares } S_0 \subseteq \mathcal{S} : \mathcal{S} \subseteq \overline{S_0} \\
&\Leftrightarrow \exists \text{ abzählbares } S_0 \subseteq \mathcal{S} : \mathcal{S} = \overline{S_0} \\
&\Leftrightarrow \exists \text{ abzählbares } S_0 \subseteq \mathcal{S} : S_0 \text{ liegt dicht in } \mathcal{S}
\end{aligned}
$$

Beispiel 2.6 Der metrische Raum $C([0,1])$ mit Supremums-Metrik ist separabel.

Beweis. Die Menge $S_0 := \{P : P \text{ ist Polynom mit rationalen Koeffizienten}\}$ ist abzählbar. Aus dem *Approximationssatz von Weierstraß* und der Dichtheit von $\mathbb{Q}$ folgt die Behauptung. □

Definition 2.7 $\mathcal{G}_0 \subseteq \mathcal{G}$ heißt **Basis** von $\mathcal{G}$ genau dann, wenn sich jedes $G \in \mathcal{G}$ als Vereinigung von Mengen aus $\mathcal{G}_0$, so genannte **$\mathcal{G}_0$-Mengen**, darstellen lässt.

Beispiel 2.8 Die Menge $\{B(x,r) : x \in \mathcal{S}, 0 < r \in \mathbb{Q}\}$ ist Basis von $\mathcal{G}$.

Beweis. Sei $G \in \mathcal{G}$. Dann gilt:

$$
\forall x \in G \;\; \exists\, 0 < r_x \in \mathbb{Q} : B(x, r_x) \subseteq G \quad \Rightarrow \quad G = \bigcup_{x \in G} \{x\} \subseteq \bigcup_{x \in G} \underbrace{B(x, r_x)}_{\subseteq G} \subseteq G
$$

Damit gilt also Gleichheit. □

Satz 2.9 $\mathcal{S}$ separabel $\Leftrightarrow$ $\mathcal{G}$ hat eine abzählbare Basis

Beweis. (⇒) Sei $S_0 \subseteq \mathcal{S}$ abzählbar und dicht in $\mathcal{S}$. Wir zeigen, dass

$$
\mathcal{G}_0 := \{B(x,r) : x \in S_0, 0 < r \in \mathbb{Q}\} \subseteq \mathcal{G}
$$

eine Basis ist. Sei also G offen. Dann folgt aus Beispiel Beispiel 2.8, dass

$$
G = \bigcup_{x \in G} B(x, r_x), \qquad 0 < r_x \in \mathbb{Q} \;\; \forall x \in G \tag{$\star$}
$$

Da $\overline{S_0} = \mathcal{S}$ gilt:

$$
\begin{aligned}
& \forall x \in G : \exists y_x \in S_0 : d(x, y_x) < \frac{r_x}{2} \\
\Rightarrow \quad & d(x,y) \overset{\triangle-\text{Ungl.}}{\leq} d(x, y_x) + d(y_x, x) < \frac{r_x}{2} + \frac{r_x}{2} = r_x \qquad \forall y \in B\left(y_x, \frac{r_x}{2}\right) \\
\Rightarrow \quad & B\left(y_x, \frac{r_x}{2}\right) \subseteq B(x, r_x) \qquad \forall x \in G \\
\Rightarrow \quad & G \overset{(\star)}{\supseteq} \bigcup_{x \in G} \underbrace{B\left(y_x, \frac{r_x}{2}\right)}_{\supseteq \{x\}} \supseteq \bigcup_{x \in G} \{x\} = G
\end{aligned}
$$

Damit gilt also Gleichheit und $\mathcal{G}_0$ ist eine Basis. Da S_0 abzählbar ist, ist $\mathcal{G}_0$ abzählbar.

($\Leftarrow$) Sei $\mathcal{G}_0$ abzählbare Basis von $\mathcal{G}$ und sei oBdA $\emptyset \notin \mathcal{G}_0$. Wähle für jedes $G \in \mathcal{G}_0$ ein $x_G \in G$ fest aus. Setze $S_0 := \{x_G : G \in \mathcal{G}_0\}$. Dann ist S_0 auch abzählbar. Es verbleibt die Dichtheit zu zeigen. Sei $x \in \mathcal{S}$ und $\varepsilon > 0$ beliebig. Da $B(x,\varepsilon)$ offen und $\mathcal{G}_0$ eine Basis ist, gilt:

$$\exists\, \mathcal{G}_{x,\varepsilon} \subseteq \mathcal{G}_0 \text{ mit } B(x,\varepsilon) = \bigcup_{G \in \mathcal{G}_{x,\varepsilon}} G \quad \Rightarrow \quad G \subseteq B(x,\varepsilon) \qquad \forall G \in \mathcal{G}_{x,\varepsilon}$$

Wähle ein G von diesen aus. Dann gilt:

$$x_G \in G \subseteq B(x,\varepsilon) \quad \Rightarrow \quad x_G \in B(x,\varepsilon) \quad \Rightarrow \quad d(\underbrace{x_G}_{\in S_0}, x) < \varepsilon$$

□

Satz 2.10 Seien $(\mathcal{S}, d)$ und $(\mathcal{S}', d')$ metrische Räume.

(1) Auf $\mathcal{S} \times \mathcal{S}'$ sind Metriken definiert durch

$$\begin{aligned}
d_1\left((x,x'),(y,y')\right) &:= \left((d(x,y))^2 + (d'(x',y'))^2\right)^{\frac{1}{2}} && \forall (x,x'),(y,y') \in \mathcal{S} \times \mathcal{S}' \\
d_2\left((x,x'),(y,y')\right) &:= \max\left\{d(x,y), d'(x',y')\right\} && \forall (x,x'),(y,y') \in \mathcal{S} \times \mathcal{S}' \\
d_3\left((x,x'),(y,y')\right) &:= d(x,y) + d'(x',y') && \forall (x,x'),(y,y') \in \mathcal{S} \times \mathcal{S}'
\end{aligned}$$

(2) Die Metriken d_1, d_2 und d_3 induzieren dieselbe Topologie $\mathcal{G}(\mathcal{S} \times \mathcal{S}')$ auf $\mathcal{S} \times \mathcal{S}'$, die sogenannte **Produkttopologie** von $\mathcal{G}(\mathcal{S})$ und $\mathcal{G}(\mathcal{S}')$.

(3) $\mathcal{G}(\mathcal{S} \times \mathcal{S}') = \left\{\bigcup_{\substack{G \in \mathcal{O} \\ G' \in \mathcal{O}'}} G \times G' : \mathcal{O} \subseteq \mathcal{G}(\mathcal{S}), \mathcal{O}' \subseteq \mathcal{G}(\mathcal{S}')\right\}$, d.h.

$$\{G \times G' : G \in \mathcal{G}(\mathcal{S}), G' \in \mathcal{G}(\mathcal{S}')\} \qquad \text{(Menge offener Rechtecke)}$$

bildet eine Basis von $\mathcal{G}(S \times \mathcal{S}')$.

Beweis. (zu 1) Überprüfung der Eigenschaften einer Metrik (zur Übung).

(zu 2) Punktweise gelten die Beziehungen:

$$d_2 \le d_1 \le \sqrt{2} \cdot d_2, \qquad \frac{1}{\sqrt{2}} \cdot d_3 \le d_1 \le d_3, \qquad d_2 \le d_3 \le 2 \cdot d_2$$

Beachte beim Nachweis, dass die d_i's als Metriken größer oder gleich Null sind. Aus obigen Beziehungen folgt u. a.:

$$B_{d_2}\left(x, \frac{r}{\sqrt{2}}\right) \subseteq B_{d_1}(x,r)$$

denn

$$r > \sqrt{2} \cdot d_2(y,x) \ge d_1(y,x) \tag{2.1}$$

Damit folgt aus G d_1-offen schon, dass G d_2-offen ist, denn für $x \in G$ existiert $r > 0$ mit $B_{d_1}(x,r) \subseteq G$ und (2.1) liefert $B_{d_2}(x, \frac{r}{\sqrt{2}}) \subseteq G$. Damit ist x auch ein ein d_2-innerer Punkt. Die anderen Relationen gelten analog.

(zu 3) $\subseteq$**:** Sei $G^* \in \mathcal{G}(\mathcal{S} \times \mathcal{S}')$ eine offene Menge in der Produkttopologie. Dann gilt

$$\forall x^* = (x, x') \in G^* : \exists r = r_{x^*} > 0 : \quad G^* = \bigcup_{x^* \in G^*} B\left(x^*, r_{x^*}\right).$$

Wegen Punkt (2) sei oBdA. $\mathcal{S}^* := \mathcal{S} \times \mathcal{S}'$ versehen mit der Metrik d_2. Dann gilt:

$$\begin{aligned} B_{d_2}\left(x^*, r_{x^*}\right) &= \left\{(y, y') \in \mathcal{S} \times \mathcal{S}' : \max\left\{d(x, y), d'(x', y')\right\} < r_{x^*}\right\} \\ &= \left\{(y, y') \in \mathcal{S} \times \mathcal{S}' : d(x, y) < r_{x^*} \wedge d'(x', y') < r_{x^*}\right\} \\ &= \underbrace{B_d\left(x, r_{x^*}\right)}_{\in \mathcal{G}(\mathcal{S})} \times \underbrace{B_{d'}\left(x', r_{x^*}\right)}_{\in \mathcal{G}(\mathcal{S}')} \end{aligned}$$

$\supseteq$**:** Sei zunächst $G \times G'$ G, G' offen und $x^* = (x, x') \in G \times G'$. Also ist $x \in G$ und $x' \in G'$ und somit

$$\exists\, r, r' > 0 : B_d(x, r) \subseteq G \wedge B_{d'}(x', r') \subseteq G'$$

Setze $r^* := \min\{r, r'\} > 0$. Damit folgt

$$\begin{aligned} B_{d_2}\left(x^*, r^*\right) &\subseteq B_d(x, r) \times B_{d'}\left(x', r'\right) \\ &\subseteq G \times G' = G^* \\ &\Rightarrow G \times G' \in \mathcal{G}(\mathcal{S} \times \mathcal{S}') \\ &\Rightarrow \bigcup_{\substack{G \in \mathcal{O} \\ G' \in \mathcal{O}'}} G \times G' \subseteq \mathcal{G}(\mathcal{S} \times \mathcal{S}') \qquad \forall \mathcal{O} \subseteq \mathcal{G}(\mathcal{S}), \mathcal{O}' \subseteq \mathcal{G}(\mathcal{S}') \end{aligned}$$

da die Produkttopologie vereinigungsstabil ist. □

Definition 2.11 Die Metriken d_1, d_2 und d_3 heißen **Produktmetriken**. Daher alternative Schreibweise $d \times d'$, also z. B. $d \times d' := \max\{d, d'\}$ usw.

Bemerkung 2.12 Analog zu obigen Definitonen lassen sich Produktmetriken für endlich viele metrische Räume $(\mathcal{S}_i, d_i)_{i \in \{1, \ldots, k\}}$ definieren, z. B.

$$d_1 \times \cdots \times d_k := \left(\sum_{i=1}^{k} d_i^2\right)^{\frac{1}{2}},$$

die wiederum dieselbe Produkttopologie induzieren. Die bisherigen Resultate gelten analog.

Kapitel 3
ZUFALLSVARIABLEN IN METRISCHEN RÄUMEN

3.1 Messbarkeit in metrischen Räumen

Definition 3.1 Die durch den metrischen Raum $(\mathcal{S}, d)$ induzierte σ-Algebra

$$\mathcal{B} := \mathcal{B}(\mathcal{S}) := \mathcal{B}_d(\mathcal{S}) := \sigma\left((\mathcal{G}(\mathcal{S})\right) := \sigma(\mathcal{G}).$$

heißt **Borel-σ-Algebra**. Elemente $B \in \mathcal{B}(\mathcal{S})$ heißen **Borel-Mengen** in $\mathcal{S}$.

Beachte: $\mathcal{B}(\mathcal{S}) = \mathcal{B}_d(\mathcal{S})$ hängt im Allgemeinen von der Metrik d ab.

Lemma 3.2 Es gilt:

(1) $\mathcal{B}(\mathcal{S}) = \sigma\left(\mathcal{F}(\mathcal{S})\right)$

(2) Ist $f\colon (\mathcal{S}, d) \to (\mathcal{S}', d)$ stetig, so ist f auch $\mathcal{B}_d(\mathcal{S})$-$\mathcal{B}_d(\mathcal{S}')$-messbar.

(3) Sei $\mathcal{G}_0$ abzählbare Basis von $\mathcal{G}(\mathcal{S})$. Dann gilt $\sigma(\mathcal{G}_0) = \mathcal{B}(\mathcal{S})$.

Beweis. (zu a) $\subseteq$: Sei $G \in \mathcal{G}(\mathcal{S})$. Dann gilt ist $G^{\mathsf{C}} \in \mathcal{F}(\mathcal{S}) \subseteq \sigma\left(\mathcal{F}(\mathcal{S})\right)$. Da $\sigma\left(\mathcal{F}(\mathcal{S})\right)$ stabil unter Bildung von Komplementen ist, ist auch $G = \left(G^{\mathsf{C}}\right)^{\mathsf{C}} \in \sigma(\mathcal{F}(\mathcal{S}))$. Weiter folgt aus $\mathcal{G} \subseteq \sigma(\mathcal{F})$ schon $\sigma(\mathcal{G}) \subseteq \sigma(\mathcal{F})$.

$\supseteq$: analog

(zu b) Per Definition gilt $f^{-1}\left(\mathcal{B}_{d'}(\mathcal{S}')\right) = f^{-1}\left(\sigma\left(\mathcal{G}(\mathcal{S}')\right)\right)$. Ein Satz aus der Maßtheorie liefert uns weiter $f^{-1}\left(\sigma\left(\mathcal{G}(\mathcal{S}')\right)\right) = \sigma\left(f^{-1}\left(\mathcal{G}(\mathcal{S}')\right)\right)$. Bekannterweise sind für stetige Funktionen Urbilder offener Mengen wieder offen, d.h. es gilt $f^{-1}\left(\mathcal{G}(\mathcal{S}')\right) \subseteq \mathcal{G}(S)$. Somit ist $\sigma\left(f^{-1}\left(\mathcal{G}(\mathcal{S}')\right)\right) \subseteq \sigma\left(\mathcal{G}(\mathcal{S})\right) = \mathcal{B}(\mathcal{S})$.

(zu c) $\subseteq$: klar wegen $\mathcal{G}_0 \subseteq \mathcal{G}(\mathcal{S})$ und σ monoton

$\supseteq$: Sei $G \in \mathcal{G}$. Dann existieren geeignete $G_i \in \mathcal{G}_0 \subseteq \sigma(\mathcal{G}_0)$, sodass $G = \bigcup_{i\in\mathbb{N}} G_i$. Damit ist $G \in \sigma(\mathcal{G}_0)$. Aus der Stabilität unter Vereinigungen folgt die Behauptung:

$$G = \bigcup_{i\in\mathbb{N}} G_i \subseteq \sigma\left(\mathcal{G}_0\right) = \mathcal{B}(\mathcal{S}) \in \sigma(\mathcal{G}_0),$$

da σ-Algebren *abzählbar* vereinigungsstabil sind. $\square$

Satz 3.3 Sei $(\mathcal{S}, d)$ separabler metrischer Raum. Dann gilt:

$$\mathcal{B}_{d\times d}(\mathcal{S}\times\mathcal{S}) = \mathcal{B}(S)\otimes\mathcal{B}(\mathcal{S})$$

Ohne Separabilität gilt nur „$\supseteq$". Für S_1 und S_2 separabel gilt $\mathcal{B}_{d_1\times d_2}(\mathcal{S}_1\times\mathcal{S}_2) = \mathcal{B}(S_1)\otimes\mathcal{B}(\mathcal{S}_2)$ analog, auch erweiterbar auf endliche Produkte.

Beweis. Seien

$$\begin{aligned}
\pi_1\colon \mathcal{S}\times\mathcal{S}\to\mathcal{S} \text{ mit } \pi_1(x,y) &:= x \qquad \forall (x,y)\in\mathcal{S}\times\mathcal{S}\\
\pi_2\colon \mathcal{S}\times\mathcal{S}\to\mathcal{S} \text{ mit } \pi_2(x,y) &:= y \qquad \forall (x,y)\in\mathcal{S}\times\mathcal{S}
\end{aligned}$$

die **Projektionsabbildungen**. Dann gilt

$$\begin{aligned}
\mathcal{B}(\mathcal{S})\otimes\mathcal{B}(\mathcal{S}) &= \sigma(\pi_1,\pi_2) = \sigma\left(\pi_1^{-1}(\sigma(\mathcal{G}))\cup\pi_2^{-1}(\sigma(\mathcal{G}))\right) && \text{(Definition)}\\
&= \sigma\left(\sigma\left(\pi_1^{-1}(\mathcal{G})\right)\cup\sigma\left(\pi_2^{-1}(\mathcal{G})\right)\right)\\
&\overset{(\star)}{=} \sigma\left(\pi_1^{-1}(\mathcal{G})\cup\pi_2^{-1}(\mathcal{G})\right)\\
&= \sigma\left(\{G\times S, S\times G' : G, G'\in\mathcal{G}\}\right)\\
&= \sigma\left(\left\{\overbrace{G\times G'}^{=(G\times S)\cap(S\times G')} : G, G'\in\mathcal{G}\right\}\right)\\
& && \text{(„}\subseteq\text{", da } S\in\mathcal{G}\text{ ; „}\supseteq\text{", da }\sigma\text{-Algebra }\cap\text{-stabil)}\\
&\overset{(\star\star)}{=} \sigma\left(\left\{\bigcup_{\substack{G\in\mathcal{O}\\ G'\in\mathcal{O}'}} G\times G' : \mathcal{O},\mathcal{O}'\subseteq\mathcal{G}\right\}\right)\\
&= \sigma\left(\mathcal{G}(\mathcal{S}\times\mathcal{S})\right) && \text{(Satz 2.10, Punkt (3))}\\
&\overset{\text{Def}}{=} \mathcal{B}(\mathcal{S}\times\mathcal{S})
\end{aligned}$$

Zum Nachweis von $(\star)$:

$\supseteq$: Setze $\mathcal{E} := \underbrace{\sigma\left(\pi_1^{-1}(\mathcal{G})\right)}_{\supseteq\pi_1^{-1}(\mathcal{G})}\cup\underbrace{\sigma\left(\pi_2^{-1}(\mathcal{G})\right)}_{\supseteq\pi_2^{-1}(\mathcal{G})} \supseteq \pi_1^{-1}(\mathcal{G})\cup\pi_2^{-1}(\mathcal{G}) =: \mathcal{H}$. Dann ist $\sigma(\mathcal{E})\supseteq\sigma(\mathcal{H})$.

$\subseteq$: Es ist $\pi_1^{-1}(\mathcal{G}) \subseteq \left(\pi_1^{-1}(\mathcal{G})\cup\pi_2^{-1}(\mathcal{G})\right) = \mathcal{H}$. Also ist auch $\sigma\left(\pi_1^{-1}(\mathcal{G})\right)\subseteq\sigma(\mathcal{H})$. Analog erhalten wir auch $\sigma\left(\pi_2^{-1}(\mathcal{G})\right)\subseteq\sigma(\mathcal{H})$. Dann gilt

$$\mathcal{E} = \underbrace{\sigma\left(\pi_1^{-1}(\mathcal{G})\right)}_{\subseteq\sigma(\mathcal{H})}\cup\underbrace{\sigma\left(\pi_2^{-1}(\mathcal{G})\right)}_{\subseteq\sigma(\mathcal{H})}\subseteq\sigma(\mathcal{H})\ ,$$

also auch $\sigma(\mathcal{E})\subseteq\sigma(\mathcal{H})$.

Bleibt Nachweis von $(\star\star)$:

$\subseteq$: ist klar mit $\mathcal{O}$ und $\mathcal{O}'$ einelementig (gilt auch ohne Separabilität)

$\supseteq$: Gemäß Satz 2.9 existiert abzählbare Basis $\mathcal{G}_0$ von $\mathcal{G}$. Seien $\mathcal{O}, \mathcal{O}\cdot \subseteq \mathcal{G}$. Sei

$$G^* = \bigcup_{\substack{G \in \mathcal{O} \\ G' \in \mathcal{O}'}} G \times G' = \bigcup_{\substack{G, G' \text{ offen} \\ G \times G' \subseteq G^*}} G \times G' \overset{(!)}{=} \bigcup_{\substack{G_0, G_0' \in \mathcal{G}_0 \\ G \times G_0' \subseteq G^*}} G_0 \times G_0'$$

eine abzählbare Vereinigung, da $\mathcal{G}_0$ Basis ist, also abzählbar. Somit gilt dann $G^* \in \sigma(\{G \times G' : G, G' \in \mathcal{G}\})$. $\square$

Definition 3.4 Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine Abbildung $X \colon \Omega \to \mathcal{S}$, die $\mathcal{A}$-$\mathcal{B}(\mathcal{S})$-messbar ist, heißt **Zufallsvariable** (ZV) in dem metrischen Raum $(\mathcal{S}, d)$ über $(\Omega, \mathcal{A})$.

Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$, also $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Das Bildmaß

$$\mathbb{P} \circ X^{-1} := \mathbb{P}_X := \mathcal{L}(X) := \mathcal{L}(X \mid \mathbb{P})$$
$$\left(\mathbb{P} \circ X^{-1}\right)(B) := \mathbb{P}\left(X^{-1}(B)\right) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in B\}\right) =: \mathbb{P}(X \in B) \qquad \forall B \in \mathcal{B}(\mathcal{S})$$

heißt **Verteilung** von X unter $\mathbb{P}$.

Satz 3.5 Sei $(\mathcal{S}, d)$ ein separabler metrischer Raum und seien X, Y Zufallsvariablen in $(\mathcal{S}, d)$ über $(\Omega, \mathcal{A})$. Dann ist $d(X, Y)$ eine reelle Zufallsvariable.

Beweis (gute Prüfungsfrage). Die Abbildungen $X, Y \colon (\Omega, \mathcal{A}) \to (\mathcal{S}, \mathcal{B}(\mathcal{S}))$ sind messbar genau dann, wenn die Abbildung $(X, Y) \colon (\Omega, \mathcal{A}) \to \Big(\mathcal{S} \times \mathcal{S}, \underbrace{\mathcal{B}(\mathcal{S}) \otimes \mathcal{B}(\mathcal{S})}_{=\mathcal{B}(\mathcal{S} \times \mathcal{S})}\Big)$ messbar ist. Jede Metrik ist bekanntlich stetig, also auch $d \colon (\mathcal{S} \times \mathcal{S}, \mathcal{G}(\mathcal{S} \times \mathcal{S})) \to \mathbb{R}^1$. Dann folgt aus Lemma 3.2, dass $d \colon \mathcal{B}(\mathcal{S} \times \mathcal{S}) \to \mathcal{B}(\mathbb{R})$ messbar ist. Damit folgt die Behauptung, denn $d(X, Y) = d \circ (X, Y)$ ist messbar als Komposition von messbaren Abbildungen. $\square$

3.2 Fast sichere Konvergenz

Definition 3.6 Seien X, X_n $(n \in \mathbb{N})$ Zufallsvariablen in einem separablen, metrischen Raum $(\mathcal{S}, d)$ über $(\Omega, \mathcal{A}, \mathbb{P})$.

$$X_n \overset{n\to\infty}{\longrightarrow} X \quad \mathbb{P}\text{-fast sicher} \quad :\Leftrightarrow \quad \mathbb{P}\Bigg(\underbrace{\left\{\omega \in \Omega : d\left(X_n(\omega), X(\omega)\right) \overset{n\to\infty}{\longrightarrow} 0\right\}}_{=:M}\Bigg) = 1$$

Beachte: Die Definition von M mengentheoretisch aufgeschrieben (Schnitt $\rightsquigarrow$ „für alle"; Verei-

[1] $\mathbb{R}$ sei mit der natürlichen Topologie versehen

nigung $\rightsquigarrow$ „Es gibt“):

$$M = \bigcap_{0<\varepsilon\in\mathbb{Q}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \Big\{ \underbrace{d(X_n, X)}_{=:\zeta_n} < \varepsilon \Big\} \overset{\text{Satz 3.5}}{\in} \mathcal{A}, \text{ denn } \zeta_n^{-1}\left((-\infty, \varepsilon)\right) \in \mathcal{A}$$

Hierbei gilt $M \in \mathcal{A}$ aufgrund der Abzählbarkeit der beteiligten Schnitte und Vereinigungen. Man schreibt auch:

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1 \quad \text{oder} \quad d(X_n, X) \to 0 \ \ (n \to \infty) \ \ \mathbb{P}\text{-f. s.}$$

Die bekannten Regeln (Ergebnisse) für *reelle* Zufallsvariablen lassen sich mühelos verallgemeinern; so zum Beispiel der folgende Satz.

Satz 3.7 Der fast-sichere Grenzwert ist $\mathbb{P}$-fast sicher eindeutig:

$$X_n \overset{n\to\infty}{\longrightarrow} X \quad \mathbb{P}\text{-fast sicher und } X_n \overset{n\to\infty}{\longrightarrow} X' \quad \mathbb{P}\text{-fast sicher} \quad \Rightarrow \quad X = X' \quad \mathbb{P}\text{-fast sicher}$$

Beweis. Es ist $\{X \neq X'\} \subseteq \{X_n \not\to X\} \cup \{X_n \not\to X'\}$ und $\mathbb{P}(X_n \not\to X) + \mathbb{P}(X_n \not\to X') = 0 + 0$. Also ist auch $\mathbb{P}(X \neq X') = 0$ und mit dem Gegenereignis gilt $\mathbb{P}\left(X = X'\right) = 1 - \mathbb{P}\left(X \neq X'\right) = 1$. □

Satz 3.8 Seien X, X_n $(n \in \mathbb{N})$ Zufallsvariablen im separablen, metrischen Raum $(\mathcal{S}, d)$ und sei $f\colon (\mathcal{S}, d) \to (\mathcal{S}', d')$ $\mathcal{B}_d(S)$ - $\mathcal{B}_{d'}(S')$-messbar und stetig in X $\mathbb{P}$-fast sicher. [a] Dann gilt:

$$X_n \overset{n\to\infty}{\longrightarrow} X \quad \mathbb{P}\text{-fast sicher} \quad \Rightarrow \quad f(X_n) \overset{n\to\infty}{\longrightarrow} f(X) \quad \mathbb{P}\text{-fast sicher}$$

[a] „stetig in X $\mathbb{P}$-fast sicher“ bedeutet, dass $\mathbb{P}(\{\omega \in \Omega : f \text{ in } X(\omega) \text{ stetig}\}) = 1$. Insbesondere ist $\{\omega \in \Omega : f \text{ in } X(\omega) \text{ stetig}\}$ messbar, was keine Selbstverständlichkeit ist.

Beweis. Aufgrund der Folgenstetigkeit gilt

$$\left\{X_n \overset{n\to\infty}{\longrightarrow} X\right\} \cap \{f \text{ stetig in } X\} \subseteq \left\{f(X_n) \overset{n\to\infty}{\longrightarrow} f(X)\right\}$$

Da abzählbare Schnitte von Einsmengen stets Einsmengen sind, d. h.

$$\forall i \in \mathbb{N} : \mathbb{P}(E_i) = 1 \quad \Rightarrow \quad \mathbb{P}\left(\bigcap_{i\in\mathbb{N}} E_i\right) = 1 \qquad \forall \{E_i\} \subseteq \Omega \text{ mit } \mathbb{P}(E_i) = 1$$

folgt

$$1 = \mathbb{P}\left(\left\{X_n \overset{n\to\infty}{\longrightarrow} X \text{ und } f \text{ stetig in } X\right\}\right) \leq \mathbb{P}\left(\left\{f(X_n) \overset{n\to\infty}{\longrightarrow} f(X)\right\}\right) \leq 1$$

und somit schon $\mathbb{P}\left(\left\{f(X_n) \overset{n\to\infty}{\longrightarrow} f(X)\right\}\right) = 1$. □

Satz 3.9 (Konvergenz-Kriterium)

$$X_n \overset{n\to\infty}{\longrightarrow} X \quad \mathbb{P}\text{-fast sicher} \quad \Leftrightarrow \quad \forall \varepsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left(\sup_{m\geq n} d(X_m, X) > \varepsilon\right) = 0$$

Beweis. Man ersetze im Beweis für den Fall reeller Zufallsvariablen $|X_n - X|$ durch $d(X_n, X)$. Und beachte, dass alle Schlussfolgerungen bestehen bleiben. Hier ausführlich:

$$
\begin{aligned}
X_n \overset{n\to\infty}{\longrightarrow} X \ \mathbb{P}\text{-fast sicher} \quad &\Leftrightarrow \quad 1 = \mathbb{P}\left(\bigcup_{0<\varepsilon} \bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} d(X_n, X) < \varepsilon\right) \\
&\Leftrightarrow \quad 0 = \mathbb{P}\left(\bigcap_{0<\varepsilon} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} d(X_n, X) \geq \varepsilon\right) \\
&\Leftrightarrow \quad 0 = \mathbb{P}\left(\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} d(X_n, X) \geq \varepsilon\right) \qquad \forall \varepsilon > 0
\end{aligned}
$$

Da für $m \to \infty$ die Mengen $\left\{\bigcap_{n\geq m} d(X_n, X) \geq \varepsilon\right\}$ kleiner werden, ist dies äquivalent zu

$$
\begin{aligned}
&\forall \varepsilon > 0 \ \forall \delta > 0 \ \exists m \in \mathbb{N} : \quad \mathbb{P}\left(\bigcap_{n\geq m} d(X_n, X) \geq \varepsilon\right) < \delta \\
&\Leftrightarrow \quad \forall \varepsilon > 0 : \quad \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n\geq m} d(X_n, X) \geq \varepsilon\right) = 0 \\
&\Leftrightarrow \quad \forall \varepsilon > 0 : \quad \lim_{m\to\infty} \mathbb{P}\left(\sup_{n\geq m} d(X_n, X) \geq \varepsilon\right) = 0
\end{aligned}
$$

□

Ein sehr nützliches Kriterium ist Folgendes:

Satz 3.10

$$
\forall \varepsilon > 0 : \quad \sum_{n\in\mathbb{N}_{\geq 1}} \mathbb{P}\left(d(X_n, X) > \varepsilon\right) < \infty \quad \Longrightarrow \quad X_n \overset{n\to\infty}{\longrightarrow} X \quad \mathbb{P}\text{-fast sicher}
$$

Bemerkung. Die $\mathbb{P}(d(X_n, X) > \varepsilon)$ werden in der Statistik **Fehlerwahrscheinlichkeit** oder **tail probability** genannt. Um Fehlerwahrscheinlichkeiten abzuschätzen, also die Voraussetzung für diesen Satz für einen speziellen Fall zu zeigen, nutzt man häufig sogenannte Maximalungleichungen wie die Markov-Ungleichung und die Tschebychew-Ungleichung.

Beweis. Setze $A_n(\varepsilon) := \{d(X_n, X) > \varepsilon\} \in \mathcal{A}$ wegen Satz 3.5. Dann folgt aus dem *ersten Borel-Cantelli-Lemma* $\mathbb{P}\left(\limsup_{n\to\infty} A_n(\varepsilon)\right) = 0$ für alle $\varepsilon > 0$. Mit

$$
\begin{aligned}
\liminf_{n\to\infty} \left(A_n(\varepsilon)^{\mathsf{C}}\right) \overset{\text{Def}}{=} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} (A_n(\varepsilon))^{\mathsf{C}} &= \left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \{d(X_n, X) > \varepsilon\}\right)^{\mathsf{C}} \\
&\overset{\text{Def}}{=} \left(\limsup_{n\to\infty} A_n(\varepsilon)\right)^{\mathsf{C}}
\end{aligned}
$$

folgt dann

$$
1 = \mathbb{P}\left(\left(\limsup_{n\to\infty} A_n(\varepsilon)\right)^{\mathsf{C}}\right) = \mathbb{P}\left(\liminf_{n\to\infty} \left(A_n(\varepsilon)^{\mathsf{C}}\right)\right) \qquad \forall \varepsilon > 0
$$

Da abzählbare Durchschnitte von Eins-Mengen (also Mengen mit $\mathbb{P}$-Maß 1) wieder Eins-Mengen sind, folgt schließlich:

$$\mathbb{P}\left(\underbrace{\bigcap_{0<\varepsilon\in\mathbb{Q}}\bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m}\{d(X_n,X)\leq\varepsilon\}}_{\{X_n\to X\}=\{d(X_n,X)\to 0\}}\right)=1 \qquad \square$$

Weitere Eigenschaften der fast sicheren Konvergenz von Zufallsvariablen in metrischen Räumen finden sich z. B. in [12, Kapitel 8.2].

3.3 Stochastische Konvergenz

Definition 3.11 X_n **konvergiert stochastisch** bzw. **in Wahrscheinlichkeit** gegen X, in Symbolen:

$$X_n \xrightarrow{\mathbb{P}} X \quad :\Leftrightarrow \quad \forall\varepsilon>0:\mathbb{P}\left(\{d(X_n,X)>\varepsilon\}\right)\overset{n\to\infty}{\longrightarrow} 0$$

d. h. $d(X_n,X)\xrightarrow{\mathbb{P}} 0$.

Satz 3.12 $X_n \overset{n\to\infty}{\longrightarrow} X \quad \mathbb{P}$-fast sicher $\quad\Rightarrow\quad X_n\xrightarrow{\mathbb{P}} X$

Beweis. Gemäß Satz 3.9 gilt

$$\forall\varepsilon>0:\quad 0\leq\mathbb{P}\left(d(X_n,X)>\varepsilon\right)\leq\mathbb{P}\left(\sup_{m\geq n} d(X_m,X)>\varepsilon\right)\overset{n\to\infty}{\longrightarrow} 0 \qquad \square$$

Die Umkehrung von Satz 3.12 gilt im Allgemeinen *nicht*, aber es gilt das folgende Teilfolgenkriterium.

Satz 3.13 (Teilfolgenkriterium für stochastische Konvergenz) Folgende Aussagen sind äquivalent:

(1) $X_n\xrightarrow{\mathbb{P}} X$

(2) Zu jeder Teilfolge (TF) $(X_{n'})$ von $(X_n)_{n\in\mathbb{N}}$ existiert eine Teilfolge $(X_{n''})$ von $(X_{n'})$ derart, dass $X_{n''}\overset{n''\to\infty}{\longrightarrow} X$ $\mathbb{P}$-fast sicher.

Die Notation $(X_{n''})$ stammt aus [2].

Beweis. Wir verfahren wie im Reellen: Es gelte Punkt (1). Sei $\varepsilon>0$ beliebig und $(X_{n_k})_k$ eine Teilfolge von $(X_n)_n$. Da $\mathbb{P}\left(\{d(X_n,X)>\varepsilon\}\right)\to 0$, können wir eine Teilteilfolge $(X_{n_{k_i}})_i := (Y_i)_i$ finden mit $\mathbb{P}\left(\{d(Y_i,X)>\varepsilon\}\right)<i^{-2}$. Dann ist für alle $n\in\mathbb{N}$ und mit $n\to\infty$

$$\mathbb{P}\left(\sup_{i\geq n} d(Y_i,X)<\varepsilon\right)=\mathbb{P}\left(\bigcup_{i\geq n}\{d(Y_i,X)<\varepsilon\}\right)\leq\sum_{i=n}^{\infty}\mathbb{P}\left(\{d(Y_i,X)<\varepsilon\}\right)\leq\sum_{i=n}^{\infty} i^{-2}\to 0$$

Laut dem Konvergenz-Kriterium in Satz 3.9 bedeutet das $Y_i\to X$ $\mathbb{P}$-fast sicher.

Gelte nun Punkt (2) und wir nehmen an, es gäbe ein $\varepsilon > 0$ und eine Teilfolge $(X_{n_k})_k$ von $(X_n)_n$, sodass $\mathbb{P}(\{d(X_{n_k}, X) > \varepsilon\}) > \delta > 0$ für ein passendes δ. Dann konvergiert nach *Punkt* (2) eine Teilteilfolge $(X_{n_{k_i}})_i$ $\mathbb{P}$-fast sicher gegen X, also laut Satz 3.12 auch stochastisch. Das widerspricht aber der obigen Annahme über $(X_{n_k})_k$, die sich auch auf $(X_{n_{k_i}})_i$ überträgt. □

Mit dem Teilfolgenkriterium lassen sich Rechenregeln für fast sichere Konvergenz auf stochastische Konvergenz übertragen.

Korollar 3.14 (a) $X_n \xrightarrow{\mathbb{P}} X$ und $X_n \xrightarrow{\mathbb{P}} X' \;\Rightarrow\; X = X'$ $\mathbb{P}$-fast sicher

(b) Gelte $X_n \xrightarrow{\mathbb{P}} X$ in $(\mathcal{S}, d)$ und $f\colon (\mathcal{S}, d) \to (\mathcal{S}', d')$ messbar mit f stetig in X $\mathbb{P}$-fast sicher. Dann gilt auch $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Beweis. (zu a) Wegen $X_n \xrightarrow{\mathbb{P}} X$ existiert nach Satz 3.13 eine Teilfolge $(X_{n'}) \subseteq (X_n)_{n\in\mathbb{N}}$ mit $X_{n'} \xrightarrow{n'\to\infty} X$ $\mathbb{P}$-fast sicher. Zu $(X_{n'})$ existiert (wegen $X_n \xrightarrow{\mathbb{P}} X'$ und Satz 3.13) eine Teilfolge $(X_{n''})$ von $(X_{n'})$, sodass aus $X_{n''} \xrightarrow{n\to\infty} X'$ $\mathbb{P}$-fast sicher mit Satz 3.7 schon $X = X'$ fast sicher folgt.

(zu b) Zu einer beliebigen Teilfolge $(X_{n'})$ von (X_n) existiert eine Teilfolge $(X_{n''})$ von $(X_{n'})$ mit $X_{n''} \to X$ $\mathbb{P}$-fast sicher. Mit Satz 3.8 folgt $f(X_{n''}) \to f(X)$ $\mathbb{P}$-fast sicher und mit Satz 3.13 angewendet auf $(f(X_n))_n$ gilt $f(X_n) \xrightarrow{\mathbb{P}} f(X)$. □

3.4 Konvergenz in Produkträumen

Seien $(\mathcal{S}, d)$ und $(\mathcal{S}', d')$ separable metrische Räume. Dann ist auch $(\mathcal{S}\times\mathcal{S}', d\times d')$ ein separabler metrischer Raum. Dies folgt z. B. aus dem *Satz von der koordinatenweise Konvergenz*:

$$(a_n, a_n') \xrightarrow{d\times d'} (a, a') \;\Leftrightarrow\; (a_n) \xrightarrow{d} a \text{ und } (a_n') \xrightarrow{d'} a' \tag{3.1}$$

Es gibt auch eine „stochastische Versionen“ dieses Satzes.

Satz 3.15 (Koordinatenweise Konvergenz vom Zufallsgrößen)

(1) $(X_n, X_n') \xrightarrow{d\times d'} (X, X')$ $\mathbb{P}$-f. s. $\Leftrightarrow$ $X_n \xrightarrow{d} X$ $\mathbb{P}$-f. s. und $X_n' \xrightarrow{d'} X'$ $\mathbb{P}$-f. s.

(2) $(X_n, X_n') \xrightarrow{\mathbb{P}} (X, X')$ $\Leftrightarrow$ $X_n \xrightarrow{\mathbb{P}} X$ und $X_n' \xrightarrow{\mathbb{P}} X'$

Beweis. (zu 1) Nach Gleichung (3.1) folgt aus $(X_n, X_n') \xrightarrow{d\times d'} (X, X')$ $\mathbb{P}$-fast sicher auch $X_n \to X$ und $X_n' \to X'$ $\mathbb{P}$-fast sicher. Da der Schnitt von Eins-Mengen wieder eine Eins-Menge ist, ist dies wiederum äquivalent zu $X_n \to X$ $\mathbb{P}$-fast sicher und $X_n' \to X'$ $\mathbb{P}$-fast sicher. *Achtung* – die „Stellung“ der Fast-Sicherheit spielt eine Rolle!

(zu 2) Die linke Seite ist wegen (3.1) und Satz 3.13 äquivalent dazu, dass für alle Teilfolgen $(X_{n'}, X_{n'}') \subseteq (X_n, X_n')$ eine Teilteilfolge $(X_{n''}, X_{n''}')$ existiert mit $(X_{n''}, X_{n''}') \xrightarrow{n\to\infty} 0$ fast sicher. Also gilt wegen Punkt (1)

$$X_{n''} \to X \text{ f. s. und } X'_{n''} \to X' \text{ f. s.}$$

Somit existiert für alle Teilfolgen $(X_{n'}) \subseteq (X_n)$ eine Teilteilfolge $(X_{n''}) \subseteq (X_{n'})$ mit $X_{n''} \to X$ fast sicher. Dies ist wegen (3.1) äquivalent zu $X_n \xrightarrow{\mathbb{P}} X$. Analog gilt auch $X'_n \to X'$. □

3.4.1 Gleichheit in Verteilung

Definition 3.16 Seien X_i $(i = 1, 2)$ Zufallsvariablen in $(\mathcal{S}, d)$ über $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$. Sie heißen **gleich in Verteilung**, in Zeichen

$$X_1 \overset{\mathrm{d}}{=} X_2 \quad :\Leftrightarrow \quad \mathbb{P}_1 \circ X_1^{-1} = \mathbb{P}_2 \circ X_2^{-1}$$

Wir wollen Verteilungsgleichheit in folgendem Satz charakterisieren.

Satz 3.17 Es gilt:

(a) Seien $\mathbb{P}, \mathbb{Q}$ Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathcal{S})$. Dann gilt:

$$\mathbb{P} \equiv \mathbb{Q} \quad \Leftrightarrow \quad \int f \ \mathrm{d}\mathbb{P} = \int f \ \mathrm{d}\mathbb{Q} \qquad \forall f \in C^b(\mathcal{S}) \text{ glm. stetig}$$

(b) Es gilt

$$X \overset{\mathrm{d}}{=} Y \quad \Leftrightarrow \quad \mathbb{E}\left[f(X)\right] = \int_{\Omega_1} f(X_1) \ \mathrm{d}\mathbb{P}_1 = \int_{\Omega_2} f(X_2) \ \mathrm{d}\mathbb{P}_2 = \mathbb{E}\left[f(Y)\right]$$

für all $f \in C^b(\mathcal{S})$ gleichmäßig stetig.

Beweis. (zu a) **(⇒)** Klar.

(⇐) Aus Lemma 3.2, Punkt (1) wissen wir, dass $\mathcal{B}(\mathcal{S}) = \sigma\left(\mathcal{F}(\mathcal{S})\right)$ und $\mathbb{F}$ ist durchschnittsstabil. Wegen des *Maßeindeutigkeitssatz* reicht es zu zeigen, dass $\mathbb{P}(F) = \mathbb{Q}(F)$ für alle $F \in \mathcal{F}(S)$. Sei nun $F \subseteq \mathcal{S}$ abgeschlossen. Setze $f_k(x) := \varphi\left(k \cdot d(x, F)\right)$ (vgl. Satz 2.4, φ wie dort). Aus Lemma 2.3 folgt, dass die f_k beschränkt und gleichmäßig stetig sind mit $f_k \searrow \mathbb{1}_F$ für $k \to \infty$. Also gilt:

$$\mathbb{P}(F) = \int \mathbb{1}_F \ \mathrm{d}\mathbb{P} = \int \lim_{k\to\infty} f_k \ \mathrm{d}\mathbb{P}$$

und mit monotoner Konvergenz

$$= \lim_{k\to\infty} \int f_k \ \mathrm{d}\mathbb{P} = \lim_{k\to\infty} \int f_k \ \mathrm{d}\mathbb{Q}$$

Wiederum mit monotoner Konvergenz

$$= \int \lim_{k\to\infty} f_k \, \mathrm{d}\mathbb{Q} = \int \mathbb{1}_F \, \mathrm{d}\mathbb{Q} = \mathbb{Q}(F)$$

Da F beliebig war, folgt die Behauptung.

(zu b) Dies folgt aus Punkt (a) mit dem Transformationssatz (**??**):

$$\begin{aligned} X \overset{\mathrm{d}}{=} Y &\overset{3.16}{\Leftrightarrow} \mathbb{P}\circ X^{-1} = \mathbb{P}\circ Y^{-1} \\ &\overset{(a)}{\Leftrightarrow} \int_{\mathcal{S}} f \, \mathrm{d}(\mathbb{P}\circ X^{-1}) = \int_{\mathcal{S}} f \, \mathrm{d}(\mathbb{P}\circ Y^{-1}) \qquad \forall f \in C^b(\mathcal{S}) \text{ glm. stetig} \\ &\overset{(*)}{\Leftrightarrow} \mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(Y)\right] \qquad \forall f \in C^b(\mathcal{S}) \text{ glm. stetig} \end{aligned}$$

Verwende dafür:

$$\int_{\mathcal{S}} f \, \mathrm{d}(\mathbb{P}\circ X^{-1}) \overset{\text{Trafo}}{=} \int_{\Omega} \underbrace{f\circ X}_{=:f(X)} \, \mathrm{d}\mathbb{P} \overset{\text{Def}}{=} \mathbb{E}\left[f(X)\right] \tag{$*$}$$

□

■

LITERATURVERZEICHNIS

[1] Billingsley, P. : *Convergence of probability measures.* John Wiley & Sons, 1968

[2] Billingsley, P. : *Probability and measure.* John Wiley & Sons, 1995

[3] Borodin, A. N. ; Salminen, P. : *Handbook of Brownian motion-facts and formulae.* Birkhäuser, 2002

[4] Chow, Y. S. ; Teicher, H. : *Probability theory: independence, interchangeability, martingales.* Springer Science & Business Media, 1997

[5] Csörgő, M. ; Horváth, L. : *Limit theorems in change-point analysis.* John Wiley & Sons Chichester, 1997

[6] Czado, C. ; Schmidt, T. : *Mathematische Statistik.* Springer, 2011

[7] Dudley, R. M.: *Real Analysis and Probability.* Chapman and Hall/CRC, 1999

[8] Ferger, D. : Asymptotic distribution theory of change-point estimators and confidence intervals based on bootstrap approximation. In: *Mathematical Methods of Statistics* 3 (1994), Nr. 4, S. 362

[9] Ferger, D. : Moment equalities for sums of random variables via integer partitions and Faà di Bruno's formula. In: *Turkish Journal of Mathematics* 38 (2014), Nr. 3, S. 558–575

[10] Ferger, D. : Optimal constants in the Marcinkiewicz–Zygmund inequalities. In: *Statistics & Probability Letters* 84 (2014), S. 96–101

[11] Ferger, D. : On the supremum of a Brownian bridge standardized by its maximizing point with applications to statistics. In: *Statistics & Probability Letters* 134 (2018), S. 63–69

[12] Gänssler, P. ; Stute, W. : *Wahrscheinlichkeitstheorie.* Berlin , Heidelberg [u.a.] : Springer, 1977. – ISBN 9780387084183

[13] Hand, D. J.: Statistical Decision Theory: Estimation, Testing, and Selection by Friedrich Liese, Klaus-J. Miescke. In: *International Statistical Review* 76 (2008), Nr. 3, S. 450–450

[14] Heuser, H. : *Funktionalanalysis.* 2006

[15] Hjort, N. L. ; Pollard, D. : Asymptotics for minimisers of convex processes. In: *arXiv preprint arXiv:1107.3806* (2011)

[16] Jacod, J. ; Protter, P. : *Probability essentials.* Springer Science & Business Media, 2000

[17] Kallenberg, O. : *Foundations of modern probability.* Springer, 1997

[18] Klenke, A. : *Wahrscheinlichkeitstheorie.* Bd. 1. Springer, 2008

[19] Prokhorov, Y. V.: Convergence of random processes and limit theorems in probability theory. In: *Theory of Probability & Its Applications* 1 (1956), Nr. 2, S. 157–214

[20] Rockafellar, R. : Convex analysis. In: *Princeton Univ., Princeton, NJ* (1972)

[21] Royden, H. L. ; Fitzpatrick, P. : *Real analysis*. Bd. 32. Macmillan New York, 1988

[22] Schmidt, K. D.: *Maß und Wahrscheinlichkeit*. Springer-Verlag, 2011

[23] Schubert, H. : *Topologie eine Einführung*. 4. Aufl. Stuttgart : Teubner, 1975. – ISBN 3519122006

[24] Shorack, G. R. ; Wellner, J. A.: *Empirical processes with applications to statistics*. Bd. 59. Siam, 1986

[25] Smirnov, N. V.: Limit distributions for the terms of a variational series. In: *Trudy Matematicheskogo Instituta imeni VA Steklova* 25 (1949), S. 3–60

[26] Whitt, W. : Weak convergence of probability measures on the function space C([0,unendlich)). In: *The Annals of Mathematical Statistics* 41 (1970), Nr. 3, S. 939–944

[27] Witting, H. ; Müller-Funk, U. : *Mathematische Statistik 1 Parametrische Verfahren bei festem Stichprobenumfang*. Stuttgart : Teubner, 1985. – ISBN 3519020262

[28] Witting, H. ; Müller-Funk, U. : *Mathematische Statistik II: Asymptotische Statistik: Parametrische Statistik und nichtparametrische Modelle*. Stuttgart : Teubner, 1995