Wahrscheinlichkeitstheorie

Mathias Vetter
SoSe 2015

Inhaltsverzeichnis

1 Diskrete W'räume

Ziel der W'theorie ist die Modellierung von als zufällig interpretierten Experimenten und die Untersuchung ihrer mathematischen Eigenschaften.

Diskrete W'räume sind W'räume, deren Ereignisraum höchstens abzählbar ist.

Beispiel 1.1 (Würfeln) Der *Ereignisraum* dieses Zufallsexperiments ist $\Omega = \{1, \ldots, 6\}$. Einzelne Elemente $\omega \in \Omega$ heißen *Elementarereignisse*, Teilmengen $A \subseteq \Omega$ heißen *Ereignisse*.

Beispiele:

- Ereignis "gerade": $A_1 = \{2, 4, 6\}$
- Ereignis "durch 3 teilbar": $A_2 = \{3, 6\}$

Sinnvoll: $P(\{\omega\}) = \frac{1}{6}$ (Symmetrieüberlegungen) und $P(A) = \frac{|A|}{|\Omega|}$ für alle $A \subseteq \Omega$ (*Laplace-Verteilung*).

Hier gilt somit $P(A_1) = \frac{1}{2}$ und $P(A_2) = \frac{1}{3}$.

Beispiel 1.2 (Werfen einer Heftzwecke) Wir werfen eine Heftzwecke und notieren:

- 1, falls die Spitze schräg nach unten zeigt
- 0, falls die Heftzwecke auf dem Kopf landet.

In diesem Fall sind die Wahrscheinlichkeiten nicht durch Symmetrie bestimmbar.

Alternative: Sei $\Omega = \{0, 1\}$ und $A \subseteq \Omega$. Führt man das Experiment n-mal durch und erhält man Ergebnisse $a = (a_1, \ldots, a_n)$ mit $a_i \in \Omega$ für alle $i \in \underline{n}$, so setzt man $r_{n,a}(A) = \frac{|\{j : a_j \in A\}|}{n}$ (*relative Häufigkeit*).

Hoffnung: $r_{n,a}(A)$"$\to$"$P(A)$?! (intuitiv klar? Wie geht das formal (Konvergenz von zufälligen Objekten)?)
Man setzt etwa $P(\{1\}) = \frac{187}{500}$, falls bei 500 Würfen die Spitze 187-mal nach oben zeigte.

Definition 1.3 Es sei Ω eine (höchstens) abzählbare Menge mit $\Omega \neq \emptyset$. Eine Abbildung $P : \mathcal{P}(\Omega) \to [0, 1]$ mit

i) $P(\Omega) = 1$

ii) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ für jede Folge $(A_i)_i$ mit $A_k \cap A_l = \emptyset$ für alle $k \neq l$ (*σ-Additivität*)

heißt *Wahrscheinlichkeitsmaß* oder *W'verteilung*.

Bemerkung 1.4 Einfacher einzusehen als σ-Additivität ist *endliche Additivität*:

$$P(A \cup B) = P(A) + P(B) \qquad \text{falls } A \cap B = \emptyset.$$

i) Laplace-Verteilung: Sind A und B disjunkt, so gilt $|A \cup B| = |A| + |B|$, also:

$$P(A \cup B) = \frac{|A \cup B|}{|\Omega|} = \frac{|A|}{|\Omega|} + \frac{|B|}{|\Omega|} = P(A) + P(B).$$

ii) analog bei relativen Häufigkeiten: Sind A und B disjunkt, so treten sie nie gemeinsam ein, also:

$$r_{n,a}(A \cup B) = r_{n,a}(A) + r_{n,a}(B).$$

Wir werden später präzisieren, warum wir die stärkere σ-Additivität verwenden.

Satz und Definition 1.5 Sei $\Omega \neq \emptyset$ gegeben und $f : \Omega \to [0,1]$ mit $\sum_{\omega \in \Omega} f(\omega) = 1$. Dann existiert genau ein Wahrscheinlichkeitsmaß P auf Ω mit der Eigenschaft $P(\{\omega\}) = f(\omega)$ für alle $\omega \in \Omega$. Insbesondere gilt dann:

$$P(A) = \sum_{\omega \in A} f(\omega) \qquad \text{für alle } \omega \in \Omega \tag{1}$$

f heißt *Zähldichte* oder *W'funktion*.

Beweis:

i) *Existenz:* Definiere P wie in (1). Dann gilt wegen $f(\omega) \geq 0$ für alle $\omega \in \Omega$ (nach Voraussetzung):

$$0 \leq P(A) = \sum_{\omega \in A} f(\omega) \leq \sum_{\omega \in \Omega} f(\omega) = 1,$$

und somit $P(\Omega) = 1$. Für $(A_i)_i$ mit $A_k \cap A_l = \emptyset$ für alle $k \neq l$ gilt:

$$P\left(\bigcup_{i \geq 1} A_i\right) = \underbrace{\sum_{\omega \in \bigcup_{i \geq 1} A_i} f(\omega)}_{\text{abs. konvergent}} = \sum_{i=1}^{\infty} \sum_{\omega \in A_i} f(\omega) = \sum_{i=1}^{\infty} P(A_i).$$

Das zweite Gleichheitszeichen gilt, da ω nur in genau einem A_i liegt und absolut konvergente Reihen auch bei beliebiger Permutation ihrer Glieder gegen den gleichen GW konvergieren (Riemannscher Umordnungssatz).

ii) *Eindeutigkeit:* Für jedes Wahrscheinlichkeitsmaß Q auf Ω gilt wegen der σ-Additivität, dass

$$Q(A) = \sum_{\omega \in A} Q(\{\omega\}) = \sum_{\omega \in \Omega} f(\omega) \qquad \text{für alle } A \subseteq \Omega.$$

Also legt $Q(\{\omega\})$ für alle $\omega \in \Omega$ Q eindeutig fest. □

Bemerkung: Der Satz 1.5 besagt, daß die W'funktion das Wahrscheinlichkeitsmaß bzw. das Zufallsexperiment vollständig charakterisiert.

Beispiel 1.6

i) *Laplace-Verteilung:* Es sei Ω endlich. Die Laplace-Verteilung mit Zähldichte $f(\omega) = \frac{1}{|\Omega|}$ für alle $\omega \in \Omega$ wird verwendet, wenn symmetrische Modellierung angemessen ist.

ii) *Binomial-Verteilung:* Urne mit N Kugeln, von denen R rot und $N - R$ weiß sind. Wie groß ist die Wahrscheinlichkeit, daß unter n gezogenen Kugeln (mit Zurücklegen) genau r rote sind?

Idee: Numeriere die Kugeln ($1,\ldots,R$ rot, $R+1,\ldots,N$ weiß). Gesucht ist $P(E_r)$ mit

$$E_r = \{(\omega_1,\ldots,\omega_n) : \omega_i \in \{1,\ldots,N\} \wedge |\{i : 1 \geq \omega_i \geq R\}|= r\}.$$

ω_i ist hierbei die Zahl der gezogenen Kugel. Offenbar sind alle Vektoren $(\omega_1,\ldots,\omega_n)$ gleichwahrscheinlich, dh. $P(E_r) = \frac{|E_r|}{|\Omega|}$ mit $|\Omega| = N^n$. Zur Bestimmung von $|E_r|$ betrachte für $I \subseteq \{1,\ldots,n\}$ die Menge

$$E_I = \{(\omega_1,\ldots,\omega_n) : \omega_i \in \{1,\ldots,R\} \forall i \in I \wedge \omega_i \in \{R+1,\ldots,N\} \forall i \notin I\}.$$

Die Elemente von I geben an, in welchem Zug eine rote Kugel gezogen wurde. Es gilt, dass die E_I disjunkt sind, also folgt

$$E_r = \biguplus_{\substack{I \subset \underline{n} \\ |I|=r}} E_I$$

und $|E_I| = R^r(N-R)^{n-r}$ für alle I mit $|I| = r$.
Wie viele I gibt es? Die Kombinatorik lehrt: $\binom{n}{r}$.
Insgesamt folgt nun:

$$P(E_r) = \underbrace{\frac{1}{N^n}}_{=|\Omega|^{-1}} \underbrace{\binom{n}{r} R^r (N-R)^{n-r}}_{=|E_r|} = \binom{n}{r} \left(\frac{R}{N}\right)^r \left(\frac{N-R}{R}\right)^{n-r} \qquad \text{mit } r \in \{0,\ldots,n\}.$$

Beachte: $f(r) = \binom{n}{r} p^r (1-p)^{n-r}$ mit $p \in (0,1), r \in \{0,\ldots,n\}$ definiert nach Binomischen Lehrsatz eine Zähldichte und heißt *Binomialverteilung mit Parametern n und p.*

iii) *Hypergeometrische Verteilung:* Situation von ii), jedoch Ziehen ohne Zurücklegen. Dann gilt für $r \in \{0,..R\}$ und $n < N$:

$$f(r) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

definiert die Zähldichte der *hypergeometrischen Verteilung mit Parametern N, R und n.*

Bemerkung 1.7 Es sei eine hypergeometrische Verteilung mit Parametern $N, \lfloor Np \rfloor$ und n gegeben, wobei $p \in (0,1)$. Dann gilt (Übung!):

$$f_N(r) = \frac{\binom{\lfloor Np \rfloor}{r}\binom{N-\lfloor Np \rfloor}{n-r}}{\binom{N}{n}} \underset{N\to\infty}{\longrightarrow} \binom{n}{r} p^r (1-p)^{n-r}.$$

Satz 1.8 (Poisson-Grenzwertsatz) Sei $(p_n)_{n\in\mathbb{N}}$ eine Folge mit $0 \leq p_n \leq 1$ und es gilt $np_n \underset{n\to\infty}{\longrightarrow} \lambda > 0$. Dann gilt:

$$f_n(r) = \binom{n}{r} p^r (1-p)^{n-r} \underset{n\to\infty}{\longrightarrow} e^{-\lambda} \frac{\lambda^r}{r!} \qquad \text{für alle } r \geq 0.$$

Beweis: Da $np_n \to \lambda > 0$, folgt $p_n \to 0$. Sei $r \in \mathbb{N}$, so gilt:

$$\begin{aligned}\binom{n}{r} p^r (1-p)^{n-r} &= \frac{n!}{(n-r)!} \frac{1}{r!} \left(\frac{np_n}{n}\right)^r \left(1 - \frac{np_n}{n}\right)^n (1-p_n)^{-r} \\ &= \underbrace{\frac{n(n-1)\cdot \ldots \cdot (n-r+1)}{n^r}}_{\to 1} \frac{1}{r!} \underbrace{(np_n)^r}_{\to \lambda^r} \underbrace{\left(1 - \frac{np_n}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{(1-p_n)^{-r}}_{\to 1}.\end{aligned}$$

Insgesamt folgt: $\lim_{n\to\infty} f_n(r) = e^{-\lambda}\frac{\lambda^r}{r!}$. □

Die Verteilung mit Zähldichte $f(r) = e^{-\lambda}\frac{\lambda^r}{r!}$ $(r \geq 0)$ heißt *Poisson-Verteilung mit Parameter* λ.

Beispiel 1.9 Wir ziehen so lange aus einer Urne mit R roten und $N-R$ weißen Kugeln mit Zurücklegen, bis zum ersten Mal eine rote Kugel gezogen wird. Mit $p = \frac{R}{N}$ und $r \geq 1$ definiert

$$f(r) = p(1-p)^{r-1}$$

die Zähldichte der *geometrischen Verteilung.* r gibt hierbei die Anzahl der gesamten Züge an.

Beachte: Auch hier existiert eine Zähldichte für alle $p \in (0,1)$ (geometrische Reihe).

Bemerkung 1.10

i) Oft wird die geometrische Verteilung auch als die Anzahl der gezogenen weißen Kugeln vor der ersten roten definiert. In diesem Fall ergibt die eine Verteilung auf $\{0, 1, 2, \ldots\}$.

ii) Verallgemeinerung: *negative Binomialverteilung* (Anzahl der Züge bis k Erfolge erreicht sind, $k \geq 1$ fest).

2 Unabhängigkeit und bedingte Wahrscheinlichkeiten

Bemerkung 2.1 Wir wollen zwei Ereignisse A und B als unabhängig ansehen, wenn das Eintreten von A das Eintreten von B nicht verändert.

- Endliche Grundgesamtheit: Merkmal B tritt eingeschränkt auf A relativ genauso oft auf wie in der gesamten Menge Ω, dh.

$$\frac{|A \cap B|}{|A|} = \frac{|B|}{|\Omega|}.$$

- relative Häufigkeit: Entspricht n_c der Häufigkeit von C unter n Versuchen, so fordert man:

$$\lim_{n\to\infty} \frac{n_{A\cap B}}{n_A} = \lim_{n\to\infty} \frac{n_B}{n}.$$

In beiden Fällen erhält man als Konsequenz $\frac{P(A\cap B)}{P(A)} = P(B)$.

Definition 2.2 Sei $\Omega \neq \emptyset$ und P Wahrscheinlichkeitsmaß auf Ω, dann heißen $A, B \subset \Omega$ *unabhängig*, falls gilt:

$$P(A \cap B) = P(A)P(B).$$

Beispiel 2.3 Urne mit 4 roten und 5 weißen Kugeln. Zwei Züge:

- A: "erste Kugel rot"
- B: "zweite Kugel weiß"

Ziehen mit Zurücklegen: Es gilt $P(A) = \frac{4}{9} \cdot \frac{9}{9} = \frac{4}{9}$ und $P(B) = \frac{9}{9} \cdot \frac{5}{9} = \frac{5}{9}$.
Außerdem ist $P(A \cap B) = \frac{4}{9} \cdot \frac{5}{9} = P(A)P(B)$. Also sind die Ereignisse unabhängig.

Ziehen ohne Zurücklegen: Es ist $P(A) = \frac{4}{9} \cdot \frac{8}{8} = \frac{4}{9}$. Sei außerdem A^c das Gegenereignis von A. Dann gilt: $P(A\cap B) = \frac{4}{9} \cdot \frac{5}{8}$ und $P(A^c \cap B) = \frac{5}{9} \cdot \frac{4}{8}$. Mit endlicher Additivität folgt: $P(A\cap B) + P(A^c \cap B) = P(B)$, also hier $\frac{4}{9} \cdot \frac{5}{8} + \frac{5}{9} \cdot \frac{4}{8} = \frac{5}{9}$. Insgesamt: $P(A\cap B) = \frac{5}{9} \cdot \frac{4}{8} > \frac{4}{9} \cdot \frac{5}{9} = P(A)P(B)$. Intuition: Die fehlende rote Kugel aus dem ersten Zug macht B wahrscheinlicher.

Definition 2.4 Ereignisse $A_1, \dots, A_n$ heißen *unabhängig*, falls für jedes $k \in \underline{n}$ und alle Indices $1 \leq i_1 < \ldots < i_k \leq n$ gilt:

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}). \tag{2}$$

Korollar 2.5 Sind $A_1, \dots, A_n$ unabhängig, so auch beliebige Teilfamilien.

Lemma 2.6 Seien $A_1, \ldots, A_n$ unabhängig. Betrachte neue Ereignisse $B_1, \ldots, B_n$ mit $B_i = A_i$ oder $B_i = A_i^c$. Dann sind auch $B_1, \ldots, B_n$ unabhängig.

Beweis: Es reicht zu zeigen (Rest Induktion): $A_1, \ldots, A_n$ unabhängig $\Rightarrow$ $A_1^c, A_2, \ldots, A_n$ unabhängig. Seien $k \in \underline{n}$ und $1 \leq i_1 < \ldots < i_k \leq n$ beliebig. Ist $i_1 > 1$, so ist nichts zu zeigen (folgt aus 2.5). Sei nun $i_1 = 1$. Dann gilt:

$$P(A_1^c \cap \underbrace{A_{i_1} \cap \ldots \cap A_{i_k}}_{=:B}) = P(B) - P(A_1 \cap B) = \prod_{j=2}^{k} P(A_{i_j}) - P(A_1) \prod_{j=2}^{k} P(A_{i_j}) = (1 - P(A_1)) \prod_{j=2}^{k} (A_{i_j}).$$

Wegen $1 = P(\Omega) = P(A_1) + P(A_1^c) \Leftrightarrow P(A_1^c) = 1 - P(A_1)$, folgt die Behauptung. □

Bemerkung 2.7

i) Würde man in 2.4 statt (2) nur $P\left(\bigcap_{i \in \underline{n}} A_i\right) = \prod_{i \in \underline{n}} P(A_i)$ voraussetzen, so hießen $A_1, \ldots, A_n$ bereits dann unabhängig, wenn nur ein $A_i = \emptyset$ (da $P(\emptyset) = 0$).

ii) Ebenso ist es nicht ausreichend, nur paarweise Unabhängigkeit vorauszusetzen. Man kann etwa beim zweifachen Münzwurf zeigen zeigen:

- A :"erster Wurf Kopf"
- B :"zweiter Wurf Kopf"
- C :"Anzahl von Kopf ist gerade"

Es gilt: A, B und C sind paarweise unabhängig, aber $A \cap B \Rightarrow C$.

Sind Ereignisse nicht unabhängig, so benötigen wir einen erweiterten Wahrscheinlichkeitsbegriff um den Einfluß von A auf den Eintritt von B zu messen.

Beispiel 2.8 Zweifacher Würfelwurf, Ereignis B : "Augensumme ist 12". Offenbar gilt $P(B) = \frac{1}{36}$. Zusatzinformation: A "im ersten Wurf kam 5" ist eingetreten. Nun ist B unmöglich geworden. Notation: $P(B : A) = 0$.

Definition 2.9 Es seien $A, B \in \mathcal{P}(\Omega)$ mit $P(A) > 0$ gegeben. Dann heißt

$$P_A(B) = P(B \mid A) =: \frac{P(B \cap A)}{P(A)}$$

die *bedingte Wahrscheinlichkeit von B gegeben A.*

Bem.: Sind A und B unabhängig, so folgt $P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$.

Satz 2.10 Sei $A \in \mathcal{P}(\Omega)$ mit $P(A) > 0$ gegeben. Dann definiert $P_A : \mathcal{P}(\Omega) \to [0, 1], B \mapsto P_A(B)$ ein Wahrscheinlichkeitsmaß auf Ω mit $P_A(A) = 1$.

Beweis: Es gilt $P_A(B) \in [0, 1]$, da $0 \leq P(B \cap A) < P(A) \leq 1$. Weiter ist:

$$P_A(\Omega) = \frac{P(A \cap \Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1 = \frac{P(A \cap A)}{P(A)} = P_A(A).$$

Sei außerdem $(B_i)_i$ mit $B_i \subseteq \Omega$ für alle i paarweise disjunkt. Es gilt:

$$P_A\left(\biguplus_{i\geq 1} B_i\right) = \frac{P\left((\uplus_{i\geq 1} B_i) \cap A\right)}{P(A)} = \frac{P\left(\uplus_{i\geq 1}(B_i \cap A)\right)}{P(A)} = \frac{\sum_{i\geq 1} P(B_i \cap A)}{P(A)} = \sum_{i\geq 1} P_A(B_i).$$

□

Beispiel 2.11 Betrachte nochmal den zweifachen Würfelwurf aus 2.8. Dann gilt: $P(A \cap B) = 0$ und $P(A) = \frac{1}{6}$. Folglich ist $P(B \mid A) = 0$.

Andererseits ist $P(A^c) = \frac{5}{6}$ und $P(A^c \cap B) = \frac{1}{36}$, also $P(B \mid A^c) = \frac{1}{30}$.

Intuitiv: Gilt A^c und soll B eintreten, dann muß in ersten Wurf unter 5 und im zweiten Wurf unter 6 Möglichkeiten die "6" getroffen werden.

Satz 2.12 (Multiplikationsregel) Es seien $A_1, \ldots, A_n \in \mathcal{P}(\Omega)$ mit $P\left(\bigcap_{i\in \underline{n}} A_i\right) \neq 0$. Dann gilt:

$$P\left(\bigcap_{i\in \underline{n}} A_i\right) = P(A_1) \cdot P(A_2 \mid A_1) \cdot \ldots \cdot P(A_n \mid A_1 \cap \ldots \cap A_{n-1}).$$

Beweis: Induktion: $n = 1$ ist klar. $n \to n+1$:

$$\begin{aligned} P(A_1 \cap \ldots \cap A_{n+1}) &= \frac{P(A_1 \cap \ldots \cap A_{n+1})}{P\left(\cap_{i\in \underline{n}} A_i\right)} \cdot P\left(\bigcap_{i\in \underline{n}} A_i\right) = P(A_{n+1} \mid \cap_{i\in \underline{n}} A_i) \cdot P\left(\bigcap_{i\in \underline{n}} A_i\right) \\ &\overset{[\text{IV}]}{=} P(A_{n+1} \mid \cap_{i\in \underline{n}} A_i) \cdot P(A_n \mid \cap_{i\in\{1,\ldots,n-1\}} A_i) \cdot \ldots \cdot P(A_1). \end{aligned}$$

□

Beispiel 2.13 Ziehen ohne Zurücklegen, R rote und $N - R$ weiße Kugeln, n Züge. Dann:

$$\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_i \in \{0, 1\} \text{ für alle } i \in \underline{n}\},$$

wobei 0 für eine weiße Kugel und 1 für eine rote Kugel steht. Setze:

- $A_i^{(0)}$:"i-te gezogene Kugel ist weiß"
- $A_i^{(1)}$:"i-te gezogene Kugel ist rot"

Wegen $(\omega_1, \ldots, \omega_n) = \bigcap_{i\in \underline{n}} A_i^{(\omega_i)}$ und $P\left(A_{k+1}^{(0)} : A_1^{(\omega_1)} \cap \ldots \cap A_k^{(\omega_k)}\right) = \frac{N-R\left(k-\sum_{i\in \underline{n}} \omega_i\right)}{N-k}$ bzw. $P\left(A_{k+1}^{(1)} \mid A_1^{(\omega_1)} \cap \ldots \cap A_k^{(\omega_k)}\right) = \frac{R-\sum_{i\in \underline{n}} \omega_i}{N-k}$ ergibt sich mit $\sum_{i\in \underline{n}} \omega_i = r$:

$$P(\{(\omega_1, \ldots, \omega_n)\}) = \frac{R(R-1) \cdot \ldots \cdot (R-r+1)(N-R)(N-R-1) \cdot \ldots \cdot (N-R(n-1)+1)}{N(N-1) \cdot \ldots \cdot (N-n+1)}.$$

Satz 2.14 (Satz der totalen Wahrscheinlichkeit) Sei $\Omega = \biguplus_{i\geq 1} A_i$ eine disjunkte Zerlegung und $B \in \mathcal{P}(\Omega)$ beliebig. Dann gilt:

$$P(B) = \sum_{\substack{i\geq 1\\ P(A_i)>0}} P(B \mid A_i)P(A_i).$$

Beweis:

$$P(B) = P((\uplus_{i\geq 1} A_i) \cap B) = P(\uplus_{i\geq 1}(A_i \cap B)) = \sum_{\substack{i\geq 1\\ P(A_i)\geq 0}} P(A_i \cap B) = \sum_{\substack{i\geq 1\\ P(A_i)\geq 0}} P(B \mid A_i)P(A_i).$$

□

Korollar 2.15 (Formel von Bayes) Situation von 2.14 (mit oE. $P(B), P(A_i) > 0$ für alle i). Dann gilt:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{k\geq 1} P(B \mid A_k)P(A_k)}.$$

Beweis:

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i \mid B)P(A_i)}{\sum_{k\geq 1} P(B \mid A_k)P(A_k)}.$$

□

Beispiel 2.16 Einer von 1000 Patienten habe eine gefährliche Krankheit. Ein Test entdeckt bei 99% aller Kranken die Krankheit, jedoch auch bei 5% der Gesunden. Wie groß ist die Wahrscheinlichkeit bei einem positiven Test krank zu sein?

Es seien A : "erkrankt" und B : "positiv", dann gilt:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)} = \frac{\frac{99}{100}\frac{1}{1000}}{\frac{99}{100}\frac{1}{1000} + \frac{5}{100}\frac{999}{1000}} = \frac{99}{5094} \approx 0,0194.$$

3 diskrete Zufallsgrößen

Beispiel 3.1 Oft interessieren wir uns nicht für das konkrete Zufallsexperiment, sondern für dessen quantitative Eigenschaften:

- Münzwurf: "Kopf"$\rightarrow 1$, "Zahl"$\rightarrow 0$
- Würfel: "gerade"$\rightarrow 1$, "ungerade"$\rightarrow 0$

In beiden Fällen gilt: $P(\{0\}) = P(\{1\}) = \frac{1}{2}$. Es ist daher zweckmäßig, beide Experimente formal als dasselbe zu betrachten.

Definition 3.2 Eine Abbildung $X : \Omega \rightarrow \Omega'$ (zumeist $\mathbb{R}$) heißt *Zufallsgröße* oder auch *Zufallsvariable.*

Bemerkung 3.3 Für $A' \subseteq \Omega'$ definieren wir $\{X \in A'\} = \{\omega \in \Omega : X(\omega) \in A'\} = X^{-1}(A')$.

Definition und Satz 3.4 Sei $X : \Omega \rightarrow \Omega'$ Zufallsgröße. Dann definiert $P^X : \mathcal{P}(\Omega) \rightarrow [0,1], A' \mapsto P^X(A') = P(X^{-1}(A'))$ ein W'maß auf Ω'. P^X heißt *Verteilung von X* bzw. formal *Bildmaß von P auf X*.

Beweis:

i) $P^X(A') = P(\underbrace{\{\omega : X(\omega) \in A'\}}_{\subseteq \Omega}) \in [0,1]$.

ii) $P^X(\Omega') = P\left(\{\omega : X(\omega) \in \Omega'\}\right) = P(\Omega) = 1$.

iii) Seien $(A'_i)_i$ paarweise disjunkt. Dann sind auch die $(X^{-1}(A'_i))_i$ disjunkt. Also ist

$$P^X\left(\cup_i A'_i\right) = P(X^{-1}(\cup_i A'_i)) = P(\cup_i X^{-1}(A'_i)) = \sum_i P(X^{-1}(A'_i)) = \sum_i P^X(A'_i).$$

□

Bemerkung 3.5 Da Ω (höchstens) abzählbar ist, gilt $\Omega' \supseteq X(\Omega) = \{x_1, x_2, \ldots\}$. Die Funktion

$$f_X(x_i) := P^X(\{x_i\}) = P(\{X = x_i\})$$

heißt *W'funktion von X*. Offenbar gilt wieder: $P^X(A') = \sum_{x_i \in A'} f_X(x_i)$.

Inbesondere benötigen wir den ursprünglichen Raum (Ω, P) nicht mehr.

Beispiel 3.6 Interessieren wir uns für die Teilnahme an einem Glücksspiel, so ist der faire Einsatz für dieses Spiel von Bedeutung. Etwa Würfeln: Gewinn ist die erzielte Augenzahl.

Idee: fairer Einsatz entspricht dem mittleren Gewinn. Mittelwert: $\frac{1}{6}(1+\ldots+6) = 3,5$, da alle Augenzahlen gleichwahrscheinlich sind.

"frequentistische" Interpretation: bei häufiger Wiederholung macht man unter 3,5 Geld Gewinn, darüber Verlust.

Definition 3.7 Es sei $X : \Omega \to \mathbb{R}$ Zufallsgröße. Wir bezeichnen mit

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$$

den *Erwartungswert von* X, falls:

- $X \geq 0$ (dann möglich $\mathbb{E}[X] = \infty$) oder
- $\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) < \infty$.

In letzterem Fall heißt $X \in \mathcal{L}^1$.

Beispiel 3.8

i) Sei X binomialverteilt mit Parametern n und p, dh. kurz $X \sim Bin(n,p)$.
Formal: $\Omega = \{0, \ldots, n\}$, P wie in Bsp. 1.6(ii) und $X(\omega) = \omega$ für alle $\omega \in \Omega$.
Beachte: $k\binom{n}{k} = \frac{n!}{(n-k)!(k-1)!} = n\binom{n-1}{k-1}$. Also (mit Binomischen Lehrsatz):

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= np \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}}_{=1} = np.\end{aligned}$$

ii) *Exkurs:* St.-Petersburg-Paradoxon:
Glücksspiel: Man werfe eine faire Münze so lange, bis zum ersten Mal Kopf erscheint.
Auszahlung: k Versuche $\to$ 2^k Geld. Es gilt:

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} 2^k \frac{1}{2^k} = \infty.$$

Gegenargument zur Intuition von Erwartungswert als fairer Einsatz.

Satz 3.9 Sei $X : \Omega \to \mathbb{R}$ Zufallsgröße und $f : \mathbb{R} \to \mathbb{R}$. Dann gilt:

$$\mathbb{E}[f(X)] = \sum_{x \in X(\Omega)} f(x) P(X = x), \qquad \text{falls } f \geq 0 \text{ oder } f \circ X \in \mathcal{L}^1.$$

Beweis: Es gilt:

$$\mathbb{E}[f(X)] = \sum_{\omega \in \Omega} f(X(\omega)) P(\{\omega\}) = \sum_{x \in X(\Omega)} f(x) \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} P(\{\omega\}) \overset{(3.5)}{=} \sum_{x \in X(\Omega)} f(x) P(X = x).$$

□

Bem: Der Satz lehrt, dass für die Berechnung von E'werten der ursprüngliche W'raum nicht mehr benötigt wird.

Bemerkung 3.10 Für den Raum $\mathcal{L}^1$ gelten die folgenden Eigenschaften:

i) $\mathcal{L}^1$ ist $\mathbb{R}$-Vektorraum und es gilt für $X, Y \in \mathcal{L}^1$ und $a \in \mathbb{R}$:

$$\mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y].$$

ii) Gilt $X \leq Y$, so folgt $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Beweis:

i) folgt direkt aus Definitionen und GWS für Reihen

ii) Es gilt $\mathbb{E}[Y] = \mathbb{E}[X] + \mathbb{E}[Y - X]$ und $Y - X \geq 0$. Dann gilt nach Def. 3.7 auch $\mathbb{E}[Y - X] \geq 0$. □

Definition 3.11 Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße und $X \in \mathcal{L}^2$ (d.h. $X^2 \in \mathcal{L}^1$). Dann heißt

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Varianz von X. Ferner heißt $\sigma_X = \sqrt{Var(X)}$ *Standardabweichung von X.*

Bemerkung 3.12

i) Allgemein gilt:

$$\begin{aligned}
\mathbb{E}[|X|] \overset{(3.9)}{=} \sum_{x \in X(\Omega)} |x| P(X = x) &= \sum_{|x| \leq 1} |x| P(X = x) + \sum_{|x| > 1} |x| P(X = x) \\
&\leq \sum_{|x| \leq 1} |x| P(X = x) + \underbrace{\sum_{|x| > 1} |x|^2 P(X = x)}_{\text{endlich, da } X \in \mathcal{L}^2} \\
&\leq 1 + \mathbb{E}[X^2].
\end{aligned}$$

Also $X \in \mathcal{L}^2$, dh. $\mathbb{E}[X^2] < \infty \;\Rightarrow\; X \in \mathcal{L}^1$, dh. $\mathbb{E}[X] < \infty$. Insbesondere: $Var(X)$ ist wohldefiniert.

ii) $Var(X)$ misst eine mittlere Abweichung von X zu ihrem Erwartungswert und ist daher ein Maß für die Streuung von X (maW. ein Maß für das Risiko). Dass wir die *quadratische* Abweichung betrachten hat vor allem rechentechnische Gründe.

iii) Sinn der Standardabweichung: Größen in den richtigen Einheiten angeben.
X : Größe in m $\to$ $Var(X)$: "Größe" in m^2 $\to$ σ_X : Größe in m.

Satz 3.13 (Verschiebungssatz) Seien $X : \Omega \to \mathbb{R}$ mit $X \in \mathcal{L}^2$ und $a, b \in \mathbb{R}$. Dann gilt:

- $Var(aX + b) = a^2 Var(X)$.
- $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Beweis: Übung. □

Satz 3.14 (Markow'sche Ungleichung) Sei $X : \Omega \to \mathbb{R}$ mit $X \in \mathcal{L}^1$ oder $X \geq 0$, sowie $a > 0$. Dann gilt:

$$P(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}.$$

Beweis: Für $x \in \mathbb{R}$ mit $|x| \geq a$ gilt $\frac{|x|}{a} \geq 1$. Also:

$$P(|X| > a) = \sum_{\{x:|x|>a\}} P(X = x) \leq \sum_{\{x:|x|>a\}} \frac{|x|}{a} P(X = x) \leq \sum_{x \in \mathbb{R}} \frac{|x|}{a} P(X = x) = \mathbb{E}\left[\frac{|X|}{a}\right] = \frac{\mathbb{E}[|X|]}{a}.$$

□

Bem: Der Satz kann verallgemeinert werden zu: $\varphi(a)P(X \geq a) \leq \mathbb{E}[\varphi(X)]$ für monoton wachsende φ.

Korollar 3.15 (Tschebyschoff'sche Ungleichung) Sei $X : \Omega \to \mathbb{R}$ in $\mathcal{L}^1$ und $a > 0$. Dann gilt:

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var(X)}{a^2}.$$

Beweis: Es gilt $\{|X - \mathbb{E}[X]| \geq a\} = \{|X - \mathbb{E}[X]|^2 \geq a^2\}$, da $|X - \mathbb{E}[X]|$ und a nicht negativ. Der Rest folgt mit 3.14 und Definition der Varianz. □

4 σ-Algebren und Maße

Bemerkung 4.1 Neben diskreten W'räumen (Ω, P) interessieren wir uns für W'maße auf allgemeinen Mengen Ω, zB. $\Omega = \mathbb{R}$. Es stellt sich heraus, dass ein sinnvoller Begriff eines Maßes nicht bzw. nur eingeschränkt möglich ist. Zum Beispiel existiert keine Funktion $\mathcal{P}([0,1]) \to [0,1], [a,b] \mapsto b-a$, die die beiden Bedingungen aus 1.3 erfüllt.

Definition 4.2 Sei $\Omega \neq \emptyset$. Eine Teilmenge $\mathcal{A} \in \mathcal{P}(\Omega)$ heißt *σ-Algebra*, falls gilt:

i) $\Omega \in \mathcal{A}$.

ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$. (Komplementstabilität)

iii) $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{A}$. (abzählbare Vereinigungen)

Ein Paar $(\Omega, \mathcal{A})$ heißt *Messraum*.

Bemerkung 4.3 σ-Algebren sind unter abzählbaren Mengenoperationen abgeschlossen. Zum Beispiel gilt:

- $\emptyset = \Omega^c \in \mathcal{A}$ (1. und 2.)
- $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcap_{i \geq 1} A_i = \left(\bigcup_{i \geq 1} A^c\right)^c \in \mathcal{A}$ (2. und 3.)

Daraus folgt, dass auch endliche Vereinigungen und Schnitte von Elementen aus $\mathcal{A}$ wieder in $\mathcal{A}$ liegen. Außerdem ist zum Beispiel:

- $A \setminus B = A \cup B^c \in \mathcal{A}$, falls $A, B \in \mathcal{A}$
- symmetrische Differenzen: $A \triangle B = (A \setminus B) \cup (B \setminus A) \in \mathcal{A}$, falls $A, B \in \mathcal{A}$
- etc.

Ziel: σ-Algebren so groß und reichhaltig wie möglich zu gestalten.

Beispiel 4.4

i) $\mathcal{P}(\Omega)$ ist σ-Algebra.

ii) $\{\emptyset, \Omega\}$ ist (die triviale) σ-Algebra.

iii) Sei I Indexmenge. Ist $\mathcal{A}_i$ für jedes $i \in I$ σ-Algebra, dass ist auch $\bigcap_{i \geq 1} \mathcal{A}_i$ σ-Algebra.
Beweis: Nach Def. ist $\Omega \in \mathcal{A}$ für alle $i \in I$, also folgt $\Omega \in \bigcap \mathcal{A}_i$. Weiter sei $A \in \bigcap_{i \geq 1} \mathcal{A}_i$. Es folgt: $A \in \mathcal{A}_i$ für alle $i \in I$, also $A^c \in \mathcal{A}_i$ für alle $i \in I$ und somit $A^c \in \bigcap_{i \geq 1} \mathcal{A}_i$. Die letzte Bedingung wird analog bewiesen.

Definition 4.5 Es sei $\mathcal{C} \in \mathcal{P}(\Omega)$. Dann heißt

$$\sigma(\mathcal{C}) := \bigcap \{\mathcal{A} : \mathcal{A}\ \sigma\text{-Algebra mit } \mathcal{C} \subseteq \mathcal{A}\}$$

die *von $\mathcal{C}$ erzeugte σ-Algebra*. Sie ist die kleinste σ-Algebra, die $\mathcal{C}$ enthält.

Beispiel 4.6

i) Für $A \subseteq \Omega$ gilt $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.

ii) Es sei Ω topologischer Raum. Dann heißt

$$\mathcal{B} = \mathcal{B}(\Omega) := \sigma(\{A \subseteq \Omega \mid A \text{ offen}\}) = \sigma(\{A \subseteq \Omega \mid A \text{ abgeschlossen}\})$$

die *Borel-σ-Algebra* auf Ω. Sie ist oftmals die "richtige" σ-Algebra zur Definition von W'maßen.

:

Satz 4.7 Es gilt: $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a] \mid a \in \mathbb{Q}\})$.

Beweis: Zunächst ist zu zeigen, dass $\mathcal{B} = \sigma(\mathcal{C})$ mit $\mathcal{C} = \{(a,b) \mid a, b \in \mathbb{R} \text{ und } a \leq b\}$. Es ist klar, dass $\mathcal{C} \subseteq \{A \mid A \text{ offen}\}$, also $\mathcal{C} \subseteq \sigma(\{A \mid A \text{ offen})\}$ und somit $\sigma(\mathcal{C}) \subseteq \sigma(\{A \mid A \text{ offen}\}) = \mathcal{B}$. Sei nun A offene Menge, so gilt $A \in \bigcup\{(a,b) \mid (a,b) \in A \text{ und } a \leq b \in \mathbb{R}\}$, also $\mathcal{B} \subseteq \sigma(\mathcal{C})$, da $\{A \mid A \text{ offen}\} \in \sigma(\mathcal{B})$.

Setze nun $\mathcal{D} := \{(-\infty, a] \mid a \in \mathbb{Q}\}$. Es ist klar, dass $\mathcal{D} \subseteq \sigma(\mathcal{C}) \subseteq \mathcal{B}$. Andererseits sei (a,b) mit $a \leq b$ beliebig. Wähle $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \subseteq \mathbb{Q}$, sodass gilt $(a_n)_{n\in\mathbb{N}} \searrow a$ und $(b_n)_{n\in\mathbb{N}} \nearrow b$ ($\searrow, \nearrow$ bedeuten strenge Monotonie). Dann gilt $(a,b) = \bigcup_{n\geq 1}(a_n, b_n] = \bigcup_{n\geq 1}((-\infty, b_n] \setminus (-\infty, a_n])$. Es folgt $\mathcal{C} \subseteq \sigma(\mathcal{D})$, also $\mathcal{B} = \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{D})$. $\square$

Bemerkung 4.8 Oft ist es nicht leicht möglich, ein Maß direkt auf einer σ-Algebra zu definieren. Man beschränkt sich zunächst auf ein Teilsystem und setzt das Maß dann geeignet fort.

Definition 4.9 Ein System $\mathcal{R} \in \mathcal{P}(\Omega)$ heißt *Ring*, falls gilt:

i) $\emptyset \in \mathcal{R}$,

ii) $A, B \in \mathcal{R} \;\Rightarrow\; A \setminus B \in \mathcal{R}$,

iii) $A, B \in \mathcal{R} \;\Rightarrow\; A \cup B \in \mathcal{R}$.

Gilt auch $\Omega \in \mathcal{R}$, so heißt $\mathcal{R}$ *Algebra*.

Bemerkung 4.10 Im Unterschied zu σ-Algebren bestehen Algebren i.A. nur aus ihren Erzeugern und endlichen Vereinigungen/Schnitten.

Beispiel 4.11

i) Der kleinste Ring auf Ω besteht nur aus $\emptyset$.

ii) σ-Algebren sind immer auch Algebren und somit auch Ringe.

iii) $\{A \subseteq \Omega \mid |A| < \infty \vee |A^c| < \infty\}$ ist Algebra und genau dann σ-Algebra, wenn $|\Omega| < \infty$ (Übungsblatt 3).

iv) Die Menge aller endlichen Vereinigungen von halboffenen Intervallen $(a,b]$ in $\mathbb{R}$ (oder $\mathbb{R}^d$) ist ein Ring und heißt *Ring der Figuren* (in $\mathbb{R}^d$ werden Intervalle komponentenweise aufgefasst). Seine Elemente werden *Figuren* genannt.

Definition 4.12 Es sei $\mathcal{A}$ σ-Algebra auf Ω. Eine Abbildung $\mu : \mathcal{A} \to [0,\infty]$ heißt *Maß*, falls gilt:

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

ii) $\mu\Big(\bigcup_{i\geq 1} A_i\Big) = \sum_{i\geq 1} \mu(A_i)$ für alle $(A_i)_{i\in\mathbb{N}}$ mit $A_k \cap A_l = \emptyset$, falls $k \neq l$.

Gilt zusätzlich $\mu(\Omega) = 1$, so heißt μ *Wahrscheinlichkeitsmaß*. Bezeichnung dann meist P.
Das Tripel $(\Omega, \mathcal{A}, \mu)$ heißt *Maßraum* bzw. *Wahrscheinlichkeitsraum* bei $(\Omega, \mathcal{A}, P)$.

Bemerkung 4.13

i) Ist $\mathcal{A}$ ein Ring oder eine Algebra und gelten 1) und 2) aus Def. 4.12, wann immer für eine disjunkte Folge $(A_n)_n$ auch $\bigcup_{n\geq 1} A_n \in \mathcal{A}$ ist, dann heißt μ ein *Prämaß*.

ii) Ist $\mathcal{A}$ Ring oder Algebra und gilt neben $\mu(\emptyset) = 0$ nur die endliche Additivität, dh. $\mu\left(\bigcup_{i\in\underline{n}} A_i\right) = \sum_{i\in\underline{n}} \mu(A_i)$ für alle $A_1, \ldots, A_n$ disjunkt, so heißt μ ein *Inhalt*.

iii) Jedes Maß/Prämaß ist ein Inhalt.

Lemma 4.14 Für Maße μ bzw. W'maße (bzw. endliche Maße) P gilt:

i) $\mu(A \cup B) = \mu(A) + \mu(B)$ für disjunkte A, B.

ii) $\mu\left(\bigcup_{i\geq 1} A_i\right) \leq \sum_{i\geq 1} \mu(A_i)$ (σ-Subadditivität)

iii) $\mu(A) \leq \mu(B)$, falls $A \subseteq B$ (Monotonie)

iv) Es gilt $P(A) \in [0,1]$

v) $P(A^c) = 1 - P(A)$

vi) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

vii) $P(A \setminus B) = P(A) - P(A \cap B)$.

Beweis: Übungszettel 4.

Definition 4.15 Seien $A_1, A_2, \ldots$ beliebige Mengen. Wir sagen:

- $A_n \nearrow A \;:\Leftrightarrow\; A_1 \subseteq A_2 \subseteq \ldots$ und $\bigcup_{n\geq 1} A_n = A$. (*Konvergenz von oben*)
- $A_n \searrow A \;:\Leftrightarrow\; A_1 \supseteq A_2 \supseteq \ldots$ und $\bigcap_{n\geq 1} A_n = A$. (*Konvergenz von unten*)

Die Pfeile bedeuten hier keine strenge Monotonie.
Außerdem:

- $\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m$ "unendlich viele A_n treten ein"
- $\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{m\geq n} A_m$ "alle bis auf endliche viele A_n treten ein"

Satz 4.16 Seien $\mathcal{A}$ Algebra und μ endlicher Inhalt auf $\mathcal{A}$, dh. $\mu(\Omega) < \infty$. Dann sind äquivalent:

i) μ ist ein Prämaß.

ii) $(A_n)_n \in \mathcal{A}$ mit $A_n \nearrow A \in \mathcal{A} \Rightarrow \mu(A_n) \nearrow \mu(A)$. (*Stetigkeit von unten*)

iii) $(A_n)_n \in \mathcal{A}$ mit $A_n \searrow A \in \mathcal{A} \Rightarrow \mu(A_n) \searrow \mu(A)$. (*Stetigkeit von oben*)

iv) $(A_n)_n \in \mathcal{A}$ mit $A_n \searrow \emptyset \Rightarrow \mu(A_n) \searrow 0$.

v) $(A_n)_n \in \mathcal{A}$ mit $A_n \nearrow \Omega \Rightarrow \mu(A_n) \nearrow \mu(\Omega)$.

Bemerkung 4.17

i) Es existert eine Version von 4.16 ohne 5) für Inhalte auf Ringen.

ii) Die Endlichkeit von μ wird nicht für jede Teilaussage benötigt. Wird die Endlichkeit nicht vorausgesetzt, so gilt:

$$v) \Leftrightarrow i) \Leftrightarrow ii) \Rightarrow iii) \Leftrightarrow iv).$$

Beweis von 4.16:

ii) $\Leftrightarrow$ iii): Es gilt $A_n \nearrow A \Leftrightarrow A_n^c \searrow A^c$ und somit

$$\mu(A_n) \nearrow \mu(A) \quad \Leftrightarrow \quad \mu(A_n^c) = \mu(\Omega) - \mu(A_n) \searrow \mu(\Omega) - \mu(A) = \mu(A^c).$$

iv) $\Leftrightarrow$ v): analog.

v) $\Rightarrow$ ii): Gelte $A_n \nearrow A$. Setze dann $B_n = A_n \cup A^c$. Offenbar $B_n \nearrow (A \cup A^c) = \Omega$. Also folgt:

$$\mu(A_n) + \mu(A^c) = \mu(B_n) \nearrow \mu(\Omega) = \mu(A) + \mu(A^c).$$

ii) $\Rightarrow$ i): Seien $(A_i)_i$ disjunkt mit $A = \bigcup_{i \geq 1} A_i \in \mathcal{A}$. Setze $B_n := \bigcup_{i=1}^n A_i$, dann gilt $B_n \nearrow A$, also $\mu(B_n) \nearrow \mu(A)$. Beachte: μ ist Inhalt, dh. $\mu(B_n) = \sum_{i=1}^n \mu(A_i)$. Da alle Summanden nicht negativ sind, folgt

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

i) $\Rightarrow$ v): Gelte $A_n \nearrow \Omega$. Setze $B_1 := A_1$ und $B_n := A_n \setminus A_{n-1}$, folglich sind die B_i disjunkt und $\Omega = \bigcup_{i\geq 1} B_i$ sowie $A_n = \bigcup_{i=1}^n B_i$. Dann folgt:

$$\mu(A_n) = \sum_{i=1}^n \mu(B_i) \nearrow \sum_{i \geq 1} \mu(B_i) \overset{\text{i)}}{=} \mu\left(\bigcup_{i\geq 1} B_i\right) = \mu(\Omega).$$

$\square$

Definition 4.18 Es seien $(A_n)_n$ und $A \subseteq \Omega$.

i) Die Abbildung

$$1_A : \Omega \to \{0,1\}, \omega \mapsto \begin{cases} 1 & , \omega \in A \\ 0 & , \omega \notin A \end{cases}$$

heißt *Indikatorfunktion von A*.

ii) $A_n \to A :\Leftrightarrow 1_{A_n}(\omega) \to 1_A(\omega) \; \forall \omega \in \Omega$ (punktweise Konvergenz) (*Konvergenz von Mengenfolgen*)

Bem: Indikatorfunktionen charakterisieren die Mengen eindeutig.

Satz 4.19 (Stetigkeit im Maß) Seien $(\Omega, \mathcal{A}, P)$ W'raum und $(A_n)_n \in \mathcal{A}$ mit $A_n \to A$. Dann gilt $A \in \mathcal{A}$ und $P(A_n) \to P(A)$.

Beweis: Es gilt (Übungsaufgabe):

$$1_{\limsup_{n\to\infty}(A_n)}(\omega) = \inf_{n\geq 1} \sup_{m\geq n} 1_{A_m}(\omega) = \limsup_{n\to\infty} 1_{A_n}(\omega) \; \forall \omega \in \Omega$$

und $\limsup_{n\to\infty} A_n \in \mathcal{A}$. Analoges gilt für $\liminf$.
Also:

$$1_A(\omega) = \lim_{n\to\infty} 1_{A_n}(\omega) = \limsup_{n\to\infty} 1_{A_n}(\omega) = 1_{\limsup_n A_n}(\omega) \;\Rightarrow\; A = \limsup_n A_n \in \mathcal{A}.$$

Analog: $A = \liminf_n A_n$. Außerdem:

$$B_n = \bigcap_{m=n}^{\infty} A_m \text{ erfüllt } B_n \nearrow \liminf_n A_n = A$$
$$C_n = \bigcup_{m=n}^{\infty} A_m \text{ erfüllt } C_n \searrow \limsup_n A_n = A.$$

Also folgt mit Satz 4.16 $P(B_n) \nearrow P(A)$ und $P(C_n) \searrow P(A)$ und somit wegen Monotonie:

$$P(B_n) \subseteq P(A_n) \subseteq P(C_n) \;\Rightarrow\; P(A_n) \to P(A).$$

$\square$

5 Fortsetzung von Prämaßen zu Maßen: Existenz

Definition 5.1 Eine Abbildung $\eta : \mathcal{P}(\Omega) \to [0,\infty]$ heißt *äußeres Maß*, falls gilt:

i) $\eta(\emptyset) = 0$.

ii) $A \subseteq B \Rightarrow \eta(A) \leq \eta(B)$. (*Monotonie*)

iii) $\eta\Big(\bigcup_{i\geq 1} A_i\Big) \leq \sum_{i\geq 1} \eta(A_i)$. (*Subadditivität*)

Satz 5.2 Sei $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ mit $\emptyset \in \mathcal{E}$ gegeben und $\mu : \mathcal{E} \to [0,\infty]$ mit $\mu(\emptyset) = 0$. Dann definiert

$$\eta(A) = \eta_\mu(A) := \inf \left\{ \sum_{i\geq 1} \mu(B_n) : B_n \in \mathcal{E}, A \subseteq \bigcup_{n\geq 1} B_n \right\} \qquad \text{mit } \inf \emptyset = \infty$$

ein äußeres Maß.
Beweis:

i) klar, da $\bigcup_{i\geq 1} \emptyset$ Überdeckung von $\emptyset$.

ii) klar: Jede Überdeckung von B ist Überdeckung von A.

iii) Seien A_i gegeben mit $\eta(A_i) < \infty$ für alle i. Zu $\varepsilon > 0$ beliebig existieren dann $B_{ik} \in \mathcal{E}$ mit $A_i \subseteq \bigcup_{k\geq 1} B_{ik}$ und $\sum_{k\geq 1} \mu(B_{ik}) \leq \eta(A_i) + \frac{\varepsilon}{2^i}$. Also $\bigcup_{i\geq 1} A_i \subseteq \bigcup_{i,k} B_{ik}$ und

$$\eta\left(\bigcup_{i\geq 1} A_i\right) \leq \sum_{i,k} \mu(B_{ik}) \leq \sum_{i\geq 1} \left(\mu(A_i) + \frac{\varepsilon}{2^i}\right) = \sum_{i\geq 1} \eta(A_i) + \varepsilon.$$

□

Definition 5.3 Sei $\eta : \mathcal{P}(\Omega) \to [0,\infty]$ ein äußeres Maß. Dann heißt $A \in \mathcal{P}(\Omega)$ *η-messbar*, falls gilt:

$$\eta(Q) = \eta(Q \cap A) + \eta(Q \setminus A) \qquad \text{für alle } Q \in \mathcal{P}(\Omega). \tag{3}$$

Bemerkung 5.4

i) Wegen Subadditivität von η ist (3) äquivalent zu

$$\eta(Q) \geq \eta(Q \cap A) + \eta(Q \setminus A) \qquad \text{für alle } Q \in \mathcal{P}(\Omega).$$

ii) Jede Nullmenge A (dh. $\eta(A) = 0$) erfüllt

$$\eta(Q) = \eta(Q) + \eta(A) \geq \eta(Q \setminus A) + \eta(A \cap Q).$$

$\to$ Nullmengen sind η-messbar.

Definition 5.5 Ein Maßraum $(\Omega, \mathcal{A}, \mu)$ heißt *vollständig*, falls gilt: Ist $A \in \mathcal{A}$ mit $\mu(A) = 0$ und $B \subseteq A$, so folgt $B \in \mathcal{A}$.

Satz 5.6 (Carathéodory) Ist $\eta : \mathcal{P}(\Omega) \to [0,\infty]$ äußeres Maß, so ist $\mathcal{A}_\eta = \{A \subseteq \Omega \mid A \text{ ist } \eta\text{-messbar}\}$ eine σ-Algebra. Die Einschränkung $\eta|_{\mathcal{A}_\eta}$ ist ein Maß. Der Maßraum $(\Omega, \mathcal{A}_\eta, \eta|_{\mathcal{A}_\eta})$ ist vollständig.

Beweis:

i) ($\mathcal{A}_\eta$ ist Algebra). Beachte $Q \setminus A = Q \cap A^c$, also $A \in \mathcal{A}_\eta$. Trivialerweise gilt $\emptyset \in \mathcal{A}_\eta$ (Bem. 5.4 2)), dann auch $\emptyset^c = \Omega \in \mathcal{A}_\eta$.

   Wir zeigen nur noch: $A, B \in \mathcal{A}_\eta \;\Rightarrow\; A \cup B \in \mathcal{A}_\eta$. Dann folgt $A \setminus B = (A^c \cap B)^c \in \mathcal{A}_\eta$. Seien also $A, B \in \mathcal{A}_\eta$ und $Q \subseteq \Omega$. Dann wissen wir:

$$\eta(Q) = \eta(Q \cap A) + \eta(Q \cap A^c) \qquad \text{und} \qquad \eta(Q \cap A^c) = \eta(Q \cap A^c \cap B) + \eta(Q \cap A^c \cap B^c).$$

   Beachte: $A \cup B = A \cup (A^c \cup B)$, also

$$\eta(Q \cap (A \cup B)) \leq \eta(Q \cap A) + \eta(Q \cap A^c \cap B). \tag{4}$$

   Man erhält:

$$\eta(Q) = \eta(Q \cap A) + \eta(Q \cap A^c \cap B) + \eta(Q \cap A^c \cap B^c) \geq \eta(Q \cap (A \cup B)) + \eta(Q \cap (A \cup B)^c).$$

   Also $A \cup B \in \mathcal{A}_\eta$ nach Bem. 5.4 1).

ii) $\big((A_n)_n \in \mathcal{A}_\eta$ disjunkt $\Rightarrow$ $A = \bigcup_{n \geq 1} \in \mathcal{A}_\eta$ und $\eta(A) = \sum_{\substack{n \geq 1 \\ M \in \mathcal{A}_\eta}} \eta(A_n)\big)$. Sind $M, N \subseteq \mathcal{A}_\eta$ disjunkt, so gilt:

$$\eta(Q \cap (M \cup N)) = \eta(Q \cap (M \cup N) \cap M) + \eta(Q \cap (M \cup N) \cap M^c) = \eta(Q \cap M) + \eta(Q \cap N)$$

   Induktiv: $\bigcup_{k \in \underline{n}} A_k \in \mathcal{A}_\eta$ (1. Schritt) und

$$\eta\left(Q \cap \bigcup_{k \in \underline{n}} A_k\right) = \sum_{k \in \underline{n}} \eta(Q \cap A_k). \tag{5}$$

   Daher:

$$\begin{aligned} \eta(Q) &= \eta\left(Q \cap \bigcup_{k \in \underline{n}} A_k\right) + \eta\left(Q \cap \left(\bigcup_{k \in \underline{n}} A_k\right)^c\right) \overset{(5)}{\geq} \sum_{k \in \underline{n}} \eta(Q \cap A^c) + \eta(Q \cup A^c) \\ \Rightarrow\; \eta(Q) &\geq \sum_{i \geq 1} \eta(Q \cap A_k) + \eta(Q \cap A^c) \geq \eta(Q \cap A) + \eta(Q \cap A^c). \end{aligned}$$

   Mit 5.4 i) folgt $A \in \mathcal{A}_\eta$ und in der letzten Ungleichung steht überall "=". Setze dann $Q = A$ ein.

iii) Seien $A_1, A_2, \ldots \in \mathcal{A}_\eta$ beliebig. Dann definiere $B_1 := A_1$ und $B_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$ sind disjunkt, ebenfalls in $\mathcal{A}_\eta$ und $A = \bigcup_{n \geq 1} A_n = \bigcup_{n \geq 1} B_n \in \mathcal{A}_\eta$.

iv) Vollständigkeit folgt aus 5.4 2) und Monotonie. ☐

Wir betrachten nun ein konkretes äußeres Maß.

Satz 5.7 Sei $\mathcal{R}$ ein Ring über Ω, $\mu : \mathcal{R} \to [0,\infty]$ ein Inhalt und η das äußere Maß gemäß 5.2. Dann gilt $\mathcal{R} \subseteq \mathcal{A}_\eta$.

Beweis: Zu zeigen ist: Für $A \in \mathcal{R}$ und $Q \in \Omega$ mit $\eta(Q) < \infty$ gilt: $\eta(Q) \geq \eta(Q \setminus A) + \eta(Q \cap A)$.

Sei also $(B_n)_n \in \mathcal{R}$ mit $Q \subseteq \bigcup_n B_n$. Dann gilt $Q \cap A \subseteq \bigcup_{n\geq 1}(B_n \cap A)$ und $Q \setminus A \subseteq \bigcup_{n \geq 1}(B_n \setminus A)$, also insbesondere $\eta(Q \cap A) \leq \eta\left(\bigcup_{n\geq 1}(B_n \cap A)\right) \leq \sum_{n\geq 1} \eta(B_n \cap A)$.

Außerdem ist $A \cap B_n \in \mathcal{R}$, also nach Def: $\eta(A \cap B_n) \leq \mu(A \cap B_n)$. Analog $Q \setminus A$. Insgesamt folgt

$$\eta(Q \cap A) + \eta(Q \setminus A) \leq \sum_{n\geq 1} (\mu(A \cap B_n) + \mu(B_n \setminus A)) = \sum_{n \geq 1} \mu(B_n).$$

Wegen $\eta(Q) = \inf\{\sum_n \mu(B_n) \mid Q \subseteq \bigcup_n B_n\}$ folgt die Behauptung. □

Korollar 5.8 Sei $\mathcal{R}$ Ring über Ω und $\mu : \mathcal{R} \to [0,\infty]$ ein Prämaß. Dann ist $\eta_\mu : \mathcal{A}_{\eta_\mu} \to [0,\infty]$ eine Fortsetzung von μ, dh.

i) $\mathcal{R} \subseteq \mathcal{A}_{\eta_\mu}$

ii) $\eta_\mu(A) = \mu(A)$ für alle $A \in \mathcal{R}$

iii) $\left(\Omega, \mathcal{A}_{\eta_\mu}, \eta_\mu\big|_{\mathcal{A}_{\eta_\mu}}\right)$ ist vollständig.

Insbesondere $\sigma(\mathcal{R}) \subseteq \mathcal{A}_{\eta_\mu}$, dh. man erhält immer ein Maß au der von $\mathcal{R}$ erzeugten σ-Algebra.

Beweis: i) und iii) wurden schon vorher bewiesen (5.7 bzw. 5.6).

ii): Immer gilt $\eta_\mu(A) \leq \mu(A)$, da $A, \emptyset, \emptyset, \ldots$ A überdeckt. Umgekehrt: μ ist Prämaß, also folgt aus $A \subseteq \bigcup_n B_n$ mit $B_n \in \mathcal{R}$, dass $A \in \mathcal{R}$. Dann folgt $\mu(A) \leq \sum_{n\geq 1} \mu(B_n)$. Also

$$\mu(A) \leq \inf\left\{\sum_n \mu(B_n) : A \subseteq \bigcup_n B_n\right\} = \eta_\mu(A).$$

□

6 Fortsetzung von Prämaßen zu Maßen: Eindeutigkeit

Beispiel 6.1 Es sei $\mathcal{R} = \{\emptyset\}$ und μ ein Prämaß auf $\mathcal{R}$, dh. $\mu(\emptyset) = 0$. Offenbar $\sigma(\mathcal{R}) = \{\emptyset, \Omega\}$. Für jede Wahl von $\mu(\Omega) \in [0, \infty]$ ist $(\Omega, \sigma(\mathcal{R}), \mu)$ ein Maßraum. Folglich ist die Fortsetzung von μ nicht eindeutig.

Bemerkung 6.2 Wir möchten klären, in welchen Fällen die Fortsetzungen eindeutig sind, dh. wann

$$\mu(A) = \nu(A) \qquad \text{für alle } A \in \mathcal{R} \tag{6}$$

auch $\mu(A) = \nu(A)$ für alle $A \in \sigma(\mathcal{R})$ impliziert.

Idee: Kann man zeigen, dass

$$\{A \subseteq \Omega \mid (6) \text{ gilt}\} \tag{7}$$

eine σ-Algebra bildet, so folgt (6) für alle $A \in \sigma(\mathcal{R})$, da (6) für alle $A \in \mathcal{R}$ gilt und $\sigma(\mathcal{R})$ die kleinste σ-Algebra ist, die $\mathcal{R}$ enthält.

Problem: Oft ist nicht leicht zu zeigen, daß (7) eine σ-Algebra ist.

Definition 6.3 Eine Teilmenge $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ heißt *Dynkin-System*, falls gilt:

i) $\Omega \in \mathcal{D}$

ii) $A \in \mathcal{D} \;\Rightarrow\; A^c \in \mathcal{D}$

iii) Sind $(A_i)_i \in \mathcal{D}$ disjunkt, so gilt auch $\biguplus_i A_i \in \mathcal{D}$.

Wann ist ein Dynkin-System eine σ-Algebra?

Definition 6.4 Ein Mengensystem $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ heißt *durchschnittstabil*, falls gilt:

$$A, B \in \mathcal{E} \;\Rightarrow\; A \cap B \in \mathcal{E}.$$

Satz 6.5 Ist ein Dynkin-System $\mathcal{D}$ durchschnittstabil, so ist es eine σ-Algebra.

Beweis: Sei $(A_i)_i$ Folge mit $A_i \in \mathcal{D}$. Setze $B_1 := A_1$ und $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i$. Zeige induktiv: $B_n \in \mathcal{D}$ und $\bigcup_{i \in \underline{n}} A_i \in \mathcal{D}$. Dann $\bigcup_{i \geq 1} B_1 \in \mathcal{D}$, weil alle $B_i \in \mathcal{D}$ und disjunkt.

$n - 1$: klar, da $B_1 = A_1 \in \mathcal{D}$.

$n - 1 \;\Rightarrow\; n$: Es gilt $B_n = A_n \cap (A_1 \cup \ldots \cup A_{n-1})^c$ und nach IV ist $A_1 \cup \ldots \cup A_{n-1} \in \mathcal{D}$, folglich $B_n \in \mathcal{D}$, da $\mathcal{D}$ durchschnittstabil. Außerdem:

$$A_1 \cup \ldots \cup A_n = \underbrace{(A_1 \cup \ldots \cup A_{n-1})}_{\in \mathcal{D}} \cup \underbrace{B_{n-1}}_{\in \mathcal{D}} \in \mathcal{D}$$

□

Satz und Definition 6.6 Es sei $\mathcal{C}$ Mengensystem und $d(\mathcal{C}) = \bigcap\{\mathcal{D} : \mathcal{C} \subset \mathcal{D}, \mathcal{D} \text{ Dynkin-System}\}$ das von $\mathcal{C}$ *erzeugtes Dynkin-System.* Ist $\mathcal{C}$ durchschnittstabil, so gilt $d(\mathcal{C}) = \sigma(\mathcal{C})$.

Beweis: Trivial: $d(\mathcal{C}) \subseteq \sigma(\mathcal{C})$, da jede σ-Algebra ein Dynkin-System ist. Bleibt zu zeigen: $\sigma(\mathcal{C}) = d(\mathcal{C})$. Es genügt zu zeigen, dass $d(\mathcal{C})$ eine σ-Algebra ist, denn dann schneidet man in der Definition von $\sigma(\mathcal{C})$ auch über $d(\mathcal{C})$. Nach 6.5 reicht es zu zeigen, dass $d(\mathcal{C})$ durchschnittstabil ist.

Setze

$$\mathcal{A} := \{A \subseteq \Omega : A \cap C \in d(\mathcal{C})\ \forall C \in \mathcal{C}\}.$$

Zeige $\mathcal{A}$ ist Dynkin-System.

- $\Omega \in \mathcal{A}$, da $\Omega \cap C = C$ und somit $C \in d(\mathcal{C})$.
- Sei $A \in \mathcal{A}$, also gilt $A \cap \Omega \in d(\mathcal{C})$ für alle $C \in \mathcal{C}$. Es gilt $A^c \cap C = (C^c \cup (A \cap C))^c$. Es sind C^c und $A \cap C$ disjunkt und $C^c, A \cap C \in d(\mathcal{C})$, also $C^c \cup (A \cap C) \in d(\mathcal{C})$ und somit auch das Komplement. Folglich $A^c \cap C \in d(\mathcal{C})$.
- Sei $(A_n)_n \in \mathcal{A}$ disjunkte Folge aus $\mathcal{A}$, dh. $A_n \cap C \in d(\mathcal{C})$ für alle $n \in \mathbb{N}$. Zu zeigen $\bigcup_n A_n \cup C \in d(\mathcal{C})$. Es gilt $\bigcup_n A_n \cap C = \bigcup_n \underbrace{(A_n \cap C)}_{\in d(\mathcal{C})} \in d(\mathcal{C})$, da $A_n \cap C$ disjunkt für alle $n \in \mathbb{N}$.

Insgesamt folgt, dass $\mathcal{A}$ Dynkin-System ist.

Außerdem: $\mathcal{C} \subseteq \mathcal{A}$, da $\mathcal{C}$ durchschnittstabil, also $d(\mathcal{C}) \subseteq \mathcal{A}$. Setze weiter

$$\bar{\mathcal{A}} := \{A \subseteq \Omega : A \cap B \in d(\mathcal{C})\ \forall B \in d(\mathcal{C})\}.$$

Mit analoger Argumentation wie oben wird gezeigt, dass $\bar{\mathcal{A}}$ Dynkin-System ist. Zudem: $\mathcal{C} \subseteq \bar{\mathcal{A}}$, da $d(\mathcal{C}) \subseteq \mathcal{A}$ ist. Folglich $d(\mathcal{A}) \subseteq \bar{\mathcal{A}}$, also ist $d(\mathcal{C})$ durchschnittstabil. □

Korollar 6.7 (Dynkin-System-Argument) Seien $(\Omega, \mathcal{A})$ Messraum, $\mathcal{E} \in \mathcal{P}(\Omega)$ durchschnittstabil und $\mathcal{A} = \sigma(\mathcal{E})$. Gilt eine Aussage A für alle $E \in \mathcal{E}$ und bildet $\mathcal{D} := \{E \subseteq \mathcal{A} : A(E) \text{ wahr}\}$ ein Dynkin-System, so gilt A für alle $E \subseteq \mathcal{A}$.

Beweis: Es gilt $\mathcal{E} \subseteq \mathcal{D} \subseteq \mathcal{A}$ nach Voraussetzung und $d(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}$ nach 6.6. Es folgt:

$$\mathcal{A} = d(\mathcal{E}) \subseteq \mathcal{D} \subseteq \mathcal{A}.$$

□

Definition 6.8 Ein Maß μ auf $(\Omega, \mathcal{A})$ heißt σ-*endlich* über $\mathcal{E} \subseteq \mathcal{A}$, falls eine Folge $(E_n)_n$ aus $\mathcal{E}$ existiert mit $E_n \nearrow \Omega$ und $\mu(E_n) < \infty$ für alle $n \in \mathbb{N}$ gilt.

Beispiel: Das Maß aus 6.1 und das Maß ν, gegeben $\nu(\emptyset) = 0$ und $\mu(A) = \infty$, sind nicht σ-endlich.

Satz 6.9 (Eindeutigkeitssatz) Seien μ und ν Maße auf $(\Omega, \mathcal{A})$ und $\mathcal{E}$ ein durchschnittstabiler Erzeuger von $\mathcal{A}$. Gilt:

i) $\mu(A) = \nu(A)$ für alle $A \in \mathcal{E}$ und

ii) μ und ν sind σ-endlich,

so folgt $\mu = \nu$, dh. $\mu(A) = \nu(A)$ für alle $A \in \mathcal{A}$.

Beweis: Seien $(E_n)_n, (F_n)_n$ Folgen aus $\mathcal{E}$ mit $E_n \nearrow \Omega$ und $F_n \nearrow \Omega$. Dann auch $E_n \cap F_n \in \mathcal{E}$ und $(E_n \cap F_n) \nearrow \Omega$. Wähle also ohne Einschränkung dieselbe Folge für μ und ν (sind die beiden Folgen nicht gleich, wähle die Schnittfolge) in 6.8. Beachte mit 4.16: $\mu(\Omega) = \lim_{n\to\infty} \mu(E_n) = \lim_{n\to\infty} \nu(E_n) = \nu(\Omega)$.

1. Fall: $\mu(\Omega) = \nu(\Omega) < \infty$. Zeige $\mathcal{D} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$ ist Dynkin-System.

- $\Omega \in \mathcal{D}$: s.o.
- Sei $A \in \mathcal{D}$, so gilt $\mu(A^c) = \mu(\Omega) - \mu(A) = \nu(\Omega) - \nu(A) = \nu(A^c)$, folglich $A^c \in \mathcal{D}$.
- Seien $(A_i)_i \in \mathcal{D}$ disjunkt. Dann folgt:
$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i) = \sum \nu(A_i) = \nu\left(\bigcup_i A_i\right).$$

Es folgt, dass $\mathcal{D}$ Dynkin-System. Behauptung folgt mit Dynkin-System-Argument.

2. Fall: $\mu(\Omega) = \nu(\Omega) = \infty$. Setze $\mu_n(A) := \mu(A \cap E_n)$ und $\nu_n(A) := \nu(A \cap E_n)$ für alle $A \in \mathcal{A}$. Dann gilt: μ_n und ν_n sind endliche Maße (da $\mu(E_n) < \infty$), also $\mu_n = \nu_n$ für alle $n \in \mathbb{N}$ nach Fall 1.

Es folgt mit 4.16: $\mu(A) = \lim_{n\to\infty} \mu(A \cap E_n) = \lim_{n\to\infty} \nu(A \cap E_n) = \nu(A)$. □

7 Wahrscheinlichkeitsmaße über $\mathbb{R}$

Eine wichtige Klasse von Zufallsexperimenten sind solche mit Werten in den reellen Zahlen, dh. wir interessieren uns für den Messraum $(\mathbb{R}, \mathcal{B})$ mit $\mathcal{B}$ aus Beispiel 4.6.

Definition 7.1 Es sei P W'maß auf $(\mathcal{R}, \mathcal{B})$. Dann heißt $F : \mathbb{R} \to [0,1], x \mapsto P\big((-\infty, x]\big)$ die *Verteilungsfunktion von P*.

Satz 7.2 Die Verteilungsfunktion legt P eindeutig fest.

Beweis: Seien P und Q W'maße mit Verteilungsfunktion F. Beachte: $\{(-\infty, a] : a \in \mathbb{R}\}$ erzeugt $\mathcal{B}$ nach 4.7 (es reicht sogar, die Verteilungsfunktion nur für alle $a \in \mathbb{Q}$ zu kennen) und ist durchschnittstabil. Außerdem gilt $P(\Omega) = Q(\Omega) = 1$ und es existieren Folgen $(a_n)_n$ mit $(-\infty, aa_n] \nearrow \mathbb{R}$ (wähle bspw. $a_n = n$). Die Behauptung folgt mit 6.9. □

Im Folgenden interessieren wir uns für Bedingungen unter denen $F : \mathbb{R} \to [0,1]$ ein W'maß P definiert, sd. F die zugehörige Verteilungsfunktion ist.

Notwendige Bedingungen:

- F ist monoton wachsend.
- $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to\infty} F(x) = 1$. (Satz 4.16)

Satz 7.3 Sei $\mathcal{I} = \{(a,b] : a \le b\}$ die Menge aller halboffenen Intervalle und $\mathcal{R}(\mathcal{I})$ der Ring der Figuren aus Beispiel 4.11 (iv).

i) Ist $\mu : \mathcal{R}(\mathcal{J}) \to [0, \infty]$ ein endlicher Inhalt, so ist

$$F : \mathbb{R} \to \mathbb{R}, x \mapsto \begin{cases} \mu((0,x]) & , x \ge 0 \\ -\mu((x,0]) & , x < 0 \end{cases}$$

monoton wachsend.

ii) Sei $F : \mathbb{R} \to \mathbb{R}$ monoton wachsend. Dann lässt sich jedes $A \in \mathcal{R}(\mathcal{J})$ eindeutig als $A = \bigcup_{i=1}^n (a_i, b_i]$ mit disjunkten $(a_i, b_i]$ mit $b_i \neq a_{i+1}$ darstellen und $\mu(A) = \sum_{i=1}^n (F(b_i) - F(a_i))$ definiert einen endlichen Inhalt auf $\mathcal{R}(\mathcal{J})$.

(iii) Sind F und G monoton wachsend, so definieren sie dasselbe μ genau dann, wenn $F - G = \text{const.}$.

Beweis:

i) Fallunterscheidung:

$$\begin{aligned} 0 \le x \le y :&\ \ \mu((0,y]) = \mu((0,x]) + \mu((x,y]) \ge \mu((0,x]) \\ x \le y < 0 :&\ \ -\mu((x,0]) = -\mu((x,y]) - \mu((y,0]) \le -\mu((y,0]) \\ x < 0 \le y :&\ \ -\mu((x,0]) \le 0 \le \mu((0,y]) \end{aligned}$$

ii) Zeige zunächst die Darstellung: Sei $A = \bigcup_{j=1}^{m}(c_j, d_j]$ (oBdA $d_1 \leq \cdots \leq d_m$). Induktion nach m:

$$A = \bigcup_{j=1}^{m}(c_j, d_j] = \bigcup_{i=1}^{n}(a_i, b_i] \text{ mit disjunkten } (a_i, b_i] \text{ für alle } i \in \underline{n} \text{ und } b_1 \leq \cdots \leq b_n.$$

$m = 1$ ist klar. Für $m \to m+1$ führen wir eine Fallunterscheidung:

$$c_{m+1} \geq b_n: \quad A = \bigcup_{i=1}^{n}(a_i, b_i] \cup (c_{m+1}, d_{m+1}]$$
$$b_n > c_{m+1} \geq a_n: \quad A = \bigcup_{i=1}^{n-1}(a_i, b_i] \cup (a_n, d_{m+1}] \text{ usw.}$$

Ansonsten gilt: $\mu(A) \geq 0$, da F monoton wachsend und $\mu(\emptyset) = 0$, da $\emptyset = (a, a]$. Folglich ist

$$\mu(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mu(A_i)$$

für $A_1, \ldots, A_n$ disjunkt nach Konstruktion.

iii) Sei $a \in \mathbb{R}$ beliebig dann gilt für alle $x \geq a$:

$$\mu_F = \mu_G \;\Rightarrow\; \mu_F((a, x]) = \mu_G((a, x]) \;\Leftrightarrow\; F(x) - F(a) = G(x) - G(a) \;\Leftrightarrow\; F(x) - G(x) = F(a) - G(a).$$

Umgekehrt: Ist $F - G = \text{const.}$, so ist $\mu_F = \mu_G$ nach Definition in ii).

Satz 7.4 Sei $F : \mathbb{R} \to \mathbb{R}$ monoton und μ_F der zugehörige endliche Inhalt aus Satz 7.3. Dann ist μ_F genau dann ein Prämaß auf dem $\mathcal{R}(\mathcal{J})$, falls F rechtsseitig stetig ist.

Für den Beweis von Satz 7.4 benötigen wir ein Hilfsresultat.

Definition 7.5 Ein Mengensystem $\mathcal{K} \subseteq \mathcal{P}(\Omega)$ heißt kompakte Klasse, falls aus $(K_i)_{i \in \mathbb{N}} \in \mathcal{K}$ mit $\bigcap_{i \in \mathbb{N}} K_i = \emptyset$ folgt, dass endlich viele $K_{j_1}, \ldots, K_{j_r} \in \mathcal{K}$ existieren mit $K_{j_1} \cap \cdots \cap K_{j_r} = \emptyset$.

Bemerkung 7.6 Im $\mathbb{R}^d$ ist jede Teilmenge der kompakten Mengen eine kompakte Klasse.

Beweis: Übung.

Satz 7.7 Sei $\mathcal{R}$ ein Ring über Ω und $\mu : \mathbb{R} \to [0, \infty]$ ein endlicher Inhalt. Existiert eine kompakte Klasse $\mathcal{K} \subseteq \mathcal{P}(\Omega)$, so dass für alle $\varepsilon > 0$ und $A \in \mathcal{R}$ ein $K \in \mathcal{K}$ und $B \in \mathcal{R}$ existieren mit $B \subseteq K \subseteq A$ und $\mu(A \setminus B) < \varepsilon$, dann ist μ ein Prämaß.

Beweis: Wir zeigen mit Satz 4.16: $A_n \searrow \emptyset$ impliziert $\mu(A_n) \longrightarrow 0$, d. h. zu zeigen ist: Für alle $\varepsilon > 0$ existiert ein $n_0 \in \mathbb{N}$ mit $\mu(A_{n_0}) < \varepsilon$. Wähle zu jedem $n \in \mathbb{N}$ Mengen $B_n \subseteq K_n \subseteq A_n$ und $\mu(A_n \setminus B_n) < \frac{\varepsilon}{2^n}$. Dann:

$$\bigcap_{i \in \mathbb{N}} K_i = \emptyset, \text{ d.h. es existiert ein } n_0 \in \mathbb{N} \text{ mit } \bigcap_{i \in \mathbb{N}}^{n_0} K_i = \emptyset. \Rightarrow B_1 \cap \cdots \cap B_{n_0} = \emptyset.$$

Wir erhalten:

$$\begin{aligned}\mu(A_{n_0}) &= \mu(\bigcap_{k=1}^{n_0} B_k) + \mu(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k) = \mu(\bigcup_{k=1}^{n_0}(A_{n_0} \setminus B_k)) \\ &\leq \mu(\bigcup_{k=1}^{n_0}(A_k \setminus B_k)) \leq \sum_{k=1}^{n_0} \mu(A_k \setminus B_k) < \sum_{k=1}^{n_0} \frac{\varepsilon}{2^k} < \varepsilon.\end{aligned}$$

□

Beweis von Satz 7.4:
"$\Rightarrow$": Sei μ_F ein Prämaß. Wähle dann zu $b \in \mathbb{R}$ beliebig eine Folge $b_n \searrow b$. Klar:

$$\bigcap_{n\in\mathbb{N}}(b, b_n] = \emptyset \overset{4.16}{\Rightarrow} 0 \leftarrow \mu_F((b, b_n]) = F(b_n) - F(b) \Rightarrow \text{F ist rechtsstetig.}$$

"$\Leftarrow$": Seien $\varepsilon > 0$ und $(a, b] \in \mathcal{I}$ mit $b - a > 0$. Da F rechtsstetig in a ist, existiert $\delta > 0$ mit $\delta < b - a$, so dass $F(a + \delta) - F(a) < \varepsilon$. Also:

$$(a + \delta, b] \subseteq [a + \delta, b] \subseteq (a, b] \text{ und } \mu_F((a, b] \setminus (a + \delta, b]) = \mu_F((a, a + \delta)) = F(a + \delta) - F(a) < \varepsilon$$

Analog: Ist $A = \biguplus_{i=1}^{n}(a_i, b_i]$ disjunkte Vereinigung, approximiere jedes $(a_i, b_i]$ mit Güte $\frac{\varepsilon}{n}$. Die Aussage folgt dann aus Satz 7.7. □

Korollar 7.8 $F : \mathbb{R} \to [0, 1]$ ist genau dann die Verteilungsfunktion eines W'maßes auf $(\mathbb{R}, \mathcal{B})$, falls gilt:

- F ist monoton wachsend.
- F ist rechtsstetig.
- Es gilt: $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$.

Beweis: Satz 7.3, Satz 7.4 und Korollar 5.8. □

Lemma 7.9 Sei F die Verteilungsfunktion von P auf $(\mathbb{R}, \mathcal{B})$. Dann gilt:

i) $P((x, y]) = F(y) - F(x)$.
ii) $P([x, y]) = F(y) - F(x-)$ mit $F(x-) = \lim_{z\nearrow x} F(z)$.
iii) $P([x, y)) = F(y-) - F(x-)$.
iv) $P((x, y)) = F(y-) - F(x)$.
v) $P(\{x\}) = F(x) - F(x-)$.

Insbesondere: F stetig $\Leftrightarrow P(\{x\}) = 0$ für alle $x \in \mathbb{R}$.
Beweis: Zeige nur (i) und (ii).

i)

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

ii) Mit Stetigkeit von unten folgt:

$$P([x,y]) = \lim_{n\to\infty} P\left(\left(x-\frac{1}{n},y\right]\right) = \lim_{n\to\infty}\left(F(y)-F\left(x-\frac{1}{n}\right)\right) = F(y) - \lim_{n\to\infty} F\left(x-\frac{1}{n}\right).$$

□

Beispiel 7.10

i) Ist $\Omega \subseteq \mathbb{R}$ diskret und P' W'maß auf Ω. Dann wird durch $P(B) = P'(B\cap\Omega)$ für alle $B \subseteq \mathbb{R}$ ein W'maß P auf $\mathbb{R}$ definiert. Die zugehörige Treppenfunktion ist stückweise konstant mit Sprunghöhen $F(x) - F(x-) = P'(\{x\})$ für alle $x \in \Omega$.

   Bsp: $\Omega = \{0, 1, 2\}$, P Laplace-Maß.

ii) Sei $z \in \mathbb{R}$. Dann heißt

$$\delta_z : \mathcal{B} \to [0,1], B \mapsto \delta_z(B) = \begin{cases} 1 & , z \in B \\ 0 & , \text{sonst} \end{cases}$$

   das *Dirac-Maß zu z*.

iii) Für jede nicht-negative L-integrierbare Abbildung $f : \mathbb{R} \to [0,\infty)$ mit $\int_{-\infty}^{\infty} f(t)dt = 1$ definiere $F : \mathbb{R} \to [0,1], x \mapsto \int_{-\infty}^{x} f(t)dt$. Dann ist F Verteilungsfunktion zu W'maß P auf $\mathbb{R}$. f heißt dann *Dichte* (genauer *Lebesgue-Dichte*) von P.

   Beweis: Satz 7.8, denn F erfüllt die dort gegebenen Bedingungen.

   (a) *Gleichverteilung auf* $[a,b]$ (Verallgemeinerung der Laplace-Verteilung)

$$f(x) = \begin{cases} 0 & , x \notin [a,b] \\ c & , \text{sonst} \end{cases} \quad \text{für } c = \frac{1}{b-a} \quad \Rightarrow \quad F(x) = \begin{cases} 0 & , x < a \\ \frac{x-a}{b-a} & , x \in [a,b] \\ 1 & , x > b \end{cases}$$

   Notation: $U(a,b)$

   (b) *Exponentialverteilung* zum Parameter $\lambda > 0$

$$f(x) = \begin{cases} 0 & , x < 0 \\ \lambda e^{-\lambda x} & , x \geq 0 \end{cases} \quad \Rightarrow \quad F(x) = \begin{cases} 0 & , x < 0 \\ 1 - e^{-\lambda x} & , x \geq 0 \end{cases}$$

   Notation: $EXP(\lambda)$
   Anwendung: Lebensdauermodellierung

   (c) *Gammaverteilung* zu den Parametern $\alpha, \beta > 0$

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & , x \geq 0 \\ 0 & , \text{sonst} \end{cases} \qquad \text{mit } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

   Es gilt: $\Gamma(1,\beta) = EXP(\beta)$ (hierbei ist Γ die Gammaverteilung)
   Notation: $\Gamma(\alpha,\beta)$
   Anwendung: Wartezeitmodellierung (Lebensdauer mit Gedächtnis)

(d) *Normalverteilung* zu den Parametern $\mu \in \mathbb{R}, \sigma^2 > 0$

$$f(x) = \varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Notation: $\mathcal{N}(\mu, \sigma^2)$
Anwendung: Grenzwertsatz, angewandte Statistik

(e) *Log-Normalverteilung* zu den Parametern $\mu \in \mathbb{R}, \sigma^2 > 0$

$$f(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{1}{x}\varphi_{\mu,\sigma^2}(\log x) & , \text{sonst} \end{cases}$$

Anwendung: Aktienkursmodellierung

(f) *Cauchy-Verteilung* mit Parametern $\alpha \in \mathbb{R}, \beta > 0$

$$f(x) = \frac{1}{\beta\pi} \cdot \frac{1}{1 + \frac{(x-\alpha)^2}{\beta}}$$

Notation: $Cauchy(\alpha, \beta)$
Anwendung: Gegenbeispiele

8 Messbare Funktionen

Bemerkung 8.1 Im diskreten Fall ist eine Zufallsgröße eine Abbildung $X : \Omega \to \Omega'$. Ist P W'maß auf Ω, dann wird P^X definiert als W'maß auf Ω' durch

$$P^X(B) := P(X \in B) = P(\underbrace{\{\omega \in \Omega : X(\omega) \in B\}}_{(*)}) = P(X^{-1}(B)).$$

Problem bei Übertragung auf allgemeinen Fall: Ist $(*)$ ein Element der σ-Algebra auf Ω, dh. ist $P^X(B)$ überhaupt "wohldefiniert"?

Definition 8.2 Es seien $(E, \mathcal{E})$ und $(F, \mathcal{F})$ Messräume. Eine Abbildung $X : E \to F$ heißt *($\mathcal{E}$-$\mathcal{F}$)-messbar*, falls gilt $\forall B \in \mathcal{F} : X^{-1}(B) =: \{X \in B\} \in \mathcal{E}$.

Gilt $(E, \mathcal{E}) = (\Omega, \mathcal{A})$ und ist $(\Omega, \mathcal{A}, P)$ W'raum, so heißt X auch *Zufallsvariable* oder *Zufallsgröße*.

Bemerkung 8.3 Situation oft: $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$ oder $(F, \mathcal{F}) = \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$, wobei $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ und $\overline{\mathcal{B}} = \{B \in \overline{\mathbb{R}} : B \cap \mathbb{R} \in \mathcal{B}\}$ (Kompaktifizierung). Zum Teil ist der Begriff Zufallsgröße für diesen Fall reserviert (Zufallsvariable beschreibt dann den allgemeineren Fall).

Satz 8.4 Es seien $(E, \mathcal{E})$ und $(F, \mathcal{F})$ Messräume, $\mathcal{C} \subseteq \mathcal{F}$ und $\mathcal{F} = \sigma(\mathcal{C})$. Dann gilt:

$$X : E \to F \text{ messbar } \Leftrightarrow \forall B \in \mathcal{C} : X^{-1}(B) \in \mathcal{E} \tag{8}$$

Beweis:
"$\Rightarrow$": trivial
"$\Leftarrow$": Es gelte die rechte Seite von (8). Setze $\mathcal{F}' := \{B \in \mathcal{F} : X^{-1}(B) \in \mathcal{E}\}$. Zu zeigen ist $\mathcal{F}' = \mathcal{F}$. Offensichtlich gilt $\mathcal{F}' \subseteq \mathcal{F}$. Es gilt $\mathcal{C} \in \mathcal{F}'$ (nach (8)). Ferner ist $\mathcal{F}'$ σ-Algebra (als Übung zu zeigen). Also folgt:

$$\mathcal{F} \overset{[\text{Vor.}]}{=} \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{F}') = \mathcal{F}'.$$

$\square$

$\to$ Messbarkeit auf Erzeugendensystem reicht.

Korollar 8.5 Sei $(E, \mathcal{E})$ Messraum und $X : E \to \overline{\mathbb{R}}$ Abbildung. X ist genau dann messbar, wenn gilt:

$$\{X \leq a\} = \{\omega \in E : X(\omega) \leq a\} \in \mathcal{E} \qquad \text{für alle } a \in \mathbb{R}.$$

Beweis:$\square$

Bemerkung 8.6 Analoges gilt für $\{X < a\} = X^{-1}([-\infty, a)), \{X \geq a\}, \ldots$

Beweis: $\{X \leq a\} = X^{-1}\left((-\infty, a]\right)$ und $\mathcal{C} = \{(-\infty, a] : a \in \mathbb{R}\}$ erzeugt $\overline{\mathcal{B}}$, analog zu 4.7.

Satz 8.7 Sei $(E, \mathcal{E})$ Messraum und $(X_n)_{n\in\mathbb{N}}$ messbare Abbildungen von E nach $\overline{\mathbb{R}}$. Es gilt:

i) $\sup_{n\in\mathbb{N}} X_n, \inf_{n\in\mathbb{N}} X_n, \limsup_{n\in\mathbb{N}} X_n, \liminf_{n\in\mathbb{N}} X_n$ sind messbar.

ii) Falls $X_n(\omega) \to X(\omega)$ punktweise konvergiert für alle $\omega \in \Omega$, so ist auch X messbar.

Achtung: Ist $(X_i)_{i\in I}$ eine Familie von messbaren Abbildungen wie oben, so ist $\sup_{i\in I} X_i$ iA. nicht messbar (I überabzählbar).
Exkurs: Ist $\mathcal{V} \subseteq \mathbb{R}$ eine nicht-Borelmenge, dh. $\mathcal{V} \in \mathcal{P}(\mathbb{R}) \setminus \mathcal{B}$ und für alle $i \in V$ sei $X_i = 1_{\{i\}}$. Diese sind messbar (s.u.), aber $\sup_{i\in I} X_i = 1_{\mathcal{V}}$ und, da $\mathcal{V} \notin \mathcal{B}$ ist 1_V nicht messbar (vgl. 8.11).

Es folgt insbesondere, dass nichtmessbare Mengen nicht abzählbar sind.

Beweis:

i) Nach 8.5 genügt es zu zeigen, dass $\{\sup_n X_n > a\} \in \mathcal{E}$ für alle $a \in \mathbb{R}$. Es gilt:

$$\{\sup_n X_n > a\} = \{\omega \mid \exists n \in \mathbb{N} : X_n(\omega) > a\} = \bigcup_{n\in\mathbb{N}} \underbrace{\{\omega : X_n(\omega) > a\}}_{\in\mathcal{E}} \in \mathcal{E},$$

da X_n messbar und $\mathcal{E}$ σ-Algebra. Weiter ist:

$$\{\limsup_n X_n > a\} = \{\omega \mid \exists n \in \mathbb{N} \forall k \geq n : X_n(\omega) > a\} = \bigcap_{n\in\mathbb{N}} \bigcup_{\substack{k\geq n\\ k\in\mathbb{N}}} \underbrace{\{X_k > a\}}_{\in\mathcal{E}} \in \mathcal{E}$$

(s. Bem. oben).

ii) Es gilt: $X = \limsup_n X_n = \liminf_n X_n$. Behauptung folgt mit i). □

Lemma 8.8 Seien $X : (E, \mathcal{E}) \to (F, \mathcal{F})$ und $Y : (F, \mathcal{F}) \to (G, \mathcal{G})$ messbare Abbildungen. Dann ist auch $Y \circ X$ messbar.

Beweis: Sei $A \in \mathcal{G}$. Es gilt: $(Y \circ X)^{-1}(A) = X^{-1}(\underbrace{Y^{-1}(A)}_{\in\mathcal{F}}) \in \mathcal{E}$. □

Definition 8.9 Sei E eine Menge. $\mathcal{U} \subseteq \mathcal{P}(E)$ heißt *Topologie*, falls gilt:

i) $\forall I\ \forall (U_i)_{i\in I} \in \mathcal{U}^I : \bigcup_{i\in I} U_i \in \mathcal{U}$. (beliebige Vereinigungen)

ii) $\forall (U_i)_{i\in \underline{n}} \in \mathcal{U}^n : \bigcap_{i\in I} U_i \in \mathcal{U}$ mit $n \in \mathbb{N}$. (endliche Schnitte)

Die Elementen von $\mathcal{U}$ nennt man *offene Mengen*, $(E,\mathcal{U})$ *topologischen Raum*.

Weiter heißt $\mathcal{E} := \sigma(\mathcal{U})$ *Borel-σ-Algebra*.

Bem: Das Konzept des topologischen Raums ist eine Verallgemeinerung des Konzepts des metrischen Raums.

Satz 8.10 Seien $(E,\mathcal{U})$ und $(F, \mathcal{V})$ topologische Räume, sowie $\mathcal{E} = \sigma(\mathcal{U})$ und $\mathcal{F} = \sigma(\mathcal{V})$ die Borel-σ-Algebren. Ist eine Abbildung $X : E \to F$ stetig, dann ist X messbar.

Beweis: Es gilt: X stetig $\Leftrightarrow$ $\forall V \in \mathcal{V} : X^{-1}(V) \in \mathcal{U}$. Mit Satz 8.4 folgt die Behauptung. □

Korollar 8.11 Sei $(E, \mathcal{E})$ Messraum und $(F, \mathcal{F}) \in \{(\mathbb{R}, \mathcal{B}), (\overline{\mathbb{R}}, \overline{\mathcal{B}})\}$.

i) Es gilt für alle $A \subseteq E : 1_A$ ist messbar $\Leftrightarrow$ $A \in \mathcal{E}$.

ii) Sind $X_1, \ldots, X_n : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B})$ und $f : (\mathbb{R}^n, \mathcal{B}^n) \to (F, \mathcal{F})$ messbar, so ist auch $f(X_1, \ldots, X_n) : E \to F, x \mapsto f(X_1(x), \ldots, X_n(x))$ messbar.

iii) Sind $X, Y : (E, \mathcal{E}) \to (F, \mathcal{F})$ messbare Abbildungen, so sind auch $X+Y, X \cdot Y, X \vee Y := \max\{X, Y\}$, $X \wedge Y := \min\{X, Y\}$ und $\frac{X}{Y}$ (soweit definiert) messbar.

Beweis:

i) Sei $B \in \mathcal{B}$. Es gilt:

$$(1_A)^{-1}(B) = \begin{cases} \emptyset & \text{falls } 0, 1 \notin B \\ E & \text{falls } 0, 1 \in B \\ A^c & \text{falls } 0 \in B \wedge 1 \notin B \\ A & \text{sonst} \end{cases},$$

da $\emptyset, E \in \mathcal{E}$ und $A^c \in \mathcal{E} \;\Rightarrow\; A \in \mathcal{E}$ folgt.

ii) Unter Benutzung von 8.8 ist nur zu zeigen: $(X_1, \ldots, X_n) \,:\, E \to \mathbb{R}^n$ ist messbar. Wie in 4.7 gilt: $\mathcal{B}^n = \mathcal{B}(\mathbb{R}^n)$ ist erzeugt von den Figuren $\prod_{i=1}^n (-\infty, a_i]$ mit $a_i \in \mathbb{R}$. Mit 8.4 ist zu zeigen $X^{-1}\left(\prod_{i=1}^n\right)(-\infty, a_i]) \in \mathcal{E}$. Es gilt:

$$X^{-1}\left(\prod_{i \in \underline{n}} (-\infty, a_i]\right) = \bigcap_{i \in \underline{n}} \underbrace{X_1^{-1}((-\infty, a_i])}_{\in \mathcal{E}} \in \mathcal{E},$$

da $\mathcal{E}$ σ-Algebra.

iii) Es gilt:

$$x \overset{w_1}{\mapsto} (x, x) \overset{w_2}{\mapsto} (X(x), Y(x)) \overset{w_3}{\mapsto} X(x) + Y(x)$$

Offensichtlich sind w_1 und w_3 stetig sowie w_2 messbar nach Voraussetzung. Somit ist $X+Y$ messbar als Komposition messbarer Abbildungen. Der Rest folgt analog. □

Satz und Definition 8.12 Sei $(E, \mathcal{E}, \mu)$ Maßraum, $(F, \mathcal{F})$ Messraum und $X : E \to F$ messbar. Dann ist $\mu^X : F \to [0, 1], B \mapsto \mu(X^{-1}(B))$ ein Maß auf F, das sogenannte *Bildmaß*.

Ist $\mu = P$ ein W'maß, so heißt P^X die *Verteilung von X*.

Beweis:

i) $\mu^X(\emptyset) = \mu(X^{-1}(\emptyset)) = \mu(\emptyset) = 0$

ii) Seien $B_1, B_2, \ldots \in \mathcal{F}$ disjunkt. Dann gilt:

$$\mu^X\left(\biguplus_{i \geq 1} B_i\right) = \mu\left(X^{-1}\left(\biguplus_i B_i\right)\right) = \mu\left(\biguplus_i X^{-1}(B_i)\right) = \sum_i \mu\left(X^{-1}(B_i)\right) = \sum_i \mu^X(B_i).$$

□

Bemerkung 8.13 Hat man $X : (\Omega, \mathcal{A}, \mu) \to (F, \mathcal{F}, P^X)$, so arbeitet man meist nur mit dem Bildraum und "vergisst" $(\Omega, \mathcal{A}, P)$.

9 Integration bezüglich Maßen

Bemerkung 9.1 Im diskrete Fall gilt für $X : \Omega \to \mathbb{R}$, dass $\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)P(\omega)$ der Erwartungswert ist. Diese Formel lässt sich nicht direkt auf überabzählbares Ω übertragen. Also ist ein neues Konzept notwendig.

In diesem Kapitel sei $(\Omega, \mathcal{A}, \mu)$ Maßraum.

Definition 9.2

i) Eine Abbildung $X : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B})$ mit $X \geq 0$ heißt *Elementarfunktion* oder *einfach*, falls
   (a) $|Bild\, X| < \infty$, dh. $X = \sum_{i\in \underline{n}} a_i 1_{A_i}$ mit $a_i \geq 1$ und $A_1, \ldots, A_n \in \mathcal{A}$ disjunkt mit $\biguplus_i A_i = \Omega$
   (b) X messbar.
ii) Für X wie oben setzen wir

$$\int X \, d\mu := \sum_{i\in \underline{n}} a_i \mu(A_i). \tag{9}$$

Bemerkung 9.3 Problem bei (9): Wohldefiniert? (Abhängigkeit von der Darstellung von X?)

Nach Voraussetzung gilt $X(\Omega) = \{c_1, \ldots, c_l\}$ mit $c_1 < \ldots < c_l$. Dann ist

$$X = \sum_{k=1}^{l} c_k 1_{\{X=c_k\}}$$

die *Normalform von X*. Man zeigt damit (leicht): alle andere Darstellungen haben den gleichen Wert.

Exkurs zu Integralen:

Riemann-Integral Approximation über Funktionen, die nur endlich oft springen und sonst konstant sind (Treppenfunktionen).

Lebesgue-Integral Approximation über Funktionen $X = \sum_{i=1}^{n} a_i 1_{A_i}$ mit $\biguplus_i A_i = \Omega$, die nur endlich viele Werte annehmen. Diese Funktionen können unendlich oft springen.

Satz 9.4 Seien X und Y einfache Funktionen.

i) Für $\alpha, \beta \in \mathbb{R}$ gilt: $\int (\alpha X + \beta Y)\, d\mu = \alpha \int X \, d\mu + \beta \int Y \, d\mu$. (Linerarität)
ii) $X \leq Y \;\Rightarrow\; \int X \, d\mu \leq \int Y \, d\mu$ (Monotonie)

Beweis:

i) Klar ist $\int \alpha X \, d\mu = \alpha \int X \, d\mu$ (nach Definition). Zu zeigen bleibt: $\int (X + Y)\, d\mu = \int X \, d\mu + \int Y \, d\mu$. Seien oBdA X, Y in Normalform, also $X = \sum_{i=1}^{n} a_i 1_{A_i}$ und $Y = \sum_{j=1}^{m} b_j 1_{A_j}$. Dann gilt:

$$X = \sum_{i=1}^{n} a_i \sum_{j=1}^{m} 1_{(A_i\cap B_j)} \qquad \text{und} \qquad Y = \sum_{j=1}^{m} b_i \sum_{i=1}^{n} 1_{(A_i\cap B_j)},$$

da $\Omega = \biguplus_i A_i = \biguplus_j B_j$. Mit $X = \sum_{i=1}^n \sum_{j=1}^m c_{ij} 1_{(A_i \cap B_j)}$ und $Y = \sum_{j=1}^m \sum_{i=1}^n d_{ij} 1_{(A_i \cap B_j)}$ besitzen X und Y dieselbe Darstellung bezüglich Indikatoren und die Behauptung folgt nach Definition.

ii) Angenommen, es gilt $X = \sum_{i=1}^n a_i 1_{A_i}$ und $Y = \sum_{i=1}^n b_i 1_{A_i}$. Dann folgt, dass $Y - X = \sum_{i=1}^n (b_i - a_i) 1_{A_i}$ eine einfache Funktion ist mit $Y - X \geq 0$. Also folgt nach Definition

$$\int (Y - X)\, d\mu \geq 0 \;\Leftrightarrow\; \int Y\, d\mu \geq \int X\, d\mu.$$

□

Definition 9.5 Sei $X : (\Omega, \mathcal{A}, \mu) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ messbar und nicht-negativ. Dann heißt

$$\int X(\omega)\, d\mu(\omega) = \sup \left\{ \int Y\, d\mu : Y \text{ einfach mit } Y \leq X \right\}$$

das *Integral von X bezüglich μ*. Notationsvarianten (gelten auch für 9.2):

- $\int X\, d\mu$
- $\int X(\omega)\, \mu(d\omega)$.

Bemerkung 9.6

i) Da $Y \equiv 0$ einfach mit $Y \leq X$ ist, folgt $\int X\, d\mu \geq 0$.

ii) Sei X einfach. Dann gilt $\int Y\, d\mu \leq \int X\, d\mu$ für alle einfachen Y mit $Y \leq X$ nach Satz 9.4. Folglich stimmen die Definitionen von $\int X\, d\mu$ überein.

Definition 9.7 Es sei $X : (\Omega, \mathcal{A}, \mu) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ messbar. Dann heißen $X^+ := X \vee 0$ bzw. $X^- := (-X) \vee 0$ *Positivteil* bzw. *Negativteil* von X.

Bemerkung 9.8 Offenbar: $X^+, X^- \geq 0$ und $X^+ - X^-$ sowie $|X| = X^+ + X^-$ messbar.

Definition 9.9 Es sei $X : (\Omega, \mathcal{A}, \mu) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ messbar mit $\int X^+\, d\mu \leq \infty$ oder $\int X^-\, d\mu \leq \infty$.

i) Dann heißt

$$\int X\, d\mu := \int X^+\, d\mu - \int X^-\, d\mu$$

das *Integral von X bezüglich μ*. (Werte $+\infty$ und $-\infty$ sind möglich!)

ii) X heißt *integrierbar*, falls $\int X^+\, d\mu, \int X^-\, d\mu < \infty$.

Der Raum der aller integrierbaren Funktionen heißt

$$\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{A}, \mu) = \{X : (\Omega, \mathcal{A}, \mu) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}}) : X \text{ integrierbar}\}.$$

Satz 9.10 Sei $X : (\Omega, \mathcal{A}, \mu) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ messbar und nicht-negativ. Dann existiert eine Folge $(X_n)_n$ nicht-negativer, einfacher Funktionen mit $X_n \nearrow X$, dh. $X_n(\omega) \to X(\omega)$ für alle ω, wobei $X_n(\omega)$ monoton wachsend.

Beweis: Wähle

$$X_n(\omega) = \begin{cases} k2^{-n} & \text{falls } k2^{-n} \leq X(\omega) < (k+1)2^{-n} \\ n & \text{falls } X(\omega) \geq n \end{cases}.$$

für $k \in \{0, 1, \ldots, n2^n - 1\}$. Sei n fest. Wegen $k2^{-n} = (2k)2^{-(n+1)}$ ist $X_n(\omega) \leq X_{n+1}(\omega)$. Außerdem gilt $\sup_n X_n(\omega) = X(\omega)$. □

Bem: Die Umkehrung folgt direkt aus 8.7 ii).

Satz 9.11 Sei $X : (\Omega, \mathcal{A}, \mu) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ messbar und nicht-negativ. Ist $(X_n)_n$ Folge einfacher Funktionen mit $X_n \nearrow X$, so folgt:

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

Beweis: Die Folge $\left(\int X_n \, d\mu\right)_n$ ist monoton wachsend, da $(X_n)_x$ monoton wachsend ist (9.4 ii)), und konvergiert gegen $a \in \overline{\mathbb{R}}$. Nach Definition 9.5 gilt $\int X \, d\mu \geq a$.

Um die andere Ungleichung zu zeigen, genügt ein Beweis von

$$0 \leq Y \leq X \text{ mit } Y \text{ einfach } \Rightarrow \int Y \, d\mu \leq a.$$

Dann folgt $\int X \, d\mu \leq a \;\Rightarrow\; \int X \, d\mu = a = \lim_{n\to\infty} \left(\int X_n \, d\mu\right)_n$.

Sei also $Y = \sum_{i=1}^m a_i 1_{A_i}$ in Normalform, also $A_i = Y^{-1}(\{a_i\})$. Wähle zu $\varepsilon > 0$:

$$Y_{n,\varepsilon} = (1-\varepsilon) Y 1_{\{(1-\varepsilon)Y \leq X_n\}} = \sum_{i \in \underline{m}} (1-\varepsilon) a_i 1_{\{(1-\varepsilon)Y \leq X_n\} \cap A_i}.$$

Offenbar: $Y_{n,\varepsilon} \leq X$, also

$$\int Y_{n,\varepsilon} \, d\mu = (1-\varepsilon) \sum_{i \in \underline{m}} a_i \mu\left(\{(1-\varepsilon)Y \leq X_n\} \cap A_i\right) \overset{\text{Mon.}}{\leq} \int X_n \, d\mu.$$

Gilt nun für jedes $\varepsilon > 0$:

$$\{\omega : (1-\varepsilon)Y(\omega) \leq X_n(\omega)\} \to \Omega, \tag{10}$$

so folgt

$$\begin{aligned}(1-\varepsilon) \int Y \, d\mu &= (1-\varepsilon) \sum_{i \in \underline{m}} a_i \mu(A_i) = \lim_{n\to\infty} (1-\varepsilon) \sum_{i \in \underline{m}} a_i \underbrace{\mu\big(\overbrace{\{(1-\varepsilon)Y \leq X_n\} \cap A_i}^{\to A_i}\big)}_{\lim \mu(\{(1-\varepsilon)Y \leq X_n\} \cap A_i) = \mu(A_i)} \\ &\leq \lim_{n\to\infty} \int X_n \, d\mu = a\end{aligned}$$

und somit $\int Y \, d\mu \leq a$.

Um den Beweis abzuschließen, muss nur noch (10) gezeigt werden:
Fall 1: Ist $Y(\omega) = 0$, so gilt (10) wegen $X_n(\omega) \geq 0$.
Fall 2: Ist $Y(\omega) > 0$, so folgt, dass $(1-\varepsilon)Y(\omega) \leq (1-\varepsilon)X(\omega)$ und $(X_n)_n$ ist monoton wachsend mit $X_n(\omega) \to X(\omega)$. Es folgt $(1-\varepsilon)Y(\omega) < X_n(\omega)$ für n groß genug. □

Bemerkung 9.12 Will man zeigen, dass eine Eigenschaft (*) für alle messbaren Funktionen $X : (\Omega, \mathcal{A}, \mu) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ gilt, ist nach 9.10 die folgende "maßtheoretische Induktion" nützlich. Zeige:

i) (*) gilt für Indikatorfunktionen.

ii) (*) bleibt auch unter Linearkombinationen (von Indikatorfunktionen) gültig ($\to$ einfache Funktionen)

iii) Gilt (*) für eine Folge $(X_n)_n$ mit $X_n \geq 0$ und $X_n \nearrow X$, so gilt (*) auch für X.

iv) Gilt (*) für X^+ und X^-, so auch für $X = X^+ - X^-$.

Definition 9.13 Sei $A \in \mathcal{A}$. Wir notieren $\int_A X \, d\mu = \int X 1_A \, d\mu$, falls das letztere Integral existiert.

Satz 9.14 Es seien $X, Y : (\Omega, \mathcal{A}, \mu) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ messbar. Dann gilt:

i) Für $0 \leq X \leq Y$ gilt:

$$\int X \, d\mu \leq \int Y \, d\mu.$$

Außerdem $Y \in \mathcal{L}^1 \;\Rightarrow\; X \in \mathcal{L}^1$.

ii) $X \geq 0 \;\Rightarrow\; \int X \, d\mu \geq 0$.

iii) $\mathcal{L}^1$ ist ein Vektorraum und dort ist das Integral linear.

iv) $X \in \mathcal{L}^1 \;\Leftrightarrow\; |X| \in \mathcal{L}^1$. In diesem Fall gilt: $\left|\int X \, d\mu\right| \leq \int |X| \, d\mu$.

v) Ist X beschränkt und μ endlich, so folgt $X \in \mathcal{L}^1$.

vi) Seien $X, Y \geq 0$ oder $X, Y \in \mathcal{L}^1$. Dann:

$$X = Y \ \mu\text{-fast überall (dh. } \mu(\{\omega : X(\omega) \neq Y(\omega)\}) = 0) \;\Leftrightarrow\; \int_A X \, d\mu = \int_A Y \, d\mu \text{ für alle } A \in \mathcal{A}.$$

Beweis:

i) Wähle $X_n, Y_n \geq 0$ einfach mit $X_n \nearrow X, Y_n \nearrow Y$ und $X_n \leq Y_n$ nach 9.10. Es folgt mit 9.4 und 9.11:

$$\int X_n \, d\mu \leq \int Y_n \, d\mu \;\Rightarrow\; \int X \, d\mu \leq \int Y \, d\mu.$$

Weiter gilt:

$$Y \in \mathcal{L}^1 \;\Leftrightarrow\; \int Y \, d\mu < \infty \;\Rightarrow\; \int X \, d\mu < \infty \;\Leftrightarrow\; X \in \mathcal{L}^1.$$

ii) folgt direkt aus i)

iii) Übung!

iv) Aus $X \in \mathcal{L}^1 < \infty$ folgt $\int X^+ \, d\mu < \infty$ und $\int X^- \, d\mu < \infty$. Mit iii) folgt

$$\int |X| \, d\mu = \int (X^+ - X^-) \, d\mu < \infty \;\Leftrightarrow\; |X| \in \mathcal{L}^1.$$

Umgekehrt sei $|X| \in \mathcal{L}^1$, so folgt:

$$\begin{aligned}\int |X| \, d\mu < \infty \;\Leftrightarrow\; & \int X^+ \, d\mu + \int X^- \, d\mu = \int (X^+ + X^-) \, d\mu < \infty \\ \Rightarrow\; & \int X \, d\mu = \int (X^+ - X^-) \, d\mu < \infty \;\Leftrightarrow\; X \in \mathcal{L}^1.\end{aligned}$$

Die Ungleichung folgt aus ii) via $|X| - X \geq 0$ und $|X| - (-X) \geq 0$.

v) Sei X beschränkt, also $|X| \leq c$, so folgt

$$\int |X| \, d\mu \leq \int c \, d\mu = \int c 1_\Omega \, d\mu = c\mu(\Omega) < \infty.$$

vi) "$\Rightarrow$": Seien $X, Y \geq 0$ und $B = \{\omega : X(\omega) \neq Y(\omega)\}$. Dann gilt:

$$\int_A Y \, d\mu = \int_{A \cap B} Y \, d\mu + \int_{A \cap B^c} Y \, d\mu = \int_{A \cap B} Y \, d\mu + \int_{A \cap B^c} X \, d\mu.$$

Wähle $0 \geq Y_n \nearrow Y$ einfach. Dann: $Y_n \leq c_n$ beschränkt. Es folgt:

$$0 \overset{\text{ii)}}{\leq} \int_{A \cap B} Y_n \, d\mu \leq c_n \mu(A \cap B) = 0 \overset{[9.11]}{\Rightarrow} \lim_{n \to \infty} \int_{A \cap B} Y_n \, d\mu = 0.$$

Insgesamt folgt:

$$\int_A Y \, d\mu = \int_{A \cap B^c} X \, d\mu = \int_A X \, d\mu - \int_{A \cap B} X \, d\mu \overset{\text{[analog]}}{=} \int_A X \, d\mu.$$

Allgemein: $X = Y$ μ-f.ü. $\Rightarrow X^+ = Y^+$ und $X^- = Y^-$ μ-f.ü.

"$\Leftarrow$": Zeige allgemein:

$$Z \geq 0 \wedge \int Z \, d\mu = 0 \;\Rightarrow\; \mu(A) = 0 \text{ mit } A = \{\omega : Z(\omega) = 0\}.$$

Wähle dazu $A_n = \{\omega : Z(\omega) \geq \frac{1}{n}\}$ mit $A_n \nearrow A$. Dann gilt:

$$\frac{1}{n} 1_{A_n} \leq Z \text{ und } 0 \geq \frac{1}{n}\mu(A_n) = \int \frac{1}{n} 1_{A_n} \, d\mu \leq \int Z \, d\mu = 0.$$

Es folgt $\mu(A_n) = 0$ für alle $n \in \mathbb{N}$ und somit $\mu(A) = 0$.

Sei $B = \{\omega : X(\omega) > Y(\omega)\}$. Dann folgt $(X-Y)1_B \geq 0$ und $\int (X-Y)1_B \, d\mu = 0$ nach Vorausetzung. Es folgt mit Hilfslemma $\mu(B) = 0$. Analog wird begründet, dass $\mu(\{\omega : X(\omega) < Y(\omega)\}) = 0$. $\square$

Satz 9.15 (Satz der monotonen Konvergenz) Sei $(X_n)_n$ Folge messbarer, nicht-negativer Abbildungen mit $X_n \nearrow X$ (X_n müssen nicht einfach sein!) Dann gilt:

$$\int X_n \, d\mu \nearrow \int X \, d\mu.$$

Beweis: Für jedes n wähle $Y_{n,k}$ einfach mit $0 \geq Y_{n,k} \nearrow X_n$ (Existenz nach 9.10). Setze $Z_k := \max_{n \leq k} Y_{n,k}$, also ist Z_n einfach. Für alle $n \leq k$ gilt:

$$Y_{n,k} \leq Z_k \leq X_n \leq X_k \leq X. \tag{11}$$

Beachte $Y_{n,k} \nearrow X_n, X_k \nearrow X$ und $(Z_k)_k$ monoton wachsend. Also konvergiert Z_k gegen Z und $X_n \leq Z \leq X$ wegen (11) und somit folgt $X = Z$.

Außerdem:

$$\int Y_{n,k} \, d\mu \leq \int Z_k \, d\mu \leq \int X_n \, d\mu \qquad \text{für alle } n \leq k. \tag{12}$$

Es folgt:

$$\begin{aligned} &\int X_n \, d\mu \overset{[9.11]}{=} \lim_{k \to \infty} \int Y_{n,k} \, d\mu \overset{(12)}{\leq} \lim_{k \to \infty} \int Z_k \, d\mu \overset{[9.11]}{=} \int Z \, d\mu \overset{(12)}{\leq} \lim_{n \to \infty} \int X_k \, d\mu \\ \Rightarrow &\lim_{n \to \infty} \int X_n \, d\mu \leq \int Z \, d\mu \leq \lim_{k \to \infty} \int X_k \, d\mu. \end{aligned}$$

Insgesamt folgt die Aussage. □

Korollar 9.16 Es sei $(X_n)_n$ messbare Folge mit $X_n \geq 0$. Dann gilt:

$$\int \sum_{n=1}^{\infty} X_n \, d\mu = \sum_{n=1}^{\infty} \int X_n \, d\mu.$$

Beweis: Satz 9.15 □

Lemma 9.17 (Lemma von Fatou) Für jede messbare Folge $(X_n)_n$ mit $X_n \geq 0$ gilt:

$$\int \liminf_n X_n \, d\mu \leq \liminf_n \int X_n \, d\mu.$$

Beweis: Sei $Y_n := \inf_{k \geq n} X_k$. Klar: $Y_n \geq 0, (Y_n)_n$ monoton wachsend, $Y_n \leq X_n$ und $Y_n \to \liminf_{k \to \infty} X_k$. Also folgt mit 9.15:

$$\liminf_{n \to \infty} \int X_n \, d\mu \geq \liminf_{n \to \infty} \int Y_n \, d\mu = \int \lim_{n \to \infty} Y_n \, d\mu = \int \liminf_{n \to \infty} Y_n \, d\mu.$$

□

Satz 9.18 (Satz von der majorisierten Konvergenz) Es seien $(X_n)_n, X$ und Y messbar mit $X_n \to X, |X_n| \leq Y$ und $Y \in \mathcal{L}^1$ gegeben. Dann gilt:

$$X_n, X \in \mathcal{L}^1 \qquad \text{und} \qquad \int X_n \, d\mu \to \int X \, d\mu.$$

Beweis: $|X_n| \le Y$ und $X_n \to X$ liefert $|X| \le Y$. Mit 9.14 folgt $X, X_n \in \mathcal{L}^1$.
Ansonsten gilt $X_n + Y \ge 0$, also folgt:

$$\begin{aligned}\liminf_{n\to\infty} \int X_n \, d\mu + \int Y \, d\mu &= \liminf_{n\to\infty} \int (X_n + Y) \, d\mu \\ &\overset{\text{Fatou}}{\geq} \int \liminf_{n\to\infty} (X_n + Y) \, d\mu = \int \liminf_{n\to\infty} X_n \, d\mu + \int Y \, d\mu.\end{aligned}$$

Wegen $Y \in \mathcal{L}^1$ und $X_n \to X$ folgt:

$$\liminf_{n\to\infty} \int X_n \, d\mu \ge \int \liminf_{n\to\infty} X_n \, d\mu = \int X \, d\mu.$$

Genauso liefert Fatou angewandt auf $-X_n + Y \ge 0$:

$$\begin{aligned}\liminf_{n\to\infty} \int (-X_n) \, d\mu \ge \int (-X) \, d\mu \; &\Leftrightarrow \; -\limsup_{n\to\infty} \int X_n \, d\mu \ge - \int X \, d\mu \\ &\Leftrightarrow \; \limsup_{n\to\infty} \int X_n \, d\mu \le \int X \, d\mu.\end{aligned}$$

Insgesamt folgt:

$$\liminf_{n\to\infty} \int X_n \, d\mu \ge \int X \, d\mu \ge \limsup_{n\to\infty} \int X_n \, d\mu \; \Leftrightarrow \; \lim_{n\to\infty} \int X_n \, d\mu = \int X \, d\mu.$$

□

Definition 9.19 Es seien $(\Omega, \mathcal{A}, P)$ W'raum und $X : (\Omega, \mathcal{A}, P) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ eine Zufallsgröße. Im Falle der Existenz heißt $\mathbb{E}[X] = \int X \, dP$ *Erwartungswert von* X.

Definition 9.20

i) Es sei Ω beliebig und $\tilde{\Omega} \subseteq \Omega$ mit $|\tilde{\Omega}| \le \aleph_0$. Zu $p : \tilde{\Omega} \to [0, \infty]$ heißt

$$\mu(A) = \sum_{\omega \in A \cap \tilde{\Omega}} p(\omega) \tag{13}$$

ein *diskretes Maß*.
Gilt $\sum_{\omega\in\tilde{\Omega}} p(\omega) = 1$, so ist μ ein W'maß im Sinne von 1.3.
(σ-Algebra nicht wichtig, da μ auf jeder σ-Algebra ein Maß ist.)

ii) Ist $\tilde{\Omega} = \{\omega_0\}$ und $p(\omega_0) = 1$, so heißt μ das *Dirac-Maß in* ω_0 und es gilt

$$\mu(A) = \begin{cases} 1 & , \omega_0 \in A \\ 0 & , \omega_0 \notin A \end{cases}.$$

Notation: δ_{ω_0}

iii) Gilt (13) mit $p(\omega) = 1$ für alle $\omega \in \tilde{\Omega}$, so heißt μ das *Zählmaß auf* $\tilde{\Omega}$.

Bemerkung 9.21 Ist P diskretes W'maß gemäß 9.20 a), so stimmt $\mathbb{E}[X]$ mit der ursprünglichen Definition 3.7 überein:

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)p(\omega).$$

Beweis: Sei oE. $X \geq 0$. Sonst $X = X^+ - X^-$ und Linearität nutzen. Ist $X = \sum_{i=1}^{\infty} c_i 1_{A_i}$, so gilt:

$$\int X\,dP = \sum_{i=1}^{m} c_i P(A_i) \overset{9.20\text{ a)}}{=} \sum_{i=1}^{m} c_i \sum_{\omega\in\tilde{\Omega}} 1_{A_i}(\omega)p(\omega) = \sum_{\omega\in\tilde{\Omega}} \underbrace{\sum_{i=1}^{\infty} c_i 1_{A_i}(\omega)}_{X(\omega)} p(\omega) = \sum_{\omega\in\tilde{\Omega}} X(\omega)p(\omega).$$

(Vertauschen erlaubt, da alles nicht negativ)

Gilt $X_n \nearrow X$ für einfache Funktionen X_n, so folgt:

$$\int X\,dP \overset{9.11}{=} \lim_{n\to\infty} \int X_n\,dP = \lim_{n\to\infty} \sum_{\omega\in\Omega} X_n(\omega)p(\omega) \overset{(*)}{=} \sum_{\omega\in\Omega} \lim_{n\to\infty} X_n(\omega)p(\omega) = \sum_{\omega\in\Omega} X(\omega)p(\omega).$$

(Gleichheit bei (*) wegen monotoner Konvergenz für Reihen) □

Korollar 9.22 Gilt für eine Folge $(X_n)_n$ von Zufallsgrößen $\sum_{n=1}^{\infty} \mathbb{E}\left[|X_n|\right] < \infty$, so existiert $\sum_{i=1}^{\infty} X_n$ fast sicher, dh.

$$P\left(\left\{\omega : \left|\sum_{i=1}^{\infty} X_n(\omega)\right| < \infty\right\}\right) = 1,$$

und es gilt:

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[X_n].$$

Beweis: Setze $S_n = \sum_{k=1}^{n} |X_k|$. Klar: S_n ist monoton wachsend und somit $S_n \nearrow S = \sum_{k=1}^{\infty} |X_k| \geq 0$. Es folgt mit Satz der monotonen Konvergenz, daß $\mathbb{E}[S_n] \nearrow \mathbb{E}[S]$. Zudem gilt (Eindeutigkeit des GW.es):

$$\mathbb{E}[S_n] = \sum_{k=1}^{n} \mathbb{E}[|X_k|] \to \sum_{k=1}^{\infty} \mathbb{E}\left[|X_k|\right] < \infty \;\Rightarrow\; \mathbb{E}[S] = \sum_{i=1}^{\infty} \mathbb{E}[|X_k|] < \infty$$

Wegen $1_{\{S=\infty\}} \leq \varepsilon S$ für alle $\varepsilon > 0$, folgt:

$$P(S = \infty) = \mathbb{E}\big[1_{\{S=\infty\}}\big] \leq \varepsilon \mathbb{E}[S] \qquad \forall \varepsilon > 0$$

und somit $P(S = \infty) = 0$. Also konvergiert S f.s. absolut und damit auch

$$T_n = \sum_{k=1}^{n} X_k \to T = \sum_{k=1}^{\infty} X_k \text{ (f.s.).}$$

Zudem gilt $|T_n| \leq S_n \leq S$ und $S \in \mathcal{L}^1$, also folgt:

$$\mathbb{E}\left[\sum_{k=1}^{\infty} X_k\right] = \mathbb{E}\Big[\lim_{n\to\infty} T_n\Big] \overset{\text{maj.K.}}{=} \lim_{n\to\infty} \mathbb{E}[T_n] = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}[X_k] = \sum_{k=1}^{\infty} \mathbb{E}[X_k].$$

□

Bemerkung 9.23 Man betrachtet formal meist Äquivalenzklassen von Zufallsgrößen, dh. man setzt $X \sim Y \;:\Leftrightarrow\; X = Y$ fast sicher bzw. fast überall gleich. Das Verhalten auf Nullmengen spielt also für die Äquivalenz keine Rolle.

Definition 9.24 Für $p \in [0, \infty)$ definiere

$$\mathcal{L}^p := \{X : (\Omega, \mathcal{A}, \mu) \to \big(\overline{\mathbb{R}}, \overline{\mathcal{B}}\big) : X^p \in \mathcal{L}^1\}$$

sowie

$$L^p := \mathcal{L}^p / \sim .$$

Bemerkung 9.25

i) L^p ist mit $\|X\|_p = \left(\int |X|^p \, d\mu\right)^{\frac{1}{p}}$ ein normierter Raum. $\mathcal{L}^p$ ist wegen $\|X\|_p = 0 \;\not\Rightarrow\; X = 0$ kein normierter Raum.

   Bsp: Ist N Nullmenge, so ist $1_N \neq 0$, aber $\int 1_N \, dP = 0$.

ii) Analog zu Kapitel 3 lässt sich $Var(X)$ definieren. Es gelten dieselben Rechenregeln (ua. Verschiebungssatz, Markow'sche und Tschebyschoff'sche Ungleichung).

Satz 9.26 (Transformationssatz) Seien $X : (\Omega, \mathcal{A}, \mu) \to (E, \mathcal{E})$ und $h : (E, \mathcal{E}) \to \big(\overline{\mathbb{R}}, \overline{\mathcal{B}}\big)$ messbar. Dann gilt mit dem Bildmaß $\mu^X(B) = \mu\big(X^{-1}(B)\big) = \mu(X = B)$ auf $(E, \mathcal{E})$:

$$h(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu) \;\Leftrightarrow\; h \in \mathcal{L}^1(E, \mathcal{E}, \mu^X).$$

In diesem Fall oder falls $h \geq 0$ gilt:

$$\int_\Omega h(X) \, d\mu = \int_E h \, d\mu^X. \tag{14}$$

(Verallgemeinerung der Integration durch Substitution)

Beweis: Es genügt, (14) für $h \geq 0$ zu zeigen. Sei dazu $h = \sum_{i=1}^n a_i 1_{A_i}$. Dann gilt mit $A_i' = X^{-1}(A_i)$:

$$\int_\Omega h(X) \, d\mu = \sum_{i=1}^n a_i \int_\Omega 1_{A_i}(X(\omega)) \, d\mu(\omega) = \sum_{i=1}^n a_i \mu(A_i') = \sum_{i=1}^n a_i \mu^X(A_i) = \int_E h \, d\mu^X.$$

Gilt $h_n \nearrow h$ für einfache Funktionen, dann auch $h_n \circ X \nearrow h \circ X$. Die Aussage folgt dann aus 9.11. □

Korollar 9.27 Sei $X : (\Omega, \mathcal{A}, \mu) \to \big(\overline{\mathbb{R}}, \overline{\mathcal{B}}\big)$ in $\mathcal{L}^1$ oder $\mathcal{L}^2$. Dann gilt:

$$\mathbb{E}[X] = \int_\Omega X \, dP = \int_{\overline{\mathbb{R}}} X dP^X$$

bzw.

$$Var(X) = \int_\Omega (X - \mathbb{E}[X])^2 \, dP = \int_{\overline{\mathbb{R}}} (X - \mathbb{E}[X])^2 dP^X.$$

Beweis: 9.26 □

10 Produkträume und Satz von Fubini

Erinnerung: $A_1, \ldots, A_n$ heißen unabhängig, falls $P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k})$ für alle $1 \leq i_1 \leq \ldots \leq i_k \leq n$.

Definition 10.1 Es sei $(\Omega, \mathcal{A}, P)$ W'raum und I beliebige Indexmenge.

i) Eine Familie $(A_i)_{i \in I}$ mit $A_i \in \mathcal{A}$ heißt *unabhängig*, falls:
$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \qquad \text{für alle endlichen } J \subseteq I. \tag{15}$$

ii) Eine Familie von Teil-σ-Algebren $(\mathcal{A}_i)_{i \in I}$ mit $\mathcal{A}_i \subseteq \mathcal{A}$ heißt *unabhängig*, falls (15) für alle endlichen $J \subseteq I$ und alle $A_j \in \mathcal{A}_j$ gilt.

iii) Zufallsgrößen $(X_i)_{i \in I}$ mit Werten in $(E_i, \mathcal{E}_i)$ heißen *unabhängig*, falls die $(X_i^{-1}(\mathcal{E}_i))_{i \in I} =: \sigma(X_i)$ unabhängig sind.

Satz 10.2 Es seien X und Y Zufallsgrößen mit Werten in $(E, \mathcal{E})$ und $(F, \mathcal{F})$. Dann sind äquivalent:

i) X und Y sind unabhängig.

ii) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ für alle $A \in \mathcal{E}$ und $B \in \mathcal{F}$.

iii) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ für alle $A \in \mathcal{C}$ und $B \in \mathcal{D}$, wobei $\mathcal{C}, \mathcal{D}$ durchschnittstabil und $\mathcal{E} = \sigma(\mathcal{C}), \mathcal{F} = \sigma(\mathcal{D})$.

iv) $f(X)$ und $g(Y)$ sind unabhängig für alle messbaren f, g.

v) Für alle

- messbaren und beschränkten oder
- nicht-negativen oder
- stetigen und beschränkten (falls $\mathcal{E}, \mathcal{F}$ Borel-σ-Algebren)

$f : E \to \overline{\mathbb{R}}$ und $g : F \to \overline{\mathbb{R}}$ gilt:
$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \tag{16}$$

Beweis: i) $\Leftrightarrow$ ii): nach Definition.

ii) $\Rightarrow$ iii): klar

iii) $\Rightarrow$ ii): Sei $B \in \mathcal{D}$ fest. Man zeigt leicht: $\{A \in \mathcal{E} : P(X \in A, Y \in B) = P(X \in A)P(Y \in B)\} =: M$ ist ein Dynkin-System und enthält $\mathcal{C}$ nach Voraussetzung. Mit Dynkin-System-Argument folgt $M = \mathcal{E}$. Dann $A \in \mathcal{E}$ fixieren und Argumentation wiederholen.

iv) $\Rightarrow$ i): Wähle die Identität.

i) $\Rightarrow$ iv): Sei $f : (E, \mathcal{E}) \to (E', \mathcal{E}')$ messbar. Dann gilt:
$$(f \circ X)^{-1}(\mathcal{E}') = X^{-1}(f^{-1}(\mathcal{E}')) \subseteq X^{-1}(\mathcal{E})$$
und analog $(g \circ Y)^{-1}(\mathcal{F}) \subseteq Y^{-1}(\mathcal{F})$. Unabhängigkeit folgt nach Definition.

ii) $\Rightarrow$ v): Es genügt, denn Fall beschränkter oder nicht-negativer f, g zu betrachten. Klar: (16) gilt für Indikatorfunktionen (wegen ii)) und wegen Linearität auch für einfache Funktionen. Da $f, g \geq 0$ gilt:

$$\begin{aligned}\mathbb{E}\left[f(X)g(Y)\right] &= \mathbb{E}\Big[\lim_{n\to\infty} f_n(X)g(Y)\Big] \overset{\text{mon.K.}}{=} \lim_{n\to\infty} \mathbb{E}[f_n(X)g(Y)] \\ &= \lim_{n\to\infty} \mathbb{E}[f_n(X)]\mathbb{E}[g_n(Y)] \overset{\text{mon.K.}}{=} \mathbb{E}[f(X)]\mathbb{E}[g(Y)].\end{aligned}$$

Sind f, g bschränkt, so folgt Behauptung aus $f = f^+ - f^-$ und $g = g^+ - g^-$.

v) $\Rightarrow$ i): Übung. $\square$

Definition 10.4 Es seien $(E, \mathcal{E})$ und $(F, \mathcal{F})$ Messräume. Dann heißt

$$\mathcal{E} \otimes \mathcal{F} := \sigma(\Lambda \times \Gamma \mid \Lambda \in \mathcal{E}, \Gamma \in \mathcal{F}\}$$

Produkt-σ-Algebra von $\mathcal{E}$ und $\mathcal{F}$.

Bem: IA. ist $\mathcal{E} \times \mathcal{F}$ keine σ-Algebra. Beispiel: $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$, aber es existieren keine $X, Y \in \mathcal{A} \times \mathcal{A}$ mit $X \times Y = \{A, A\} \cup \{A^c, A^c\}$.

Satz 10.5 Es sei $f : (E \times F, \mathcal{E} \otimes \mathcal{F}) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ messbar. Dann ist für jedes $x \in E$ die Abbildung (auch "*Schnitt*" genannt) $y \mapsto f(x, y)$ ($\mathcal{F}$-$\overline{\mathcal{B}}$)-messbar (plus "vertauschte Rollen").

Beweis: Sei $f(x, y) = 1_C(x, y)$ mit $C \in \mathcal{E} \otimes \mathcal{F}$. Man zeigt leicht:

$$\mathcal{H} = \{C \in \mathcal{E} \otimes \mathcal{F} : y \mapsto 1_C(x, y) \text{ ist } (\mathcal{F}\text{-}\overline{\mathcal{B}})\text{-messbar für alle } x \in E\}$$

ist eine σ-Algebra ($1_{E\times F} = 1$, Differenzen messbarer Abbildungen sind messbar, Limiten sind messbar). Außerdem: $\Lambda \times \Gamma \in \mathcal{H}$ für alle $\Lambda \in \mathcal{E}$ und $\Gamma \in \mathcal{F}$ wegen $y \mapsto 1_\Gamma(x)1_\Lambda(y)$. Folglich $\mathcal{H} = \mathcal{E} \otimes \mathcal{F}$ und Indikatorfunktionen sind messbar. Insbesondere folgt Messbarkeit von einfachen Funktionen als Linearkombinationen von messbaren Funktionen. Sei $f \geq 0$ und $f_n \nearrow f$ mit f_n einfach. Dann ist $g_n(y) = f_n(x, y)$ messbar und $g(y) = \lim_{n\to\infty} g_n(y) = f(x, y)$ messbar (da GWe von messbaren Funktionen). Allgemeiner Fall: $f = f^+ - f^-$ und somit f messbar. $\square$

Satz 10.6 (Satz von Fubini) Seien P und Q W'maße auf Messräumen $(E, \mathcal{E})$ und $(F, \mathcal{F})$. Dann gilt:

i) Es existiert ein eindeutiges W'maß R auf $(E \times F, \mathcal{E} \otimes \mathcal{F})$ mit $R(A \times B) = P(A)Q(B)$ für alle $A \in \mathcal{E}$ und $B \in \mathcal{F}$. R heißt *Produktmaß*. Notation: $P \otimes Q$.

ii) Es sei $f : (E \times F, \mathcal{E} \otimes \mathcal{F}) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ messbar und

- $f \geq 0$ oder
- $f \in \mathcal{L}^1(E \times F, \mathcal{E} \otimes \mathcal{F}, P \otimes Q)$.

Dann gilt:

(a) $x \mapsto \int f(x, y)\, dQ(y)$ ist ($\mathcal{E}$-$\overline{\mathcal{B}}$)-messbar und

(b) $\int f(x, y)\, d(P \otimes Q)(x, y) = \iint f(x, y)\, dQ(y)\, dP(x) = \iint f(x, y)\, dP(x)\, dQ(y)$.

Beweis:

i) Da $\{\Lambda \times \Gamma \mid \Lambda \in \mathcal{E}, \Gamma \in \mathcal{F}\}$ durchschnittstabil ist (sowie P und Q σ-endlich sind), folgt die Eindeutigkeit nach 6.9.

   Sei nun $C \in \mathcal{E} \otimes \mathcal{F}$ und setze $C(x) = \{y : (x,y) \in \mathcal{C}\}$.

   Gilt $C = A \times B$, so folgt:

$$C(x) = \begin{cases} \emptyset & x \notin A \\ B & x \in A \end{cases}.$$

   Setze

$$R(C) := \int Q(C(x))\, dP(x) = \int_A Q(B)\, dP(x) = P(A)Q(B).$$

   Ausdehnen auf $\mathcal{E} \otimes \mathcal{F}$ und setze $\mathcal{H} = \{C \in \mathcal{E} \otimes \mathcal{F} : x \mapsto Q(C(x)) \text{ ist } \mathcal{E}\text{-messbar}\}$. Man zeigt leicht, dass $\mathcal{H}$ ein Dynkin-System ist und $\Lambda \times \Gamma \in \mathcal{H}$. Also $\mathcal{H} = \mathcal{E} \otimes \mathcal{F}$.

   Setze dann allgemein:

$$R(C) := \int Q(C(x))\, dP(x).$$

   Noch zu zeigen: R ist W'maß. Klar: $R(\emptyset) = 0$ und $R(E \times F) = P(E)Q(F) = 1$. Sei $(C_n)_n$ paarweise disjunkte Folge aus $\mathcal{E} \otimes \mathcal{F}$ und $C = \bigcup_n C_n$. Dann sind $(C_n(x))_n$ paarweise disjunkt und $Q(C(x)) = \sum_{n=1}^{\infty} Q(C_n(x))$. Zuletzt:

$$\sum_{n=1}^{\infty} R(C_n) = \sum_{n=1}^{\infty} \int Q(C_n(x))\, dP(x) \overset{9.16}{=} \int \sum_{i=1}^{\infty} Q(C_n(x))\, dP(x) = \int Q(C(x))\, dP(x) = R(C).$$

ii) Für $f(x,y) = 1_C(x,y)$ gilt die Aussage bereits nach Teil i).

   Linearität: i) gilt für einfache Funktionen. Sei also $f \geq 0$ und $f_n \nearrow f$ eine Folge einfacher Funktionen. Dann gilt:

$$\begin{aligned} \int f(x,y)\, dR(x,y) &\overset{\text{mon.K.}}{=} \lim_{n\to\infty} \int f_n(x,y)\, dR(x,y) = \lim_{n\to\infty} \int \left(\int f_n(x,y)\, dQ(y) \right) dP(x) \\ &\overset{\text{mon.K.}}{=} \int \left(\lim_{n\to\infty} \int f_n(x,y)\, dQ(y) \right) dP(x) \overset{\text{mon.K.}}{=} \int \int \lim_{n\to\infty} f_n(x,y)\, dQ(y)\, dP(x) \\ &= \iint f(x,y)\, dQ(y)\, dP(x). \end{aligned}$$

   Mit $f = f^+ - f^-$ folgt die Behauptung. $\square$

Korollar 10.7 Es seien X und Y Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(E, \mathcal{E})$ und $(F, \mathcal{F})$. Dann ist $Z = (X, Y)$ eine Zufallsgröße mit Werten in $(E \times F, \mathcal{E} \otimes \mathcal{F})$ und X, Y sind genau dann unabhängig, wenn gilt: $P^{(X,Y)} = P^X \otimes P^Y$.

Beweis: Messbarkeit von Z: Nach Satz 8.4 genügt es zu zeigen:

$$Z^{-1}(A \times B) = \big(X^{-1}(A) \cap Y^{-1}(B)\big) \in \mathcal{A} \qquad \text{für alle } A \in \mathcal{E}, B \in \mathcal{F},$$

da X, Y messbar.
Außerdem:

$$X, Y \text{ unabhängig} \Leftrightarrow P((X,Y) \in A \times B) = P(X \in A)P(Y \in B) \Leftrightarrow P^{(X,Y)}(A,B) = P^X(A)P^Y(B).$$

Die Eindeutigkeit es Produktmaßes liefert die Aussage. □

Bemerkung 10.8 Wie konstruiert man (unabhängige) Zufallsgrößen?

- Dimension 1: Zu $(E, \mathcal{E})$ und gegebenen W'maß μ wähle $\Omega = E, \mathcal{A} = \mathcal{E}, P = \mu$ und $X(\omega) = \omega$.
- Dimension 2: Zu $(E, \mathcal{E})$ und $(F, \mathcal{F})$ mit Maßen μ und ν wähle $\Omega = E \times F, \mathcal{A} = \mathcal{E} \otimes \mathcal{F}$ und $P = \mu \otimes \nu$ sowie $X(x,y) = x$ und $Y(x,y) = y$. (10.7)
- Dimension n: analog mit einer n-dimensionalen Version von Fubini.

Definition 10.9 Es seien $(\Omega_n, \mathcal{A}_n, P_n)$ für jedes $n \in \mathbb{N}$ W'räume. Dann bezeichnet $\Omega = \prod_{n=1}^{\infty} \Omega_n$ das abzählbare karthesische Produkt der $(\Omega_n)_n$ und

$$\mathcal{A} = \bigotimes_{n=1}^{\infty} \mathcal{A}_n = \sigma(\{A_1 \times \ldots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \ldots : A_i \in \mathcal{A}_i, k \geq 1\})$$

die Produkt-σ-Algebra von $(\mathcal{A}_n)_n$.

Satz 10.10 Auf $(\Omega, \mathcal{A})$ wie in 10.9 existiert ein eindeutiges W'maß $P = \bigotimes_{n=1}^{\infty} P_n$, sd. gilt:

$$P(A_1 \times \ldots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \ldots) = P(A_1) \cdot \ldots \cdot P(A_k) \qquad \text{für alle } A_i \in \mathcal{A}_i, k \geq 1.$$

Beweis: Maßtheorie. □

Bemerkung 10.11 Es seien $(X_n)_n$ Zufallsgrößen auf $(\Omega_n, \mathcal{A}_n, P_n)$. Dann gilt für die Zufallsgrößen $\tilde{X}_n$ auf dem Produktraum $(\Omega, \mathcal{A}, P)$ mit $\tilde{X}_n(\omega) = X(\omega_n)$, wobei $\omega = (\omega_1, \omega_2, \ldots)$:

$$\tilde{X}_n \text{ sind unabhängig} \qquad \text{und} \qquad P^{\tilde{X}_n} = P_n^{X_n}.$$

Beweis: Es gilt:

$$\tilde{X}_n^{-1}(B_n) = \Omega_1 \times \ldots \times \Omega_{n-1} \times X_n^{-1}(B_n) \times \Omega_{n+1} \times \ldots,$$

also $P(\tilde{X}_n^{-1}(B_n)) = P_n(X_n^{-1}(B_n))$ nach Definition des Produktmaßes.
Außerdem:

$$\begin{aligned} P(\tilde{X}_1 \in B_1, \ldots, \tilde{X}_k \in B_k) &= P(\bigcap_{n=1}^{k} \tilde{X}_n^{-1}(B_n)) = P(\tilde{X}_1^{-1}(B_1) \cap \ldots \cap \tilde{X}_k^{-1}(B_k) \cap \Omega_{k+1} \cap \ldots) \\ &= P_1(X_1^{-1}(B_1)) \cdot \ldots \cdot P_k(X_k^{-1}(B_k)) = P(\tilde{X}_1^{-1}(B_1)) \cdot \ldots \cdot P(\tilde{X}_k^{-1}(B_k)). \end{aligned}$$

□

Bem: Die Hauptaussage von 10.11 ist, dass es möglich ist, einen W'raum zu konstruieren, auf dem abzählbar viele Zufallsgrößen unabhängig sind. Damit ist 10.11 eine direkte Verallgemeinerung von 10.8.

Lemma 10.12 (Lemma von Borel-Cantelli) Es sei $(A_n)_n$ Folge von Ereignisse in $(\Omega, \mathcal{A}, P)$.

i) Ist $\sum_{i=1}^{\infty} P(A_n) < \infty$, dann gilt:
$$P(A_n \text{ unendlich oft}) = P(\limsup_{n\to\infty} A_n) = 0.$$

ii) Sind die $(A_n)_n$ unabhängig und gilt $\sum_{n=1}^{\infty} P(A_n) = \infty$, so folgt:
$$P(A_n \text{ unendlich oft}) = P(\limsup_{n\to\infty} A_n) = 1.$$

Beweis:

i) Es sei
$$\limsup_n A_n = \bigcap_{k\geq 1} \bigcup_{m\geq n} A_m =: \bigcap_{n\geq 1} B_n$$
und somit $B_n \supseteq B_{n+1}$. Es folgt $P\Big(\bigcap_{n\geq 1} B_n\Big) = \lim_{n\to\infty} P(B_n)$. Mit
$$P(B_n) \overset{\text{Subadd.}}{\leq} \sum_{m\geq n} P(A_m) \to 0$$
(nach Voraussetzung) folgt die Aussage.

ii) Es gilt $1 - x \leq e^{-x}$ (MWS der Diff'rechnung) für alle $x \in \mathbb{R}$, also folgt für $0 < n < m$:
$$P\left(\bigcap_{k=n}^{m} A_k^c\right) \overset{\text{Unabh.}}{=} \prod_{k=n}^{m} P(A_k^c) = \prod_{k=n}^{m} (1 - P(A_k)) \leq \prod_{k=n}^{m} exp(-P(A_k)) = exp\left(-\sum_{k=n}^{m} P(A_k)\right) \to 0$$
für $m \to \infty$ (exp ist die Exponentialfunktion). Also:
$$P\left(\bigcup_{k=n}^{\infty} A_k\right) = 1 - P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \overset{4.16}{=} 1 \qquad \text{für alle } n \in \mathbb{N}.$$
Wie im ersten Teil folgt $P(\limsup_n A_n) = 1$. □

Definition 10.13 Es seien $(X_n)_n$ Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(\mathbb{R}, \mathcal{B})$ und $\mathcal{C}_n = \sigma(X_k : k \geq n) = \sigma\big(\bigcup_{k=n}^{\infty} X_k^{-1}(\mathcal{B})\big)$. Dann heißt
$$\mathcal{C}_\infty = \bigcap_{n\geq 1} \mathcal{C}_n$$
die *σ-Algebra der terminalen Ereignisse*.

Idee: $\mathcal{C}_\infty$ enthält alle Informationen, die nicht von endlich vielen X_n abhängen.

Satz 10.14 (Null-Eins-Gesetz) Sind $(X_n)_n$ unabhängig, so gilt $P(C) \in \{0,1\}$ für alle $C \in \mathcal{C}_\infty$.

Beweis: Sei $\mathcal{D}_n = \sigma(X_k : k < n)$. Nach Voraussetzung sind $\mathcal{C}_n$ und $\mathcal{D}_n$ unabhängig, dh.

$$P(A \cap B) = P(A)P(B) \qquad \text{für alle } A \in \mathcal{C}_n, B \in \mathcal{D}_n. \tag{17}$$

Insbesondere: Ist $A \in \mathcal{C}_\infty$, so gilt (17) für alle $B \in \mathcal{D}_n$ (n beliebig), also auch für alle $B \in \bigcup_n \mathcal{D}_n$. Beachte: $\mathcal{D}_n$ ist durchschnittstabil und die Teilmengen B, die (17) für ein $A \in \mathcal{C}_\infty$ erfüllen, bilden ein Dynkin-System. Also:

$$P(A \cap B) = P(A)P(B) \qquad \text{für alle } A \in \mathcal{C}_\infty, B \in \sigma\left(\bigcup_n \mathcal{D}_n\right).$$

Zuletzt folgt $\mathcal{C}_\infty \subseteq \sigma(\bigcup_n \mathcal{D}_n)$ und somit $P(A) = P(A)^2$ für alle $A \in \mathcal{C}_\infty$, also $P(A) \in \{0,1\}$. $\square$

Bemerkung 10.15

i) Es gilt:

$$\{\omega : \lim_{n\to\infty} X_n(\omega) \text{ existiert}\} \in \mathcal{C}_\infty.$$

Für jede Folge unabhängiger Zufallsgrößen gilt also: Entweder sie konvergiert fü. oder sie divergiert f.s.

ii) Jede ($\mathcal{C}_\infty$-$\mathcal{B}$)-messbare Zufallsgröße ist f.s. konstant, insbesondere

$$\limsup_n X_n, \liminf_n X_n, \limsup_n \frac{1}{n}\sum_{i=1}^n X_i, \liminf_n \frac{1}{n}\sum_{i=1}^n X_i, \dots$$

11 Lebesgue-Integrale

Definition 11.1 Ein Maß λ auf $(\mathbb{R}, \mathcal{B})$ heißt *Lebesgue-Maß*, falls gilt:

$$\lambda((a,b]) = b - a.$$

Satz 11.2 Das Lebesgue-Maß existiert und ist eindeutig.

Beweis: Es sei $\mathcal{R}(\mathcal{J})$ der Ring der Figuren (s. Bsp. 4.11 iv)). Nach Satz 7.4 definiert $F(x) = x$ via

$$\mu(A) = \sum_{i \in \underline{n}} (F(b_i) - F(a_i)) = \sum_{i \in \underline{n}} (b_i - a_i) \qquad \text{für } A = \dot{\biguplus_{i \in \underline{n}}} (a_i, b_i]$$

ein Prämaß auf $\mathcal{R}(\mathcal{J})$ (7.3 und 7.4).

Die Existenz von λ folgt dann aus 5.8, da $\sigma(\mathcal{R}(\mathcal{J})) = \mathcal{B}$ (4.7). Die Eindeutigkeit folgt aus 6.9, da $\mathcal{R}(\mathcal{J})$ durchschnittstabil und λ σ-endlich über $\mathcal{R}(\mathcal{J})$. $\square$

Bemerkung 11.3 Man kann zeigen: Ist $f : [a,b] \to \mathbb{R}$ messbar und (eigentlich) R-integrierbar, so ist f auch bzgl. des L-Maßes integrierbar und beide Integrale stimmen überein.

Notation: $\int_a^b f(x)\,dx := \int_{[a.b]} f\,d\lambda$.

Definition 11.4 Eine messbare Abbildung $f \geq 0$ heißt *Dichte eines W'maßes P* (bzw. der Zufallsgröße X), falls für alle $x \in \mathbb{R}$ gilt:

$$P((-\infty, x]) = (\text{bzw. } P^X((-\infty,x]) =) \int_{-\infty}^x f(t)\,dt.$$

Satz 11.5 Eine messbare Abbildungen $f \geq 0$ ist genau dann Dichte eines W'maßes auf $(\mathbb{R}, \mathcal{B})$, wenn gilt:

$$\int_{-\infty}^{\infty} f(t)\,dt = 1.$$

Beweis:

"$\Rightarrow$": Es gilt $f 1_{(-\infty,n]} \nearrow f$ und $P((-\infty,n]) \to P((-\infty,\infty)) = P(\mathbb{R}) = 1$ nach Stetigkeit von unten. Also:

$$\int_{-\infty}^{\infty} f(t)\,dt = \lim_{n\to\infty} P((-\infty,n]) = 1.$$

"$\Leftarrow$": Definiere $F(x) = \int_{-\infty}^x f(t)\,dt$. Zu zeigen nach 7.8: F ist Verteilungsfunktion.

- F ist monoton:

$$x \leq y \;\Rightarrow\; f 1_{(-\infty,x]} \leq f 1_{(-\infty,y)} \;\Rightarrow\; \int f 1_{(-\infty,x]}\,d\lambda \leq \int f 1_{(-\infty,y]}\,d\lambda.$$

- F ist rechtsstetig: $x_n \searrow x \;\Rightarrow\; f 1_{(-\infty,x_n]} \searrow f 1_{(-\infty,x]}$. Außerdem ist $f 1_{(-\infty,x_n]} \leq f$ integrierbar, also folgt mit Monotonie (majorisierte Konvergenz):

$$F(x_n) = \int f 1_{(-\infty,x_n]}\,d\lambda \searrow \int f 1_{(-\infty,x]}\,d\lambda = F(x).$$

- $\lim_{n\to-\infty} F(x) = 0$ und $\lim_{n\to\infty} F(x) = 1$: analog. $\square$

Bemerkung 11.6

i) *9.14 (vi)* liefert: Dichten sind eindeutig bis auf λ-f.-ü. Gleichheit, d.h. f, f' definieren genau dann dasselbe W'maß, wenn $\lambda(f \neq f') = 0$.

ii) Vorteile von Lebesgue gegenüber Riemann:

- Liefert handhabbare Konvergenzsätze.
- Es ist möglich, $\int_A f(t)\, dt$ für viele Mengen A, statt nur für endliche Intervallvereinigungen, zu erklären.

iii) Ist f Dichte einer Verteilungsfunktion F und stetig, so gilt:

$$F'(x) = f(x) \qquad \text{für alle } x \in \mathbb{R}.$$

Warnung: Ist F eine stetige Verteilungsfkt., die an λ-f.-a. Stellen diff'bar ist, so ist F' keine Dichte f.

Satz 11.7 Sei $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ Zufallsgröße mit Dichte f und $g : \mathbb{R} \to \mathbb{R}$ messbar. Dann:

- $g \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, P^X) \;\Leftrightarrow\; fg \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, \lambda)$
- Falls $g \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, P^X)$ oder $g \geq 0$, so: $\mathbb{E}[g(X)] = \int g(x)f(x)dx$

Beweis: Mit maßtheoretischer Induktion:

- Sei $C \in \mathcal{B}, g = 1_C$. Dann gilt:

$$\mathbb{E}[g(X)] = \int g(X)dP \overset{9.26}{=} \int g dP^X = P^X(C) = \int_C f(x)dx = \int f(x) \cdot 1_C(x)dx$$

- $g = \sum_{i \in \underline{n}} a_i 1_{C_i}$ einfach: folgt direkt aus der Linearität.
- $g \geq 0, g_n \nearrow g, g_n$ einfach:

$$\mathbb{E}[g(X)] = \mathbb{E}[\lim_{n\to\infty} g_n(X)] = \lim_{n\to\infty} \mathbb{E}[g_n(X)] = \lim_{n\to\infty} \int g_n(x)f(x)\, dx$$

- Für $g = g^+ - g^-$, nutze Linearität. □

Korollar 11.8 Im Kontext von 11.7 gilt bei Integrierbarkeit:

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)\, dx$$

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{\mathbb{R}} x^2 f(x)\, dx - \left(\int_{\mathbb{R}} xf(x)\, dx\right)^2$$

Beispiel 11.9

i) Sei $X \sim N(\mu, \sigma^2)$. Dann:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx \overset{y=x-\mu}{=} \int_{-\infty}^{\infty} (y+\mu)\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-y^2}{2\sigma^2}}\, dy \\
&= \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-y^2}{2\sigma^2}}\, dy + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-y^2}{2\sigma^2}}\, dy = \mu \int_{-\infty}^{\infty} f_{0,\sigma^2}(y)\, dy = \mu \\
Var(X) &= \int_{-\infty}^{\infty} (X-\mu)^2 f_{\mu,\sigma^2}(x)\, dx \overset{y=\frac{x-\mu}{\sigma}}{=} \int_{-\infty}^{\infty} \frac{\sigma^2 y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy \overset{(17)}{=} \\
&= \sigma^2 \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy = \int_{-\infty}^{y} f_{0,1}(y)\, dy = \sigma^2
\end{aligned}$$

Wobei wir implizit benutzen:

$$\int_{-\infty}^{\infty} y \cdot y e^{-\frac{y^2}{2}}\, dy \overset{\text{part. Int.}}{=} \left[-y e^{-\frac{y^2}{2}}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy. \tag{17}$$

□

ii) Sei $X \sim Cauchy(1,1)$. Dann:

$$\mathbb{E}(X^+) = \int_{-\infty}^{\infty} X^+ f(x)dx = \int_0^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \frac{1}{\pi(1+x^2)} dx \geq \frac{1}{\pi}\int_1^{\infty} \frac{x}{2x^2} dx = \frac{1}{2\pi}\int_1^{\infty} \frac{1}{x} dx = \infty$$

$\mathbb{E}(X^-) = \infty$ folgt genauso, also ist X nicht integrierbar.

Satz 11.10 Sei $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ mit Dichte f_X, $g : \mathbb{R} \to \mathbb{R}$ monoton, $Y := g \circ X = g(X)$. Dann:

$$F_Y(y) = P(Y \leq y) = \int_{A_y} f_X\, dx, \qquad \text{wobei } A_y := \{x : g(x) \leq y\}.$$

Beweis: Folgt aus 11.7. Wähle $X \to 1_{(-\infty,y)}(g(X))$. □

Beispiel 11.11 Sei X $U(0,1)$, $X > 0$, setze $Y := -\frac{1}{\lambda}\log(X)$. Dann:

$$Y(x) \leq y \Leftrightarrow -\frac{1}{\lambda}\log(x) \leq y \Leftrightarrow x \geq e^{-\lambda y} \Rightarrow P(Y \leq y) = P(X \geq e^{-\lambda y}) = \begin{cases} 0 & \text{für } y < 0 \\ 1 - e^{-\lambda y} & \text{für } y \geq 0 \end{cases},$$

d.h. Y ist $Exp(\lambda)$-verteilt.

Satz 11.12 Sei $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ eine Zufallsvariable mit stetiger Dichte f_X und $g : \mathbb{R} \to \mathbb{R}$. Seien ferner $I_1, \ldots, I_n \subset \mathbb{R}$ disjunkte Intervalle mit $\mathbb{R} = \biguplus_{i \in \underline{n}} I_i$ derart, dass g auf jedem $\mathring{I}_i$ stetig differenzierbar ist und streng monoton. Dann besitzt $Y := g(X)$ die Dichte

$$f_Y(y) = \sum_{i=1}^{n} f_X(k_i(y))|k_i'(y)| 1_{g(\mathring{I}_i)}(y), \qquad \text{wobei } k_i: \text{ Umkehrfunktion von g auf } \mathring{I}_i.$$

Beweis: Vergleiche Serie 6., A.2. □

Beispiel 11.13 Sei $X \sim N(0,1)$, $Y := X^2$, wähle in 11.12. $g(x) = x^2$, $I_2 = (-\infty, 0)$, $I_1 = [0, \infty)$. Dann: $k_1 : [0, \infty] \to \mathbb{R}$, $y \mapsto \sqrt{y}$, $k_2 : [0, \infty] \to \mathbb{R}$, $y \mapsto -\sqrt{y}$.
Es folgt dann $|k_i'(y)| = \frac{1}{2\sqrt{y}}$ für $i \in \{1, 2\}$. Also:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2\pi}} \frac{1}{\sqrt{2y}} \cdot 1_{(0,\infty)}(y) + \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2\pi}} \frac{1}{\sqrt{2y}} \cdot 1_{(0,\infty)}(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2\pi}} \cdot 1_{(0,\infty)}(y)$$

Die Verteilung von Y heißt χ_1^2-Verteilung.

Die bisherigen Resultate lassen sich im Wesentlichen auf Zufallsgrößen mit Werten in $(\mathbb{R}^n, \mathcal{B}^n)$ übertragen.

Bemerkung 11.14 Es gilt $\mathcal{B}^n = \bigotimes_{i \in \underline{n}} \mathcal{B}$.

Beweis: Blatt 9, Aufgabe 1 (plus Induktion)

Definition 11.15 Es sei P ein W'maß auf $(\mathbb{R}^n, \mathcal{B}^n)$. Dann heißt

$$F : \mathbb{R}^n \to [0, 1], x \mapsto F(x) = P\left(\prod_{i \in \underline{n}} (-\infty, x_i]\right)$$

die *Verteilungsfunktion* von P (mit $x = (x_1, \dots, x_n)$).

Bemerkung 11.16

i) Wie im eindimensionalen Fall: F charakterisiert P eindeutig.

ii) Es gibt keine "schöne" Charakterisierung von Verteilungsfunktionen mehr wie in 7.8. Grund: Monotonie von $F : \mathbb{R}^n \to \mathbb{R}$ nicht definiert. Was ähnlich geht: Arbeit mit Dichten.

Definition 11.17 Ein Maß λ^n auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit

$$\lambda^n\left(\prod_{i \in \underline{n}} (a_i, b_i]\right) = \prod_{i \in \underline{n}} (b_i - a_i)$$

heißt *Lebesgue-Maß.*

Satz 11.18 Das Lebesgue-Maß existiert und ist eindeutig. Zudem gilt:

$$\lambda^n\left(\prod_{i \in \underline{n}} A_i\right) = \prod_{i \in \underline{n}} \lambda(A_i) \qquad \text{für alle } A_i, \dots, A_n \in \mathcal{B}.$$

λ^n ist also das n-fache Produkt der eindimensionalen Lebesgue-Maße.

Beweis: Die Eindeutigkeit folgt nach 6.9 wie üblich. Die Existenz folgt mit Satz von Fubini (10.6), der mit demselben Beweis auch für σ-endliche Maße (wie das Lebesgue-Maß) gilt. □

Bemerkung 11.19 Das Lebesgue-Integral verallgemeinert auch im $\mathbb{R}^n$ das Riemann-Integral. Schreibweise für $f : \mathbb{R}^n \to \mathbb{R}$ messbar:

$$\int f(x)\,dx := \int f\,d\lambda^n(x),$$

wann immer das Integral existiert. Insbesondere mit Fubini für $f \geq 0$ oder $f \in \mathcal{L}(\mathbb{R}^n, \mathcal{B}^n, \lambda^n)$:

$$\int f(x)\,dx = \int \left(\dots \left(\int f(x_1, \dots, x_n)\,dx_n \right) \dots \right) dx_1.$$

Definition 11.20 Eine Abbildung $f : (\mathbb{R}^n, \mathcal{B}^n) \to \left(\overline{\mathbb{R}}, \overline{\mathcal{B}}\right)$ messbar heißt *Dichte* eines W'maßes P (oder einer Zufallsgröße X) auf $(\mathbb{R}^n, \mathcal{B}^n)$, falls gilt:

$$P(A)(= P^X(A)) = \int_A f\,d\lambda^n = \int_A f(x)\,dx \qquad \text{für } A \in \mathcal{B}^n.$$

Bemerkung 11.21

i) Dichten sind wieder λ^n-f.ü. eindeutig.

ii) $f \geq 0$ messbar definiert genau dann die Dichte eines W'maßes, falls $\int f(x)\,dx = 1$.

Satz und Definition 11.22 Sei $(X, Y) : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^2, \mathcal{B}^2)$ eine Zufallsgröße mit Dichte f. Dann gilt:

i) X besitzt die Randdichte $f_X(x) = \int f(x, y)\,dy$

ii) X, Y sind genau dann unabhängig, wenn $f(x, y) = f_X(x) f_Y(y)$ λ^2-f.ü.

iii) Sei $y \in \mathbb{R}$ mit $f_Y(y) > 0$ gegeben. Dann definiert

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}$$

eine Dichte auf $(\mathbb{R}, \mathcal{B})$, die sogenannte *bedingte Dichte von X gegeben $Y = y$*.

Bemerkung 11.23

i) Es existiert eine n-dimensionale Version dieses Satzes.

ii) Wegen $P(Y = y) = 0$ ist $P(X \in A : Y = y)$ formal nicht definiert. Trotzdem existiert eine analoge Interpretation:

- Allgemein: Z eindimensional mit Dichte f_Z, so gilt:

$$P(z \leq Z \leq z + \Delta)\frac{1}{\Delta} = \frac{1}{\Delta} \int_z^{z+\Delta} f_Z(t)\,dt \underset{\Delta \text{ klein}}{\approx} f_Z(z).$$

- Hier:

$$\frac{1}{\Delta}P(x \leq X \leq x+\Delta : y \leq Y \leq y+\Delta) = \frac{\frac{1}{\Delta^2}P(x \leq X \leq x+\Delta : y \leq Y \leq y+\Delta)}{\frac{1}{\Delta}P(y \leq Y \leq y+\Delta)}$$
$$= \frac{\frac{1}{\Delta^2}\int_x^{x+\Delta}\int_y^{y+\Delta} f(u,v)\,dv\,du}{\frac{1}{\Delta}\int_y^{y+\Delta} f_Y(t)\,dt} \underset{\Delta \text{ klein}}{\approx} \frac{f(x,y)}{f_Y(x)}.$$

Beweis von 11.22:

i) Es gilt:

$$P(X \in A) = P((X,Y) \in A \times \mathbb{R}) \overset{\text{Fubini}}{=} \int_A \left(\int_{\mathbb{R}} f(x,y)\,dy\right) dx = \int_A f_X(x)\,dx \qquad \text{für alle } A \in \mathcal{B}.$$

ii) "$\Rightarrow$": Setze

$$\mathcal{H} = \left\{C \in \mathcal{B}^2 : \int_C f(x,y)\,dx = \int_C f_X(x)f_Y(y)\,dy\,dx\right\}.$$

Leicht: $\mathcal{H}$ ist Dynkin-System. Außerdem: Wegen

$$\int_{A\times B} f(x,y)\,dy\,dx = P(X \in A, Y \in B) = P(X \in A)P(Y \in B) = \int_{A\times B} f_X(x)f_Y(y)\,dy\,dx \tag{18}$$

für alle $A, B \in \mathcal{B}$ (und 11.14) enthält $\mathcal{H}$ einen durchschnittstabilen Erzeuger von $\mathcal{B}^2$. Also $\mathcal{H} = \mathcal{B}^2$.
"$\Leftarrow$": genauso wie (18) nur mit P außen und Integral innen.

iii) Es ist $f_{X:Y=y}(x) \geq 0$ und messbar. Es folgt:

$$\int_{-\infty}^{\infty} f_{X|Y=y}(x)\,dx = \frac{1}{f_Y(y)}\underbrace{\int_{-\infty}^{\infty} f(x,y)\,dx}_{=f_Y(y)} = 1.$$

$\square$

Definition 11.24 Seien $X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Dann heißt:

$$Cov(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$$

Kovarianz von X und Y.

Bemerkung 11.25

i) $Cov(X,Y)$ existiert wegen der Cauchy-Schwarz-Ungleichung (Blatt 8, Aufgabe 4).

ii) $Cov(X,X) = Var(X)$

iii) $Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Satz 11.26 Sind $X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ unabhängig, dann gilt: $Cov(X, Y) = 0$.

Beweis: Mit $X = X^+ - X^-, Y = Y^+ - Y^-$ und 10.2 folgt:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$
$$f(X)g(Y) = f(X)g(Y).$$

□

Definition 11.27 Es seien $X, Y \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ mit $Var(X), Var(Y) \neq 0$. Dann heißt:

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Korrelationkoeffizient von X und Y.

Bem: Der Korrelationskoeffizient ist ein Maß für den linearen Zusammenhang.

Bemerkung 11.28 Es gilt:

i) $\rho \in [-1, 1]$. (folgt aus Cauchy-Schwarz-Ungleichung)

ii) X, Y unabhängig $\Rightarrow$ $\rho = 0$. (Satz 11.26)

iii) $\rho \in \{-1, 1\} \Leftrightarrow Y = \alpha X + \beta$ f.s. für $\alpha \neq 0, \beta \in \mathbb{R}$.

Definition 11.29 Es seien $X_1, \ldots, X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Dann heißen

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])^T \text{ und } Cov(X) = (Cov(X_i, X_j))_{i,j=1}^n$$

Erwartungswert und *Kovarianzmatrix* von $X = (X_1, \ldots, X_n)^T$.

Bem: Bei Normalverteilung bestimmen Erwartungswert und Kovarianzmatrix die Verteilung (vgl. 15.7)

Satz 11.30 Es sei $X = (X_1, \ldots, X_n)^T$ wie oben.

i) $Cov(X)$ ist positiv semidefinit und symmetrisch (folglich sind die Eigenwerte reell).

ii) Ist $A \in Mat(m \times n, \mathbb{R})$, so gilt für $Y = AX$:

$$\mathbb{E}[Y] = A\mathbb{E}[X] \qquad \text{und} \qquad Cov(Y) = ACov(X)A^T.$$

("...das verhält sich irgendwie friedlich.“)

Beweis: Übung (Zettel 11, Aufgabe 1).

12 Konvergenz von Zufallgrößen

Bemerkung 12.1 Es seien $(Y_n)_n$ unabhängig auf $(\Omega, \mathcal{A}, P)$ mit $P(Y_i = 0) = (Y_i = 1) = \frac{1}{2}$. Setze $X_n = \prod_{i=1}^n Y_i$.
Intuitiv: $X_n \to 0$, da irgendein $Y_i = 0$.
Problem: Für das $\omega \in \Omega$ mit $Y_i(\omega) = 1$ für alle $i \in \mathbb{N}$ gilt:

$$X_n(\omega) \to 1$$

(also keine punktweise Konvergenz gegen 0.

Von nun an: $(X_n)_n$ und X sind Zufallsgrößen auf $(\Omega, \mathcal{A}, P)$ mit Werten in $\mathbb{R}$ (oder $\mathbb{R}^d$).

Definition 12.2 $(X_n)_n$ *konvergiert P-fast sicher gegen X*, falls

$$P(\{\omega : \underbrace{\lim_{n\to\infty} X_n(\omega) = X(\omega)}_{\text{punktweise Konvergenz}}\}) = 1.$$

Notation: $X_n \to X$ f.s. (bzw. P-f.s., wenn P nicht eindeutig), auch $X_n \overset{\text{f.s.}}{\to} X$.

Beachte: $\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} \in \mathcal{A}$ gemäß Blatt 6, Aufgabe 4.

Beispiel 12.3 Wegen $P(X_n = 1) = P(Y_i = 1 \text{ für } i \in \underline{n}) = \frac{1}{2^n} \to 0$ folgt $X_n \overset{\text{f.s.}}{\to} 0$ in Bemerkung 12.1.

Definition 12.4 Sind $X_n, X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$, so *konvergiert X_n in p-ten Mittel gegen X*, falls

$$\mathbb{E}[|X - X_n|^p] \to 0 \qquad (1 \leq p < \infty).$$

Notation: $X_n \overset{\mathcal{L}^p}{\to} X$. (vgl. L_p-Konvergenz aus Analysis)

Bem: Die Grenzwerte der verschiedenen Konvergenzbegriffe sind f.s. gleich (ohne Beweis).

Bemerkung 12.5 Aus der umgekehrten Dreiecksungleichung folgt leicht:

$$X_n \overset{\mathcal{L}^1}{\to} X \;\Rightarrow\; \mathbb{E}[X_n] \to \mathbb{E}[X] \text{ und } \mathbb{E}[|X_n|] \to \mathbb{E}[|X|].$$

Die Umkehrung gilt iA. nicht.

Definition 12.6 $(X_n)_n$ *konvergiert in P-Wahrscheinlichkeit* oder *P-stochastisch gegen X*, falls für jedes $\varepsilon > 0$ gilt:

$$\lim_{n\to\infty} (|X_n - X| \geq \varepsilon) = \lim_{n\to\infty} (\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0.$$

Notation: $X_n \overset{P}{\to} X$.

Satz 12.7 Es gilt: $X_n \overset{P}{\to} X \Leftrightarrow \mathbb{E}\Big[\frac{|X_n - X|}{1+|X_n-X|}\Big] \to 0.$

Beweis: Sei oBdA $X = 0$ (ansonsten betrachte Differenzen).

"$\Rightarrow$": Es gilt:

$$\frac{|X_n|}{1+|X_n|} \leq \underbrace{\frac{|X_n|}{1+|X_n|}}_{\leq 1} 1_{\{|X_n|\geq\varepsilon\}} + \varepsilon \underbrace{1_{\{|X_n|<\varepsilon\}}}_{\geq 1} \Rightarrow \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right] \leq P(|X_n| \geq \varepsilon) + \varepsilon \text{ für alle } \varepsilon > 0$$

$$\Rightarrow \limsup_{n\to\infty} \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right] \leq \limsup_{n\to\infty} P(|X_n| \geq \varepsilon) + \varepsilon = \varepsilon \qquad \text{für alle } \varepsilon > 0$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right] = 0.$$

"$\Leftarrow$": Für $f(x) = \frac{x}{1+x}$ gilt $f'(x) = \frac{1}{(1+x)} > 0$, also ist f streng monoton. Damit folgt:

$$\frac{\varepsilon}{1+\varepsilon} 1_{\{|X_n|\geq\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|} 1_{\{|X_n|\geq\varepsilon\}} \leq \frac{|X_n|}{1+|X_n|}$$

$$\Rightarrow \frac{\varepsilon}{1+\varepsilon} \limsup_{n\to\infty} P(|X_n| \geq \varepsilon) \leq \limsup_{n\to\infty} \mathbb{E}\left[\frac{|X_n|}{1+|X_n|}\right] = 0 \qquad \text{für alle } \varepsilon > 0$$

$$\Rightarrow \lim_{n\to\infty} P(|X_n| \geq \varepsilon) = 0.$$

□

Bem: Der Satz 12.7 gilt für jede beliebige, streng monoton wachsende Funktion.

Die drei bisher etablierten Konvergenzbegriffe sind verschieden stark.

Satz 12.8

i) $X_n \overset{\mathcal{L}^p}{\to} X \Rightarrow X_n \overset{P}{\to} X.$

ii) $X_n \overset{\text{f.s.}}{\to} X \Rightarrow X_n \overset{P}{\to} X.$

Beweis:

i) Es gilt für $\varepsilon > 0$:

$$P(|X_n - X| \geq \varepsilon) = \mathbb{E}\big[1_{\{|X_n-X|\geq\varepsilon\}}\big] \leq \mathbb{E}\left[\frac{|X_n - X|^p}{\varepsilon^p} 1_{\{|X_n-X|\geq\varepsilon\}}\right] = \frac{1}{\varepsilon^p}\mathbb{E}\left[|X_n - X|^p\right] \to 0.$$

ii) Wegen $\frac{|X_n-X|}{1+|X_n-X|} \leq 1$ und $|X_n - X| \to 0$ f.s. folgt mit majorisierter Konvergenz:

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{|X_n - X|}{1+|X_n - X|}\right] = \mathbb{E}\left[\lim_{n\to\infty} \frac{|X_n - X|}{1+|X_n - X|}\right] = \mathbb{E}[0] = 0.$$

Die Aussage folgt mit 12.7. □

Bemerkung 12.9 Es gibt für beide Rückrichtungen in Satz 12.8 Gegenbeispiele (s. Übung).

Satz 12.10 Gilt $X_n \xrightarrow{P} X$, so existiert eine Teilfolge $(X_{n_k})_k$ mit $\lim_{k\to\infty} X_{n_k} = X$ P-f.s.

Beweis: Es gilt: $\forall\, k \in \mathbb{N} \exists\, n_k \in \mathbb{N} : P\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \frac{1}{k^2}$. Wähle oBdA $n_{k+1} > n_k$. Sei $\varepsilon > 0$ fest und $l \in \mathbb{N}$ mit $\varepsilon > \frac{1}{l}$. Dann gilt:

$$\sum_{k\geq l} P\left(|X_{n_k} - X| \geq \varepsilon\right) \leq \sum_{k\geq l} P\left(|X_n - X| \geq \frac{1}{k}\right) \leq \sum_{k\geq l} \frac{1}{k^2} < \infty.$$

Insbesondere $\sum_{k\geq l} P\left(|X_{n_k} - X| \geq \varepsilon\right) \leq \infty$ und mit Borel-Cantelli:

$$P(|X_{n_k} - X| \geq \varepsilon \text{ unendlich oft}) = 0 \;\Leftrightarrow\; P\left(\lim_{k\to\infty} X_{n_k} = X\right) = 1.$$

□

Satz 12.11 Gilt $X_n \xrightarrow{P} X$ und ist $|X_n| < Y$ mit $Y \in \mathcal{L}^p$, so folgt: $X \in \mathcal{L}^p$ und $X_n \xrightarrow{\mathcal{L}^p} X$.

Beweis:

- $X \in \mathcal{L}^p$: Da $|X_n| < Y \in \mathcal{L}^p$, folgt $X_n \in \mathcal{L}^p$. Weiterhin:

$$\begin{aligned}
&P(|X| \geq Y + \varepsilon) \leq P(|X| \geq |X_n| + \varepsilon) = P(|X| - |X_n| \geq \varepsilon) \overset{\Delta\text{-Ungl.}}{\leq} P(|X_n - X| \geq \varepsilon) \\
&\Rightarrow\; \forall\, \varepsilon > 0 : P(|X| \geq Y + \varepsilon) = 0 \\
&\Rightarrow\; P(|X| \geq Y) \overset{4.19}{=} \lim_{m\to\infty} P\left(|X| \geq Y + \frac{1}{m}\right) = 0 \\
&\Rightarrow\; |X| \leq Y \text{ f.s.} \;\Rightarrow\; |X| \in \mathcal{L}^p \text{ und } X \in \mathcal{L}^p.
\end{aligned}$$

- $X_n \xrightarrow{\mathcal{L}^p} X$: Angenommen, es gilt $X_n \overset{\mathcal{L}^p}{\nrightarrow} X$, so existiert eine Folge $(X_{n_k})_{k\in\mathbb{N}}$ mit $\mathbb{E}[|X_{n_k} - X|^p] \geq \varepsilon$ für ein $\varepsilon > 0$ und $\forall k \in \mathbb{N}$.

  Klar: $X_{n_k} \underset{k\to\infty}{\xrightarrow{P}} X$. Nach 12.10 existiert eine Teilfolge $X_{n_{k_l}} \underset{l\to\infty}{\longrightarrow} X$ f.s. Wegen $|X_{n_{k_l}} - X| \leq 2Y$ P-f.s., folgt mit majorisierter Konvergenz

$$\lim_{l\to\infty} \mathbb{E}\left[\left|X_{n_{k_l}} - X\right|^p\right] = \mathbb{E}\left[\lim_{l\to\infty} \left|X_{n_{k_l}} - X\right|^p\right] = 0,$$

  da $X_{n_{k_l}} \to X$ fast sicher. Widerspruch. □

Bemerkung 12.12 Ist $f : \mathbb{R} \to \mathbb{R}$ stetig, so gilt:

(i) $X_n \overset{n\to\infty}{\longrightarrow} X$ P-fast-sicher $\Rightarrow$ $f(X_n) \overset{n\to\infty}{\longrightarrow} f(X)$ P-fast-sicher.

(ii) $X_n \xrightarrow{P} X \;\Rightarrow\; f(X_n) \xrightarrow{P} f(X)$.

Beweis: Übung. □

13 Gesetze der großen Zahlen

Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von reellwertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Gesetze der großen Zahlen befassen sich mit der Konvergenz von:

$$\frac{1}{n}S_n, \qquad \text{wobei } S_n := \sum_{i\in\underline{n}} X_i$$

Satz 13.1 (Starkes Gesetz der großen Zahlen) Es seien $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten (i.i.d) Zufallsvariablen mit $\mathbb{E}[X_n] = \mu$ und $Var(X_n) = \sigma^2 < \infty$. Dann gilt:

$$\frac{1}{n}S_n = \frac{1}{n}\sum_{i\in\underline{n}} X_i \overset{\text{f.s.}}{\to} \mu \qquad \text{und} \qquad \frac{1}{n}S_n \in \mathcal{L}^2.$$

Satz 13.2 Als *schwaches Gesetz der großen Zahlen* bezeichnet man $\frac{1}{n}S_n \overset{P}{\to} \mu$, was als Korollar direkt aus Satz 13.1 folgt, sich aber auch leicht mit Hilfe des Tschebyschoff-Ungleichung zeigen lässt.

Beweis von 13.1: Sei o. E. $\mu = 0$ (sonst: $Z_i := X_i - \mu$). Dann gilt für $Y_n = \frac{1}{n}S_n$:

$$\mathbb{E}\big[Y_n\big] = \frac{1}{n}\sum_{i\in\underline{n}} \mathbb{E}\big[X_i\big] = 0$$

$$\mathbb{E}\big[{Y_n}^2\big] = \frac{1}{n^2}\sum_{i,j\in\underline{n}} \mathbb{E}\big[X_iX_j\big] = \frac{1}{n^2}\sum_{i\in\underline{n}} \mathbb{E}\big[X_i^2\big] = \frac{\sigma^2}{n} \to 0, \text{ da } \mathbb{E}\big[X_i\big]\mathbb{E}\big[X_j\big] = 0 \text{ für } i \neq j.$$

Also: $\mathbb{E}\big[{Y_n}^2\big] \to 0 \Leftrightarrow Y_n \overset{\mathcal{L}^2}{\to} 0$.

Zeige Konvergenz fast sicher. Es gilt:

$$\sum_{n=1}^{\infty} \mathbb{E}\big[{Y_{n^2}}^2\big] = \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty$$

Aus Korollar 9.22 folgt:

$$\sum_{n=1}^{\infty} \mathbb{E}\big[{Y_{n^2}}^2\big] < \infty \text{ fast sicher } \Rightarrow {Y_{n^2}}^2 \to 0 \text{ fast sicher } \Leftrightarrow Y_{n^2} \to 0 \text{ fast sicher.}$$

Sei nun $n \in \mathbb{N}$ beliebig und $p(n)$ so, dass $p(n)^2 \leq n < (p(n)+1)^2$ (also $p(n) = \lfloor\sqrt{n}\rfloor$). Dann:

$$Y_n - \frac{p(n)^2}{n}Y_{p(n)^2} = \frac{1}{n}\sum_{i=p(n)^2+1}^{n} X_i \quad \text{und}$$

$$\mathbb{E}\left[\left(Y_n - \frac{p(n)^2}{n}Y_{p(n)^2}\right)^2\right] = \frac{1}{n^2}(n - p(n)^2)\sigma^2 \leq (2p(n)+1)\frac{\sigma^2}{n^2} \leq (2\sqrt{n}+1)\frac{\sigma^2}{n^2} \leq \frac{3\sigma^2}{n^{\frac{3}{2}}}.$$

Wie nach (19):

$$\lim_{n\to\infty}\big(Y_n - \frac{p(n)^2}{n}Y_{p(n)^2}\big) = 0 \text{ fast sicher, weil } \sum_{i=1}^{\infty}\mathbb{E}\left[\left(Y_n - \frac{p(n)^2}{n}Y_{p(n)^2}\right)^2\right] < \infty.$$

Mit $\frac{p(n)^2}{n} \to 1$ und $Y_{p(n)^2} \to 0$ fast sicher, folgt $Y_n \to 0$ fast sicher. □

Bemerkung 13.3 Eine deutliche Verschärfung ist das *starke Gesetz von Kolmogorov*:
Es seien $(X_n)_{n\in\mathbb{N}}$ i.i.d. und $\mu \in \mathbb{R}$. Dann gilt:

$$\frac{1}{n}S_n \to \mu \text{ fast sicher } \Leftrightarrow \mu = \mathbb{E}[X_1]$$

Der Beweis ist aufwendig, aber machbar.

Beispiel 13.4

i) Es seien $(X_n)_{n\in\mathbb{N}}$ i.i.d. mit $P(X_1 = 1) = p$ und $P(X_1 = 0) = 1 - p$, dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to p \text{ fast sicher}$$

Also: Der Anteil der Erfolge bei Wiederholung eines Spiels konvergiert fast sicher gegen p, die Wahrscheinlichkeit eines Erfolges.

ii) Sei $f : [0,1] \to \mathbb{R}$ messbar mit $\int_0^1 |f(x)|\,dx < \infty$. Dann folgt aus Bemerkung 13.3 mit $(U_n)_{n\in\mathbb{N}}$ i.i.d. und $U_n \sim U(0,1)$:

$$\frac{1}{n}\sum_{i=1}^{n} f(U_i) \to \mathbb{E}[f(U_1)] = \int_0^1 f(x)\,dx \text{ fast sicher.}$$

Diese "Monte-Carlo-Simulation" dient der Approximation schwer zu berechnender Integrale.

14 Schwache Konvergenz

Im Gegensatz zu den Konvergenzbegriffen aus Kapitel 12, die sich mit der Konvergenz von Zufallsvariablen befassen, ist schwache Konvergenz ein Begriff für Maße bzw. Verteilungen.

Definition 14.1

i) Es seien $(\mu_n)_{n\in\mathbb{N}}, \mu$ W'maße auf $\mathbb{R}$ (bzw. $\mathbb{R}^n$). Dann *konvergiert μ_n schwach gegen μ*, falls:

$$\lim_{n\to\infty} \int f\, d\mu_n = \int f, d\mu \qquad \text{für alle } f \in C_b(\mathbb{R}),$$

wobei $C_b(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ stetig und beschränkt}\}$.

Notation: $\mu_n \xrightarrow{w} \mu$.

ii) Sind $(X_n)_{n\in\mathbb{N}}, X$ Zufallsgrößen mit Werten in $\mathbb{R}$ und Verteilungen P^{X_n} bzw. P^X, so *konvergiert $(X_n)_{n\in\mathbb{N}}$ in Verteilung gegen X*, falls $P^{X_n} \xrightarrow{w} P^X$.

Notation: $X_n \xrightarrow{w} X$ oder $X_n \xrightarrow{\mathcal{D}} X$.

Bemerkung 14.2 Das Grenzmaß μ bei schwacher Konvergenz ist eindeutig, denn

$$\int f\, d\mu = \int f\, d\nu \;\forall f \in C_b(\mathbb{R}) \;\Rightarrow\; \mu = \nu$$

(vgl. Blatt 9, Aufgabe 2 & 3).

Satz 14.3 Für reellwertige Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ und X gilt:

$$X_n \xrightarrow{w} X \;\Rightarrow\; \mathbb{E}[f(X_n)] \underset{n\in\mathbb{N}}{\longrightarrow} \mathbb{E}[f(X)] \;\forall f \in C_b(\mathbb{R})$$

Beweis: $\mathbb{E}[f(X)] = \int f \circ X\, dP = \int f\, dP^X$. □

Bemerkung 14.4 $(X_n)_{n\in\mathbb{N}}$ und X müssen nicht auf demselben W'raum definiert sein, um schwache Konvergenz zu definieren.

Satz 14.5 Es seien $(X_n)_{n\in\mathbb{N}}, X$ reellwertige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$.

i) Gilt $X_n \xrightarrow{P} X$, so folgt $X_n \xrightarrow{\mathcal{D}} X$.

ii) Gilt $X_n \xrightarrow{\mathcal{D}} X$ und $X = a$ f.s., so folgt $X_n \xrightarrow{P} X$.

Beweis:

i) Sei $f \in C_b(\mathbb{R})$. Dann:

   (a) f stetig: $f(X_n) \xrightarrow{P} f(X)$ (Bemerkung 12.12)

(b) $f(X_n) \xrightarrow{\mathcal{L}^1} f(X)$ (f beschränkt, Satz 12.11)

(c) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ (Bemerkung 12.5) $\Rightarrow$ $X_n \xrightarrow{\mathcal{D}} X$.

ii) $f(x) = \frac{|x-a|}{1+|x-a|}$ ist stetig und beschränkt durch 0 und 1. Also:

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \overset{X=a}{=} 0 \text{ f.s.} \overset{\text{S.12.7}}{\Longrightarrow} X_n \xrightarrow{P} a.$$

□

Bemerkung 14.6 Zusammenfassend ergibt sich für reellwertige $(X_n)_{n\in\mathbb{N}}$ auf $(\Omega, \mathcal{A}, P)$: BILD.

Satz 14.7 (Charakterisierung der schwachen Konvergenz) Seien $(X_n)_n$ und X reellwertig mit Verteilungsfunktion $(F_n)_n$ bzw. F.

i) Aus $X_n \xrightarrow{\mathcal{D}} X$ folgt $F_n(x) \underset{n\to\infty}{\longrightarrow} F(x)$ für alle x aus der (dichten) TM $D := \underbrace{\{x : F(x) = F(-x)\}}_{\text{Menge der Stetigkeitsstellen}}$.

ii) Gilt $F_n(x) \underset{\text{p.w.}}{\longrightarrow} F(x)$ für alle $x \in \Delta$ dicht, so folgt $X_n \xrightarrow{\mathcal{D}} X$.

Beweis:

i) Definiere zu $x \in \mathbb{R}$ und $p \geq 1$ folgende Funktionen:

$$g_p = \begin{cases} 1 & , y \in \left(-\infty, x - \frac{1}{p}\right) \\ p(x-y) & , y \in \left[x - \frac{1}{p}, x\right] \\ 0 & , y \in (x, \infty) \end{cases} \quad \text{und } f_p = \begin{cases} 1 & , y \in (-\infty, x) \\ p(x-y)+1 & , y \in \left[x, x - \frac{1}{p}\right] \\ 0 & , y \in \left(x - \frac{1}{p}, \infty\right) \end{cases}.$$

Idee: f_p und g_p schachteln $1_{\{y\leq x\}}(y)$ ein. Dann folgt:

$$\mathbb{E}[g_p(X)] \underset{\substack{X_n \xrightarrow{\mathcal{D}} X, \\ g_p \in C_b(\mathbb{R})}}{\longleftarrow} \mathbb{E}[g_p(X_n)] \leq \underbrace{\mathbb{E}\left[1_{\{X_n \leq x\}}\right]}_{F_n(x)} \leq \mathbb{E}[f_p(X_n)] \to \mathbb{E}[f_p(X)].$$

Außerdem: $f_p(y) \underset{p\in\infty}{\longrightarrow} 1_{(-\infty,x]}(y)$ und $g_p(y) \underset{p\to\infty}{\longrightarrow} 1_{(-\infty,x)}(y)$. Mit majorisierter Konvergenz folgt:

$$\lim_{p\to\infty} \mathbb{E}[g_p(X)] = F(-x) \qquad \text{und} \qquad \lim_{p\to\infty} \mathbb{E}[f_p(X)] = F(x).$$

Insgesamt folgt:

$$F(-x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x)$$

und mit $x \in D$ die Behauptung.

ii) Sei $f \in C_b(\mathbb{R}), f \neq 0$ und $\varepsilon > 0$ beliebig. Da Δ dicht, existieren $r, s \in \Delta$ mit

$$P(X \in (r, s]) = 1 - F(s) + F(r) < \varepsilon.$$

Wegen $F_n(x) \to F(x)$ für $x \in \Delta$ folgt:

$$\exists\, N_1 \in \mathbb{N} : P(X_n \in (r,s]) = 1 - F_n(s) + F_n(r) < 2\varepsilon \qquad \text{für alle } n \geq N_1$$

Auf $[r,s]$ ist f gleichmäßig stetig, also existieren $r = r_0 < r_1 < \ldots < r_k = s$ (wobei $r_j \in \Delta$) mit $|f(x) - f(r_j)| \leq \varepsilon$ für alle x mit $x \in [r_{j-1}, r_j)$.
Also: Mit $g(x) = \sum_{j=1}^{k} f(r_j) 1_{(r_{j-1}, r_j]}(x)$ gilt $|f(x) - g(x)| \leq \varepsilon$ für $x \in (r,s]$ und

$$\Big|\mathbb{E}[f(X_n)] - \mathbb{E}[g(X_n)]\Big| \leq \varepsilon P(X_n \in (r,s]) + \|f\|_\infty P\left(X_n \notin (r,s]\right) \leq \varepsilon + \|f\|_\infty P(X_n \overset{???}{\in} (r,s]).$$

Analoge Formel: X_n durch X ersetzen. Zudem existiert ein $N_2 \in \mathbb{N}$ mit

$$|F_n(r_j) - F(r_j)| \leq \frac{\varepsilon}{2k\|f\|_\infty} \qquad \text{für alle } n \geq N_2.$$

Also:

$$\Big|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]\Big| = \left|\sum_{j=1}^{k} f(r_j)\left(F_n(r_j) - F_n(r_{j-1}) - (F(r_j) - F(r_{j-1}))\right)\right| \leq \varepsilon \qquad \text{für alle } n \geq N_2.$$

Insgesamt gilt für $n \geq N := \max\{N_1, N_2\}$:

$$\begin{aligned}\Big|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]\Big| &\leq \Big|\mathbb{E}[f(X_n)] - \mathbb{E}[g(X_n)]\Big| + \Big|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]\Big| + \Big|\mathbb{E}[g(X)] - \mathbb{E}[f(X)]\Big| \\ &\leq 2\varepsilon\|f\|_\infty + \varepsilon + (\varepsilon\|f\|_\infty + \varepsilon) = 3(1 + \|f\|_\infty)\varepsilon.\end{aligned}$$

□

Beispiel 14.8

i) Es seien $(\mu_n)_n$ Dirac-Maße, dh. für $\alpha_n \in \mathbb{R}$ gilt $\mu_n(\{\alpha_n\}) = 1$, also $F_n(x) = 1_{[a_n,\infty)}(x)$. Dann folgt: $\mu_n \overset{w}{\to} \mu \;\Leftrightarrow\; \alpha_n \to \alpha$ mit $\mu(\{\alpha\}) = 1$, da punktweise Konvergenz gegen eine Verteilungsfunktion $F(x)$ die Form $F(x) = 1_{[\alpha,\infty)}(x)$ impliziert.

ii) Es seien $(X_n)_n, X$ Zufallsgrößen mit Dichten f_n bzw. f. Gilt $f_n(x) \to f(x)$ für λ-f.a. x und $f_n(x) \leq g(x)$ mit $g \in \mathcal{L}^1$, so folgt mit majorisierter Konvergenz:

$$F_n(x) = \int_{-\infty}^{x} f_n(y)\, dy \to \int_{-\infty}^{x} f(y) dy = F(x).$$

Also folgt $X_n \overset{\mathcal{D}}{\to} X$.

Bem: Man kann leicht mit einem alternativen Beweis auf majorisierte Konvergenz verzichten.

Definition 14.9 Eine Folge $(\mu_n)_n$ von W'maßen auf $(\mathbb{R}, \mathcal{B})$ heißt *straff*, falls

$$\lim_{m\to\infty} \sup_n \mu_n([-m,m]^c) = 0.$$

Bsp: Sei $(\mu_n)_n$ eine Folge von Maßen auf $(\mathbb{R}, \mathcal{B})$, wobei μ_i durch die Verteilungsfunktion von $U(i, i+1)$ definiert wird für jedes $i \in \mathbb{N}$. Diese Folge $(\mu_n)_n$ ist nicht straff.

Satz 14.10 (Auswahlsatz von Helly) Eine straffe Folge $(\mu_n)_n$ besitzt eine schwach konvergente Teilfolge.

Beweis: Sei F_n die Verteilungsfunktion zu μ_n. Zu jedem $x \in \mathbb{R}$ ist $(F_n(x))_n$ eine beschränkte Folge, besitzt folglich nach Bolzano-Weierstraß eine konvergente Teilfolge $(F_{n_k}(x))_k$.

Sei nun $\mathbb{Q} = \{r_1, r_2, \ldots\}$ (Abzählung existiert da $\mathbb{Q} \sim \mathbb{N}$). Setze $G(r_1) = \lim_{k\to\infty} F_{n_{1,k}}(r_1)$, wobei wir die zu r_1 gehörende Teilfolge $n_{1,k}$ von oben wählen. Dann $G(r_2) = \lim_{k\to\infty} F_{n_{2,k}}(r_2)$ mit der zu r_2 gehörenden Teilfolge von $n_{1,k}$ und iterativ $G(r_j) = \lim_{k\to\infty} F_{n_{j,k}}(r_j)$ mit der zu r_j gehörenden Teilfolge von $n_{j-1,k}$. Offenbar gilt: $r < s \Rightarrow G(r) \leq G(s)$ für $r, s \in \mathbb{Q}$.

Setze nun $n_k := n_{k,k}$ ("Diagonal-Folgen-Argument"). Klar: $G(r_j) = \lim_{k\to\infty} F_{n_k}(r_j)$ für alle j, da n_k für $k \geq j$ eine Teilfolge von $n_{j,k}$ ist. Setze nun $F(x) = \inf_{\substack{y\in\mathbb{Q}\\ y>x}} G(y)$. Es gilt:

i) F ist monoton wachend, da G monoton.

ii) F ist rechtsstetig (wegen $y > x$).

iii) $F(x) = G(x)$ für alle $x \in \mathbb{Q}$.

Sei $\varepsilon > 0$. Dann existieren $m \in \mathbb{N}$ mit $\mu_n([-m,m]^c) \leq \varepsilon$ für alle $n \in \mathbb{N}$. Es folgt: $F_n(x) \leq \varepsilon \; \forall x < (-m)$ und $F_n(x) \geq 1 - \varepsilon \; \forall x > m$ für alle $n \in \mathbb{N}$. Dann gilt $G(x) \leq \varepsilon \; \forall x < (-m)$ und $G(x) \geq 1 - \varepsilon \; \forall x > m$, weil $F_{n_k}(x) \to F(x)$ für alle $x \in \mathbb{Q}$. Nach Definition: $F(x) < \varepsilon \; \forall x < (-m)$ und $F(x) \geq 1 - \varepsilon \; \forall x \geq m$. Also ist F Verteilungsfunktion.

Wir zeigen nun: $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ für alle x mit $F(x) = F(-x)$. Wähle solch ein x. Dann existiert zu $\varepsilon > 0$ (wegen Stetigkeit) ein $x_0 < x$ mit

$$F(x) - \varepsilon < F(x_0) \leq F(x).$$

Insbesondere existieren $y, z \in \mathbb{Q}$ mit $y < x < z$ und

$$F(x) - \varepsilon \overset{(*)}{\leq} G(y) \leq F(x) \overset{(**)}{\leq} G(z) \leq F(x) - \varepsilon,$$

wobei (*) wegen $F(x_0) = \inf_{y>x_0} G(y)$ und (**) wegen $F(x) = \inf_{z>x} G(z)$ gilt.

Also gilt für großes k:

$$\begin{aligned} &F(x) - 2\varepsilon \leq F_{n_k}(y) \leq F_{n_k}(x) \leq F_{n_k}(z) \leq F(x) + 2\varepsilon \\ &\Rightarrow \; F(x) - 2\varepsilon \leq F_{n_k}(y) \leq \liminf_{k\to\infty} F_{n_k}(x) \leq \limsup_{k\to\infty} F_{n_k}(x) \leq F(x) + 2. \end{aligned}$$

Also: $\lim_{k\to\infty} F_{n_k}(x) = F(x)$. $\square$

15 Charakteristische Funktionen

Definition 15.1 Es sei μ W'maß auf $(\mathbb{R}^n, \mathcal{B}^n)$. Dann heißt für $t \in \mathbb{R}^n$

$$\hat{\mu}(t) = \varphi_\mu(t) = \int e^{i\langle t,x\rangle}\, d\mu(x) = \int \cos(\langle t, x\rangle)\, d\mu(x) + i \int \sin(\langle t, x\rangle)\, d\mu(x)$$

die *charakteristische Funktion von* μ.

Analog: Ist X Zufallsgröße in $\mathbb{R}^n$, so gilt:

$$\varphi_X(t) = \int e^{i\langle t,x\rangle}\, d\mu^X(x) = \mathbb{E}\Big[e^{i\langle t,x\rangle}\Big].$$

Bem: Diese Funktionen sind wohldefiniert, da über eine beschränkte Funktion integriert wird.

Satz 15.2 (Eindeutigkeitssatz) Es seien μ und ν W'maße auf $(\mathbb{R}^n, \mathcal{B}^n)$. Gilt $\varphi_\mu(t) = \varphi_\nu(t)$ für alle $t \in \mathbb{R}$, so folgt $\mu = \nu$.

Beweis: Die kompakten Mengen des $\mathbb{R}^n$ bilden ein durchschittstabiles Erzeugendensystem von $\mathcal{B}^n$. Also genügt es zu zeigen: $\mu(K) = \nu(K)$ für alle K kompakt.

Wir zeigen

$$\int f\, d\mu = \int f\, d\nu \tag{*}$$

für alle stetigen Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ mit $0 \le f \le 1$ und kompaktem Träger $\overline{\{x : f(x) \neq 0\}}$. Dann gilt für

$$f_n(x) = \begin{cases} 1 & , x \in K \\ 0 & , d(x, K) \ge \frac{1}{m} \\ 1 - nd(x, K) & , \text{sonst} \end{cases}$$

mit $m \in \mathbb{N}$ aus (*), dass $\int f_m\, d\mu = \int f_m\, d\nu$. Wegen $f_m \nearrow 1_K$ folgt:

$$\mu(K) = \int 1_K\, d\mu \overset{\text{maj.K.}}{=} \lim_{m\to\infty} \int f_m\, d\mu = \lim_{m\to\infty} \int f_m\, d\nu \overset{\text{maj.K.}}{=} \int 1_K\, d\nu = \nu(K).$$

Beweis von (*): Wähle zu $\varepsilon > 0$ ein $N \in \mathbb{N}$ so, dass $\{x : f(x) \neq 0\} \in [-N, N]^n =: B_N$ und $\max\{\mu(B_N^c), \nu(B_N^c)\} < \varepsilon$.

Approximationssatz von Weierstraß: Es existiert $g : \mathbb{R}^n \to \mathbb{C}$ der Form

$$g(x) = \sum_{j=1}^{m} c_j\, exp\left(i \left\langle \frac{\pi}{N} t_j, x \right\rangle\right)$$

mit $c_j \in \mathbb{C}$ und $t_j \in \mathbb{Z}^n$, so dass gilt:

$$\sup\{|f(x) - g(x)| : x \in B_N\} \le \varepsilon.$$

Klar: $\|g\|_\infty \le 1 + \varepsilon$. Dann gilt:

$$\left|\int f\, d\mu - \int f\, d\nu\right| \overset{\Delta\text{-Ungl.}}{\le} \left|\int f\, d\mu - \int g\, \mu\right| - \underbrace{\left|\int g\, d\mu - \int g\, \nu\right|}_{=0 \text{ n.Vor.}} + \left|\int g\, d\nu - \int f\, d\nu\right|.$$

Nun folgt etwa:

$$\begin{aligned}
\left|\int (f-g)\,d\mu\right| &\le \left|\int_{B_N} (f-g)\,d\mu\right| + \left|\int_{B_N^c} f\,d\mu\right| + \left|\int_{B_N^c} g\,d\mu\right| \\
&\le \int_{B_N^c} |f-g|\,d\mu + \mu(B_N^c) + (1+\varepsilon)\mu(B_N^c) \\
&\le \varepsilon\mu(B_N^c) + (1+\varepsilon)\varepsilon.
\end{aligned}$$

$\square$

Beispiele 15.3

i) *Dirac-Maß:* $a \in \mathbb{R}^n$ und Maß δ_a liefern $\varphi_{\delta_a}(t) = e^{i\langle t,a\rangle}$.

ii) *Gleichverteilung:* Ist $X \sim U(-a,a)$, so folgt:

$$\varphi_X = \int_{-a}^{a} e^{i\langle t,x\rangle} \frac{1}{2a}\,dx \overset{\langle t,x\rangle = tx}{=} \frac{1}{2a}\left[\frac{1}{it}e^{itx}\right]_{x=-a}^{x=a} = \frac{1}{at}\frac{e^{ita}-e^{-ita}}{2i} = \frac{\sin(at)}{at}.$$

Definition 15.4

i) Das W'maß auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit Dichte

$$\varphi(x_1,\dots,x_n) = (2\pi)^{-\frac{n}{2}}\,exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$$

heißt *Standardnormalverteilung* in $\mathbb{R}^n$.

Notation: $X \sim \mathcal{N}_n(0, I_n)$.

Beachte: Dies ist die Verteilung von $(Y_1,\dots,Y_n)$ mit $y_i \sim N(0,1)$ i.i.d.

ii) Ein W'maß P auf $(\mathbb{R}^n, \mathcal{B}^n)$ heißt *Normalverteilung*, wenn $A \in Mat(n\times n, \mathbb{R})$ und $b \in \mathbb{R}^n$ existieren, so dass P das Bildmaß der Standardnormalverteilung P_{st} unter $\Phi(x) = Ax + B$ ist, also

$$P(\tilde{A}) = P_{st}^{\Phi}(\tilde{A}) = P_{st}(\Phi^{-1}(\tilde{A})) \qquad \text{für alle } \tilde{A} \in \mathcal{B}^n.$$

Mit anderen Worten: $X \sim \mathcal{N}_n(0, I_n) \;\Rightarrow\; Y = Ax + b$ normalverteilt.

Bemerkung 15.5

i) Man kann zeigen (hier ohne Beweis): P in 15.4(ii) besitzt genau dann eine Dichte, wenn $A \in GL(n,\mathbb{R})$. In diesem Fall:

$$\varphi(x,b,\Sigma) = \frac{1}{\sqrt{(2\pi)^n \det\Sigma}}exp\left(-\frac{1}{2}(x-b)^T\Sigma^{-1}(x-b)\right) \qquad \text{mit } b \in \mathbb{R}^n, \Sigma = AA^T.$$

Die Verteilung hängt also nur noch von Σ und nicht mehr von A ab.

ii) Für $X \sim \mathcal{N}_n(0, I_n)$ gilt: $\mathbb{E}[X] = 0$ und $Cov(X) = I_n$. Mit 11.30 folgt für $Y = Ax + b$:

$$\mathbb{E}[Y] = b \qquad \text{und} \qquad Cov(Y) = AA^T.$$

Beispiel 15.6

i) Für $X \sim N(0,1)$ gilt:

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int e^{itx} e^{-\frac{x^2}{2}}\, dx = e^{-\frac{t^2}{2}}.$$

Beweis:

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_X(t) \overset{\text{maj.K.}}{=} \frac{1}{\sqrt{2\pi}} \int \underbrace{ie^{itx}}_{u} \underbrace{xe^{-\frac{x^2}{2}}}_{v'}\, dx = \underbrace{\frac{1}{\sqrt{2\pi}} \left[-ie^{itx}e^{-\frac{x^2}{2}}\right]_{x=-\infty}^{x=\infty}}_{=0, \text{ da } \|e^{itx}\|_\infty = 1} + \frac{1}{\sqrt{2\pi}} \int i^2 t e^{itx} e^{-\frac{x^2}{2}}\, dx = -t\varphi_X(t)$$

Mit "Trennung der Variablen" folgt $\varphi_X(t) = e^{-t^2 \frac{1}{2}} e^C$ wobei $C = 0$, da $\varphi_X(0) = 1$. □

ii) Für $X \sim \mathcal{N}_n(0, I_n)$ gilt:

$$\varphi_X(t) = \mathbb{E}\Big[e^{i\langle t,X\rangle}\Big] \overset{(*)}{=} \mathbb{E}\Big[e^{i\sum_{j\in \underline{n}} t_j X_j}\Big] \overset{X_j \text{ unabh.}}{=} \prod_{j\in\underline{n}} \mathbb{E}\big[e^{it_jX_j}\big] \overset{\text{i)}}{=} \prod_{i\in\underline{n}} e^{-\frac{t_j^2}{2}} = e^{-\frac{\langle t,t\rangle}{2}}.$$

(Gleichheit bei (*): $X = (X_1, \ldots, X_n)$ mit $X_i \sim N(0,1)$)

iii) Allgemein: $Y = AX + b$ mit $X \sim \mathcal{N}_n(0, I_n), b \in \mathbb{R}^m$ und $A \in Mat(m \times n, \mathbb{R})$. Dann gilt:

$$\varphi_X(t) = exp\bigg(i\langle t, b\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\bigg),$$

wobei $\Sigma = AA^T$ (Übung).

Insbesondere hängt die Verteilung von Y nur von b und Σ ab.

Korollar 15.7 Zu $b \in \mathbb{R}$ und $\Sigma \in Mat(n, \mathbb{R})$ symmetrisch und positiv semidefinit existiert genau eine Normalverteilung Y mit $\mathbb{E}[Y] = b$ und $Cov(Y) = \Sigma$.

Beweis:

- Existenz: Finde ein A mit $AA^T = \Sigma$ durch Diagonalisierung (Eigenwerte sind nichtnegativ, also existiert eine Wurzel der Matrix).
- Eindeutigkeit: 15.2 und 15.6 iii).

□

Bemerkung 15.8

i) Sind X und Y unabhängig und reell, so gilt:

$$\varphi_{X+Y}(t) = \mathbb{E}\Big[e^{it(X+Y)}\Big] \overset{\text{Unabh.}}{=} \mathbb{E}\big[e^{itX}\big]\,\mathbb{E}\big[e^{itY}\big] = \varphi_X(t)\varphi_Y(t).$$

ii) Insbesondere: $X \sim N(\mu, \sigma^2)$ und $Y \sim N(\nu, \tau^2)$ unabhängig. Dann:

$$\varphi_X(t)\varphi_Y(t) \overset{\text{15.6 iii)}}{=} exp\bigg(it\mu - \frac{1}{2}\sigma^2 t^2\bigg)\, exp\bigg(it\nu - \frac{1}{2}\tau^2 t^2\bigg) = exp\bigg(it(\mu+\nu) - \frac{1}{2}(\sigma^2 + \tau^2)t^2\bigg).$$

Folglich gilt: $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$.

Lemma 15.9 Gilt für eine Folge reeller W'maße $(\mu_n)_n$:

i) $(\mu_n)_n$ ist straff und

ii) jede schwach konvergente Teilfolge von $(\mu_n)_n$ konvergiert gegen denselben Grenzwert μ,

dann gilt $\mu_n \to \mu$.

Beweis: Gilt $\mu_n \overset{w}{\nrightarrow} \mu$, so existiert ein $\varepsilon > 0, f \in C_b(\mathbb{R})$ und Teilfolge $(\mu_{n_k})_k$ mit

$$\left| \int f \, d\mu_{n_k} - \int g \, d\mu_{n_k} \right| \geq \varepsilon \qquad \text{für alle } k \geq 1.$$

Jedoch existiert nach 14.10 eine schwach konvergente Teilfolge von μ_{n_k} (wegen Straffheit) und diese konvergiert gegen μ. Widerspruch. □

Satz 15.10 (Stetigkeitssatz von Lêvy) Es seien $(\mu_n)_n$ und μ W'maße auf $(\mathbb{R}^d, \mathcal{B}^d)$ mit charakteristischen Funktionen $\varphi_n := \varphi_{\mu_n}$ und $\varphi := \varphi_\mu$. Dann sind äquivalent:

i) $\mu_n \overset{w}{\to} \mu$

ii) $\varphi_n(t) \to \varphi(t)$ für alle $t \in \mathbb{R}^d$.

Beweis:
i) "⇒" ii): $t \mapsto e^{itX}$ ist stetig und beschränkt.

ii) "⇒" i): Zeige: $(\mu_n)_n$ ist straff (*). Dann: Jede schwach konvergente Teilfolge mit $\mu_{n_k} \overset{w}{\to} \nu$ erfüllt $\varphi_{n_k}(t) \to \varphi(t)$ für alle $t \in \mathbb{R}$. Insgesamt folgt: $\mu_n \overset{w}{\to} \mu$ nach 15.9.

Beweis von (*): Zu zeigen: $\forall \varepsilon > 0 \, \exists K$ kompakt mit $\sup_n \mu(K^c) \leq \varepsilon$. Es genügt zu zeigen (Übung!):

$$\forall \varepsilon \, \exists K \text{ kompakt } \exists N \in \mathbb{N} : \sup_{n \geq N} \mu_n(K^c) \leq \varepsilon.$$

Mit Fubini gilt:

$$\frac{1}{u} \int_{-u}^{u} (1 - \varphi_n(t)) \, dt = \int_R \frac{1}{u} \int_{-u}^{u} (1 - e^{itx}) \, dt \, d\mu_n(x) \overset{15.3}{=} 2 \int_{\mathbb{R}} \underbrace{\left(1 - \frac{\sin(ux)}{ux}\right)}_{\geq 0 \text{ wegen } |\sin(y)| \leq |y|} d\mu_n.$$

Also folgt mit $X = \left\{x : |x| > \frac{2}{u}\right\}$:

$$\begin{aligned} \frac{1}{u} \int_{-u}^{u} (1 - \varphi_n(t)) \, dt &\geq 2 \int_X \left(1 - \frac{\sin(ux)}{ux}\right) d\mu_n(x) \geq 2 \int_X \left(1 - \left|\frac{\sin(ux)}{ux}\right|\right) d\mu_n(x) \\ &\geq 2 \int_X \left(1 - \frac{1}{|ux|}\right) d \overset{(**)}{\geq} \mu_n(x) \leq \mu(X) \end{aligned}$$

(Gleichheit bei (**): $1 - \frac{1}{|ux|} \geq \frac{1}{2}$ für $|x| > \frac{2}{u}$). Beachte: $\varphi_X(t) = \int e^{itx} \, d\mu_n$ ist Lipschitz-stetig mit $\varphi(0) = 1$ (Übung!). Also: $\frac{1}{u} \int_{-u}^{u} (1 - \varphi(t)) \, dt \to 0$ für $u \to 0$ (wegen Stetigkeit), dh. zu $\varepsilon > 0$ existiert $u_0 > 0$ mit

$$\frac{1}{u_0} \int_{-u_0}^{u_0} (1 - \varphi(t)) \, dt \leq \varepsilon.$$

Es folgt:

$$\mu_n\left(\left\{x : |x| > \frac{2}{u_0}\right\}\right) \leq \frac{1}{u_0}\int_{-u_0}^{u_0}(1-\varphi_n(t))\,d\mu_n \overset{\text{maj.K.}}{\longrightarrow} \frac{1}{u_0}\int_{-u_0}^{u_0}(1-\varphi(t))\,dt \leq \varepsilon.$$

Es folgt:

$$\exists\, N \in \mathbb{N} \text{ mit } \mu_n\left(\left\{x : |x| > \frac{1}{u_0}\right\}\right) \text{ für alle } n \geq N.$$

□

Korollar 15.11 Sind $(\mu_n)_n$ W'maße auf $(\mathbb{R}^d, \mathcal{B}^d)$ mit charakteristischen Funkionen φ_n und gilt

$$\varphi_n(t) \to \varphi(t) \qquad \text{für alle } t \in \mathbb{R}$$

für eine in 0 stetige Funktion φ, so ist φ die charakteristische Funktion eines W'maßes μ und $\mu_n \overset{w}{\to} \mu$.

Beweis: Wegen $1 = \varphi_n(0) \to \varphi(0)$ gilt $\varphi(0) = 1$. Außerdem: φ ist stetig in 0, folglich ist $(\mu_n)_n$ straff (Argumentation wie im Beweis von 15.10). Also existiert (14.10) eine schwach konvergente Teilfolge $\mu_{n_k} \overset{w}{\to} \nu$. Da ν W'maß, folgt $\varphi_\nu(t) = \varphi(t)$ für alle $t \in \mathbb{R}$. Die Aussage folgt aus Satz 15.10. □

Beispiel 15.12

i) Es seien $X_n \sim N(0, \sigma_n^2)$ und $X_n \overset{\mathcal{D}}{\to} X$. Aus $\varphi_n(t) = e^{-\frac{1}{2}\sigma_n^2 t^2} \to \varphi_X(t)$ für alle $t \in \mathbb{R}$ folgt mit $t = 1$: $\sigma_n^2 \to \sigma^2$ für ein $\sigma^2 \in \mathbb{R}$. Folglich: $e^{-\frac{1}{2}\sigma_n^2 t^2} \to e^{-\frac{1}{2}\sigma^2 t^2}$ und $X \sim N(0, \sigma^2)$.

ii) $X_n \sim \mathcal{N}_n(\mu_n, \sigma_n)$ und $X_n \overset{\mathcal{D}}{\to} X$ analog. (Übung!)

16 Der zentrale Grenzwertsatz

Bemerkung 16.1 Wir wissen aus dem starken Gesetz der großen Zahlen: Ist $(X_i)_i$ i.i.d. Folge mit $\mathbb{E}[X_i] = \mu$ und $Var(X_i) = \sigma^2 \leq \infty$ für alle i, so folgt:

$$\overline{X}_n := \frac{1}{n} \sum_{j=1}^{n} X_j = \frac{S_n}{n} \overset{\text{f.s.}}{\to} \mu.$$

Klar: Ist $\sigma^2 = 0$, so folgt $X_j = 0$ f.s. und somit $\frac{S_n}{n} = \mu$ f.s.
Frage: Wie hoch ist im Falle von $\sigma^2 > 0$ die Konvergenzgeschwindigkeit von $\frac{S_n}{n}$?

Satz 16.2 (zentraler Grenzwertsatz) Es sei $(X_j)_j$ wie in 16.1 und $\sigma^2 > 0$. Dann gilt für

$$Y_n = \sqrt{n} \left(\frac{\overline{X}_n - \mu}{\sigma} \right) = \frac{S_n - \mu}{\sigma \sqrt{n}}$$

die Konvergenz in Verteilung $Y_n \overset{\mathcal{D}}{\to} Y$ mit $Y \sim N(0,1)$.

Beweis: Es sei $\varphi_j = \varphi$ die charakteristische Funktion von $X_j - \mu$. Dann gilt:

$$\varphi_{Y_n}(u) = \varphi_{\frac{1}{\sqrt{n}\sigma} \sum_{j \in \underline{n}} (X_j - \mu)}(u) \overset{\text{Def. char. Fkt.}}{=} \varphi_{\sum_{j \in \underline{n}} (X_j - \mu)} \left(\frac{1}{\sigma \sqrt{n}} u \right) \overset{15.5}{=} \prod_{j \in \underline{n}} \varphi_j \left(\frac{u}{\sigma \sqrt{n}} \right) = \left(\varphi \left(\frac{u}{\sigma \sqrt{n}} \right) \right)^n.$$

Beachte: $\varphi(u) = \mathbb{E}[exp\,(iu\,(X_j - \mu))]$ erfüllt:

$$\begin{aligned} \varphi'(u) &= \mathbb{E}[i(X_j - \mu) exp(iu(X_j - \mu))] \qquad \text{(vgl. 15.6)} \\ \varphi''(u) &= \mathbb{E}[-(X_j - \mu)^2 exp(iu(X_j - \mu))]. \end{aligned}$$

Es folgt: $\varphi'(0) = 0, \varphi''(0) = -\sigma^2$ und φ'' ist stetig. Also folgt mit Satz von Taylor:

$$\varphi(u) = \varphi(0) + \varphi'(0)u + \frac{1}{2}\varphi''(0)u^2 + u^2 h(u) = 1 - \frac{\sigma^2 u^2}{2} + u^2 h(u)$$

mit $\lim_{u \to 0} h(u) = 0$. Insgesamt folgt:

$$\varphi_{Y_n}(u) = \left(\varphi \left(\frac{u}{\sigma \sqrt{n}} \right)^n \right) = exp \left(n \cdot \log \varphi \left(\frac{u}{\sigma \sqrt{n}} \right) \right) = exp \left(n \cdot \log \left(1 - \frac{u^2}{2n} + \frac{u^2}{\sigma^2 n} h \left(\frac{u}{\sigma \sqrt{n}} \right) \right) \right)$$

mit dem Hauptzweig des komplexen Logarithmus'.
Zuletzt stellen wir fest: $\frac{\log(1+X_n)}{X_n} \to 1$ für $X_n \to 0$. Da in unseren Fall $X_n = -\frac{u^2}{2n} + \frac{u^2}{\sigma^2 n} h \left(\frac{u}{\sigma \sqrt{n}} \right)$, folgt:

$$\varphi_{Y_n}(u) = exp \left(n X_n \frac{\log(1 + X_n)}{X_n} \right) \sim e^{n X_n} \to e^{-\frac{u^2}{2}} \qquad \text{für alle } u \in \mathbb{R}.$$

($\sim$ bedeutet hierbei asymptotische Äquivalenz). Die Aussage folgt nun aus 15.10 und 15.6. $\square$

Korollar 16.3 In der Situation von 16.2 gilt

$$\sqrt{n}(\overline{X}_n - \mu) \overset{\mathcal{D}}{\to} N(0, \sigma^2).$$

Bemerkung 16.4 Es ergibt sich $\sqrt{n}$ bzw. $\frac{1}{\sqrt{n}}$ als Konvergenzgeschwindigkeit in starken Gesetz der großen Zahlen.

Satz 16.5 Es sei $(X_n)_n$ unabhängig mit $\mathbb{E}[X_j] = 0$ und $\sigma_j^2 = Var(X_j) < \infty$. Gelten:

i) $\sup_j \mathbb{E}\big[|X_j|^{2+\varepsilon}\big] < \infty$ (für ein $\varepsilon > 0$) (sonst klappt Taylor nicht immer)

ii) $\sum_{j=1}^{\infty} \sigma_j^2 = \infty$ (sonst keine "Mittlungseffekte")

so folgt:

$$\frac{S_n}{\sqrt{\sum_{j \in \underline{n}} \sigma_j^2}} \xrightarrow{\mathcal{D}} N(0,1).$$

Beispiel 16.6

i) Es seien $(X_i)_i$ i.i.d. mit $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$. Dann gilt:

$$\mathbb{E}[X_i] = p, Var(X_i) = p(1-p) \text{ und } S_n \sim Bin(n,p).$$

Wir erhalten:

$$\frac{S_n}{n} \xrightarrow{\text{f.s.}} p \text{ (sGgZ)} \qquad \text{und} \qquad \sqrt{n}\left(\frac{S_n}{n} - p\right) \xrightarrow{\mathcal{D}} N(0, p(1-p)) \text{ (zGWS)}$$

ii) Es seien $X_1, \ldots, X_n$ i.i.d. reelle Zufallsgrößen mit unbekannter Verteilungsfunktion F. Ziel ist die Schätzung von F. Setze dazu:

$$F_n(x) = \frac{1}{n} \sum_{j \in \underline{n}} 1_{\{X_j \le x\}}$$

("empirische Verteilungsfunktion"). Wenn x fest ist, gilt $\mathbb{E}\big[1_{\{X_j \le x\}}\big] = P(X_j \le x) = F(x)$. Es folgt $F_n(x) \xrightarrow{\text{f.s.}} F(x)$ und $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{\mathcal{D}} N(0, F(x)(1 - F(x)))$.