Kapitel 1

Mengensysteme, Wahrscheinlichkeitsmaße

Der Großteil der folgenden fundamentalen Begriffe sind schon aus den Vorlesungen 'Stochastische Modellbildung' und Analysis 3 bekannt:

Definition 1.1 *Eine Familie $\mathcal{A}$ von Teilmengen einer Menge Ω ('Grundgesamtheit') heisst* σ-Algebra, *falls die folgenden drei Bedingungen erfüllt sind:*

1. $\Omega \in \mathcal{A}$.

2. *Für jedes $A \in \mathcal{A}$ ist auch $A^c \in \mathcal{A}$.*

3. *Für $A_1, A_2, A_3, \ldots \in \mathcal{A}$ gilt auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.*

Jedes $A \in \mathcal{A}$ heisst messbar *bzw.* Ereignis. $(\Omega, \mathcal{A})$ *heisst* Messraum.

Lemma 1.2 *Sei $(\Omega, \mathcal{A})$ ein Messraum und $A, B, A_1, A_2, A_3, \ldots \in \mathcal{A}$. Dann sind auch die folgenden Mengen messbar:*

1. $A \setminus B$.

2. $A \Delta B := (A \setminus B) \cup (B \setminus A)$ *(symmetrische Differenz).*

3. $\bigcap_{n=1}^{\infty} A_n$.

4. $\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ *(limes superior der Folge $(A_n)_{n\in\mathbb{N}}$)*

5. $\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$ *(limes inferior der Folge $(A_n)_{n\in\mathbb{N}}$)*

Beweis: Übungsaufgabe

Bemerkung 1.3 Angelehnt an den reellen Fall heisst eine Folge $(A_n)_{n\in\mathbb{N}}$ messbarer Mengen *konvergent* genau dann, wenn $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$.

Satz 1.4 *Für jedes System $\mathcal{E}$ von Teilmengen von Ω existiert eine kleinste σ-Algebra $\mathcal{A}_\sigma(\mathcal{E})$ mit $\mathcal{E} \subseteq \mathcal{A}_\sigma(\mathcal{E})$. Wir nennen $\mathcal{A}_\sigma(\mathcal{E})$ die* von $\mathcal{E}$ erzeugte σ-Algebra.

Beweis: Übungsaufgabe. Es reicht zu zeigen, dass der beliebige Durchschnitt von σ-Algebren wieder eine σ-Algebra ist.

Beispiel 1.5 Die Vereinigung von σ-Algebren muss keine σ-Algebra sein: Betrachten wir $\Omega = [0,1]$ sowie $\mathcal{A}_1 = \{\emptyset, [0,1], [0,3/4], (3/4,1]\}$ und $\mathcal{A}_2 = \{\emptyset, [0,1], [0,1/4], (1/4,1]\}$ dann ist $\mathcal{A}_1 \cup \mathcal{A}_2$ offensichtlich keine σ-Algebra.

Definition 1.6 *Sei $(\Omega, \mathcal{A})$ ein Messraum. $P : \mathcal{A} \to [0,1]$ heisst Wahrscheinlichkeitsmaß genau dann wenn $P(\Omega) = 1$ gilt und P* σ-additiv *ist, i.e. für jede Folge $A_1, A_2, \dots$ von messbaren, paarweise disjunkten Mengen gilt*

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n). \tag{1.1}$$

Das Tripel $(\Omega, \mathcal{A}, P)$ heisst Wahrscheinlichkeitsraum.

Lemma 1.7 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, dann gilt:*

(a) $P(\emptyset) = 0$.

(b) Für $A, B \in \mathcal{A}$ mit $A \subseteq B$ gilt $P(A) \le P(B)$.

(c) Für jede wachsende Folge $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ messbarer Mengen gilt $\lim_{n\to\infty} P(A_n) = P(\bigcup_{n=1}^{\infty} A_n)$ *(Stetigkeit von unten).*

(d) Für jede fallende Folge $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ messbarer Mengen gilt $\lim_{n\to\infty} P(A_n) = P(\bigcap_{n=1}^{\infty} A_n)$ *(Stetigkeit von oben).*

Die Eigenschaften (b)-(d) besagen, dass jedes Wahrscheinlichkeitsmaß monoton und stetig von unten und oben ist.

Beweis: Aussage (a) folgt direkt aus Gleichung (1.1).
(b): Wegen $B = A \cup (B \setminus A)$ und $A \cap (B \setminus A) = \emptyset$ folgt aus (1.1) unmittelbar (einfach $A_n := \emptyset$ für alle $n \ge 3$ setzen) $P(B) = P(A) + P(B \setminus A)$, also $P(B) - P(A) = P(B \setminus A) \ge 0$.
Zum Beweis von (c) kann wie folgt vorgegangen werden: Wir setzen $B_1 := A_1$ und, für jedes $n \ge 2$, $B_n := A_n \cap A_{n-1}^c$. Dann ist $B_n \in \mathcal{A}$ für jedes $n \in \mathbb{N}$, $B_i \cap B_j = \emptyset$ für $i \ne j$, und es gilt $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Anwendung der σ-Additivität liefert daher

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

Beweis von (d): Übungsaufgabe ■

Bemerkung 1.8 A-priori ist unklar, warum man anstatt mit Wahrscheinlichkeitsmaßen auf σ-Algebren (oder schwächeren Strukturen) nicht gleich mit Wahrscheinlichkeitsmaßen auf Potenzmengen $\mathfrak{p}(\Omega)$ arbeitet. Die zwei Hauptgründe dafür sind die Folgenden:
Grund 1: Es ist unmöglich, (Wahrscheinlichkeits-) Maße mit natürlichen Eigenschaften auf ganz $\mathfrak{p}(\Omega)$ zu definieren. Angenommen, wir suchen eine Funktion $\mathcal{V} : \mathfrak{p}(\mathbb{R}^3) \to [0,\infty]$, die das 'Volumen' von Mengen beschreibt und daher mindestens folgende Bedingungen erfüllen sollte:

(V1) $\mathcal{V}(A \cup B) = \mathcal{V}(A) + \mathcal{V}(B)$ für alle $A, B \subseteq \mathbb{R}^3$ mit $A \cap B = \emptyset$

(V2) $\mathcal{V}([0,1]^3) = 1$

(V3) Für jede Bewegung $f : \mathbb{R}^3 \to \mathbb{R}^3$ und jedes $A \subseteq \mathbb{R}^3$ gilt $\mathcal{V}(f(A)) = \mathcal{V}(A)$.

Hausdorff zeigte 1905, dass es keine solche Funktion $\mathcal{V}$ gibt. Banach & Tarski bewiesen 1924 ein noch wesentlich stärkeres Resultat, das sog. 'Banach-Tarksi-Paradoxon'† (interessantes Thema für eine Bachelor Arbeit):

Satz 1.9 (Banach-Tarski Paradoxon, AC) *Angenommen $K, G \subseteq \mathbb{R}^3$ sind beschränkt und haben nichtleeres Inneres. Dann existieren paarweise disjunkte Mengen $K_1, \ldots, K_n$ mit $K = \bigcup_{i=1}^n K_i$ und Bewegungen $f_1, \ldots, f_n$ sodass $f_1(K_1), \ldots, f_n(K_n)$ paarweise disjunkt sind und $G = \bigcup_{i=1}^n f_i(K_i)$ gilt.*
Mit anderen Worten: Aus einem Tennisball K kann durch geschickte Zerlegung und Bewegung ein Fussball G gebastelt werden.

Grund 2: Wahrscheinlichkeitsmaße können eindeutig von 'guten' kleinen Klassen (Halbringe etc.) auf die erzeugten σ-Algebren fortgesetzt werden (vergleiche Satz 1.11 sowie [2, 4, 5]). Als Beispiel betrachten wir $\Omega = [0,1]$ und die Klasse $\mathfrak{h}$, definiert durch

$$\mathfrak{h} = \big\{(a,b] \subseteq [0,1] : \, a \leq b\big\} \tag{1.2}$$

Offensichtlich ist $\mathfrak{h}$ keine σ-Algebra, zum Beispiel ist weder $(0,1/4] \cup (1/2,1]$ noch $[0,1]$ ein Element von $\mathfrak{h}$. $\mathfrak{h}$ hat aber die folgenden Eigenschaften: (i) $\emptyset \in \mathfrak{h}$, (ii) $A, B \in \mathfrak{h}$ impliziert $A \cap B \in \mathfrak{h}$, (iii) Gelte $A, B \in \mathfrak{h}$ und $A \subseteq B$, dann existieren disjunkte Mengen $E_1, E_2 \in \mathfrak{h}$ mit $B \setminus A = E_1 \cup E_2$. $\mathfrak{h}$ ist ein sogenannter Halbring:

Definition 1.10 *Eine Familie $\mathfrak{h}$ von Teilmengen einer Menge Ω heisst* Halbring, *falls die folgenden drei Bedingungen erfüllt sind:*

1. $\emptyset \in \mathfrak{h}$.
2. *Für $A, B \in \mathfrak{h}$ gilt auch $A \cap B \in \mathfrak{h}$.*
3. *Für $A, B \in \mathfrak{h}$ mit $A \subseteq B$ gilt: Es existieren paarweise disjunkte Mengen $E_1, \ldots, E_n \in \mathfrak{h}$ sodass $B \setminus A = \bigcup_{i=1}^n E_i$.*

Es ist klar, wie Elementen von $\mathfrak{h}$ gemäß (1.2) eine Länge zugewiesen werden kann: Wir definieren einfach

$$\lambda((a,b]) := b - a \tag{1.3}$$

Wie kann dieses λ auf kompliziertere Mengen $A \in \mathcal{A}_\sigma(\mathfrak{h})$ erweitert werden? Naheliegende Idee: Überdecke A möglichst gut durch halboffene Intervalle, i.e. für $A \subseteq [0,1]$ definiere

$$\lambda(A) := \inf\left\{\sum_{i=1}^\infty \lambda(I_i) : I_1, I_2, \ldots \in \mathfrak{h}, \, A \subseteq \bigcup_{i=1}^\infty I_i\right\} \tag{1.4}$$

Dann lässt sich zeigen (siehe Vorlesung Analysis 3/Maßtheorie), dass damit ein Wahrscheinlichkeitsmaß λ auf $\mathcal{A}_\sigma(\mathfrak{h})$ definiert ist. Im Folgenden werden wir λ als *Lebesgue'sches Maß* bezeichnen. Allgemeiner gilt folgender Satz:

† für mehr Hintergrundwissen sei auf die Vorlesung Analysis 3/Maßtheorie verwiesen

Satz 1.11 *Sei $(\Omega, \mathcal{A})$ ein Messraum und $\mathfrak{h}$ ein Halbring mit der Eigenschaft $\mathcal{A}_\sigma(\mathfrak{h}) = \mathcal{A}$. Weiters sei $\mu : \mathfrak{h} \to [0,1]$ σ-additiv[†]. Dann existiert ein eindeutiges Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{A})$, für das $P(E) = \mu(E)$ für alle $E \in \mathfrak{h}$ gilt.*

Beweis: Übersteigt leider den Rahmen dieser Vorlesung, siehe Vorlesung Analysis 3/Maßtheorie sowie [2, 4, 5, 7]. Für Interessierte sei hier nur angemerkt, dass der Beweis der Fortsetzbarkeit auf der ganz elementaren, zuvor schon erwähnten Idee beruht: Approximiere (überdecke) die Menge A, von der die Fläche (das Volumen) berechnet werden soll, durch einfache Objekte (Rechtecke), deren Fläche (Volumen) bekannt ist.

Beispiel 1.12 Die folgenden Eigenschaften des Lebesgue-Maßes sind leicht zu zeigen (Übungsaufgabe) - $\mathfrak{h}$ bezeichnet dabei wieder den Halbring der halboffenen Intervalle gemäß (1.2):

(S1) Für jedes $[a, b] \subseteq (0, 1]$ gilt $[a, b] \in \mathcal{A}_\sigma(\mathfrak{h})$

(S2) Für $[a, b] \subseteq (0, 1]$ gilt $\lambda([a, b]) = b - a$

(S3) $\lambda(\mathbb{Q} \cap (0, 1]) = 0$

Die im Folgenden für uns wichtigste σ-Algebra ist die sogenannte Borel'sche σ-Algebra:

Definition 1.13 *Gegeben sei ein metrischer Raum (Ω, d). Dann heißt die von der Familie $\mathcal{O}$ aller offenen Mengen erzeugte σ-Algebra die* σ-Algebra der Borelmengen bzw. Borel'sche σ-Algebra. *Wir bezeichnen sie im Folgenden mit $\mathcal{B}(\Omega)$.*

Lemma 1.14 Wir betrachten die folgenden Mengensysteme in $\mathbb{R}$:

$$\begin{aligned}
\mathcal{E}_1 &= \big\{(a,b) \subseteq \mathbb{R} : a \le b\big\}, & \mathcal{E}_1' &= \big\{(a,b) \subseteq \mathbb{R} : a \le b \text{ und } a,b \in \mathbb{Q}\big\} \\
\mathcal{E}_2 &= \big\{(a,b] \subseteq \mathbb{R} : a \le b\big\}, & \mathcal{E}_2' &= \big\{(a,b] \subseteq \mathbb{R} : a \le b \text{ und } a,b \in \mathbb{Q}\big\} \\
\mathcal{E}_3 &= \big\{[a,b] \subseteq \mathbb{R} : a \le b\big\}, & \mathcal{E}_3' &= \big\{[a,b] \subseteq \mathbb{R} : a \le b \text{ und } a,b \in \mathbb{Q}\big\} \\
\mathcal{E}_4 &= \big\{(-\infty,b] \subseteq \mathbb{R} : b \in \mathbb{R}\big\}, & \mathcal{E}_4' &= \big\{(-\infty,b] \subseteq \mathbb{R} : b \in \mathbb{Q}\big\} \\
\mathcal{E}_5 &= \big\{(b,\infty) \subseteq \mathbb{R} : b \in \mathbb{R}\big\}, & \mathcal{E}_5' &= \big\{(b,\infty) \subseteq \mathbb{R} : b \in \mathbb{Q}\big\} \\
\mathcal{E}_6 &= \big\{K \subseteq \mathbb{R} : K \text{ kompakt}\big\} \\
\mathcal{E}_7 &= \big\{A \subseteq \mathbb{R} : A \text{ abgeschlossen}\big\}
\end{aligned} \tag{1.5}$$

Dann sind alle 12 Familien Erzeuger der Borel'sche σ-Algebra $\mathcal{B}(\mathbb{R})$ in $\mathbb{R}$.

Beweis: Wir zeigen nur, dass $\mathcal{E}_1$ und $\mathcal{E}_2$ die Borel'sche σ-Algebra $\mathcal{B}(\mathbb{R})$ erzeugen.
(i) Da jedes Element von $\mathcal{E}_1$ offen ist gilt $\mathcal{E}_1 \subseteq \mathcal{O}$ und damit $\mathcal{A}_\sigma(\mathcal{E}_1) \subseteq \mathcal{A}_\sigma(\mathcal{O}) = \mathcal{B}(\mathbb{R})$. Umgekehrt kann jede offene Menge $U \subseteq \mathbb{R}$ als höchstens abzählbare Vereinigung $\bigcup_{i=1}^\infty I_i$ von offenen Intervallen dargestellt werden. Wegen $\bigcup_{i=1}^\infty I_i \in \mathcal{A}_\sigma(\mathcal{E}_1)$ folgt damit $U \in \mathcal{A}_\sigma(\mathcal{E}_1)$, also $\mathcal{A}_\sigma(\mathcal{O}) \subseteq \mathcal{A}_\sigma(\mathcal{E}_1)$, und damit insgesamt $\mathcal{A}_\sigma(\mathcal{E}_1) = \mathcal{B}(\mathbb{R})$.
(ii) Wegen $(a, b] = \bigcap_{n=1}^\infty (a, b + 1/n)$ gilt $\mathcal{E}_2 \subseteq \mathcal{B}(\mathbb{R})$ und damit $\mathcal{A}_\sigma(\mathcal{E}_2) \subseteq \mathcal{B}(\mathbb{R})$. Da weiters $(a, b) = \bigcup_{n=1}^\infty (a, b - 1/n]$ gilt, folgt $\mathcal{E}_1 \subseteq \mathcal{A}_\sigma(\mathcal{E}_2)$, also auch $\mathcal{B}(\mathbb{R}) = \mathcal{A}_\sigma(\mathcal{E}_1) \subseteq \mathcal{A}_\sigma(\mathcal{E}_2)$. ■

[†] i.e. für paarweise disjunkte Mengen $A_1, A_2, \ldots$ in $\mathfrak{h}$ mit $\bigcup_{n=1}^\infty A_n \in \mathfrak{h}$ gilt Gleichung (1.1)

Bemerkung 1.15 Entsprechende Aussagen gelten auch im allgemeinen Fall von $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, siehe [2, 4, 5]. So ist zum Beispiel die Klasse $\mathcal{I}_d$ aller halboffenen d-dimensionalen Rechtecke

$$\mathcal{I}_d := \Big\{ \bigtimes_{i=1}^{d} (a_i, b_i] : \forall i \in \{1, \dots, d\} : a_i \leq b_i \Big\} \tag{1.6}$$

ein Halbring, der $\mathcal{B}(\mathbb{R}^d)$ erzeugt, i.e. es gilt $\mathcal{A}_\sigma(\mathcal{I}_d) = \mathcal{B}(\mathbb{R}^d)$.

Folgerung 1.16 *Es existiert genau ein Wahrscheinlichkeitsmaß λ_d auf $([0,1]^d, \mathcal{B}([0,1]^d))$, das*

$$\lambda_d\Big(\bigtimes_{i=1}^{d} (a_i, b_i] \Big) = \prod_{i=1}^{d} (b_i - a_i)$$

*für alle $a_i, b_i \in [0,1]$ und $a_i \leq b_i$ erfüllt. Dieses Maß geht auf Henri Lebesgue zurück und trägt den Namen d-*dimensionales *Lebesgue Maß.*

Beweis: Die Familie $\mathcal{I}_d$ aller d-dimensionalen Rechtecke der Form $\bigtimes_{i=1}^{d} (a_i, b_i]$ ist ein Halbring und λ_d ist σ-additiv auf $\mathcal{I}_d$ (beides selbst verifizieren !). Damit ist Satz 1.11 anwendbar. ■

Bemerkung 1.17 Die Definition von λ_d kann auch unschwer auf ganz $\mathcal{B}(\mathbb{R}^d)$ erweitert werden (siehe Analysis 3). Das entsprechenden Maß (für das $\lambda_d(\mathbb{R}^d) = \infty$ gilt) werden wir ebenfalls mit λ_d bezeichnen.

Kapitel 2

Zufallsvariable

In der Vorlesung 'Stochastische Modellbildung' haben Sie zu einem großen Teil diskrete Zufallsvariable X, i.e. Abbildungen $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ mit abzählbarem Wertebereich $W_X = \{\alpha_1, \alpha_2, \ldots\}$, betrachtet, die Wahrscheinlichkeiten $P(X = \alpha_i)$ berechnet und nachfolgend verwendet. Dabei wurde immer implizit angenommen, dass $P(X = \alpha_i)$ wohldefiniert ist, i.e. dass $\{\omega \in \Omega : X(\omega) = \alpha_i\}$ ein Element der σ-Algebra ist, auf der P definiert ist. Mit anderen Worten: Grundannahme war

$$X^{-1}(\{\alpha_i\}) := \{\omega \in \Omega : X(\omega) = \alpha_i\} \in \mathcal{A} \quad \text{für jedes } \alpha_i \in W_X. \tag{2.1}$$

Falls Bedingung (2.1) erfüllt ist, gilt für jede Menge $E \subseteq W_X$ auch $X^{-1}(E) := \{\omega \in \Omega : X(\omega) \in E\} \in \mathcal{A}$ weil

$$\begin{aligned} X^{-1}(E) &= \{\omega \in \Omega : X(\omega) \in E\} = \bigcup_{e \in E} \{\omega \in \Omega : X(\omega) = e\} \\ &= \bigcup_{e \in E} \underbrace{X^{-1}(\{e\})}_{\in \mathcal{A}} \in \mathcal{A}. \end{aligned}$$

Sei $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ nun eine beliebige Abbildung, die den Ausgang eines Zufallsexperiments beschreibt, beispielsweise die Lebensdauer eines elektronischen Bauteiles. Um die Güte des Bauteils bewerten zu können, interessiert man sich insbesondere für die Wahrscheinlichkeit dafür, dass der Bauteil höchstens bis zum Zeitpunkt T funktioniert, i.e. für

$$P(X^{-1}(-\infty, T])) = P(X \le T).$$

Mengen der Form

$$X^{-1}((-\infty, T]) = \{\omega \in \Omega : X(\omega) \le T\}$$

müssen also $X^{-1}((-\infty, T]) \in \mathcal{A}$ erfüllen.

Frage 2.1 Für welche weiteren Menge $E \subseteq \mathbb{R}$ sollte $P(X^{-1}(E))$ für das Beispiel der Lebensdauer X bekannt sein? Wen interessiert $P(X^{-1}(E))$ für eine sehr komplizierte Menge E, zum Beispiel die Cantor Menge?

Das folgende Resultat zeigt, dass aus der Eigenschaft $X^{-1}((-\infty, T]) \in \mathcal{A}$ für jedes $T \in \mathbb{R}$ schon $X^{-1}(E) \in \mathcal{A}$ für jede Borel-Menge $E \in \mathcal{B}(\mathbb{R})$ folgt - allgemeiner gilt folgendes wichtiges Resultat:

Satz 2.2 *Gegeben sei eine Abbildung* $f:(\Omega_1,\mathcal{A}_1)\to(\Omega_2,\mathcal{A}_2)$ *sowie ein Erzeuger* $\mathcal{E}_2$ *von* $\mathcal{A}_2$*. Gilt* $f^{-1}(E_2)\in\mathcal{A}_1$ *für jedes* $E_2\in\mathcal{E}_2$ *dann folgt* $f^{-1}(A_2)\in\mathcal{A}_1$ *für jedes* $A_2\in\mathcal{A}_2$*.*

Beweis: Wir beweisen das Resultat in drei Schritten:
(i) Leicht nachzuprüfen (Übungsaufgabe): Für jedes beliebige $f:(\Omega_1,\mathcal{A}_1)\to(\Omega_2,\mathcal{A}_2)$ sind die Mengensysteme

$$f^{-1}(\mathcal{A}_2):=\big\{f^{-1}(B):\,B\in\mathcal{A}_2\big\},\quad f(\mathcal{A}_1):=\big\{B\subseteq\Omega_2:\,f^{-1}(B)\in\mathcal{A}_1\big\} \tag{2.2}$$

wieder σ-Algebren.
(ii) Wir zeigen, dass für jedes Mengensystem $\mathcal{S}_2\subseteq\mathfrak{p}(\Omega_2)$ und jedes $f:(\Omega_1,\mathcal{A}_1)\to(\Omega_2,\mathcal{A}_2)$ die folgende Gleichheit gilt:

$$\mathcal{A}_\sigma\big(f^{-1}(\mathcal{S}_2)\big)=f^{-1}\big(\mathcal{A}_\sigma(\mathcal{S}_2)\big) \tag{2.3}$$

Einerseits gilt offensichtlich $f^{-1}(\mathcal{S}_2)\subseteq f^{-1}(\mathcal{A}_\sigma(\mathcal{S}_2))$, also $\mathcal{A}_\sigma\big(f^{-1}(\mathcal{S}_2)\big)\subseteq\mathcal{A}_\sigma\big(f^{-1}(\mathcal{A}_\sigma(\mathcal{S}_2))\big)=f^{-1}(\mathcal{A}_\sigma(\mathcal{S}_2))$. Für den Beweis der anderen Inklusion betrachten wir das Mengensystem $\mathcal{C}:=f\Big(\mathcal{A}_\sigma\big(f^{-1}(\mathcal{S}_2)\big)\Big)$. Nach (i) ist $\mathcal{C}$ eine σ-Algebra und für jedes $B\in\mathcal{S}_2$ gilt offensichtlich $f^{-1}(B)\in f^{-1}(\mathcal{S}_2)\subseteq\mathcal{A}_\sigma\big(f^{-1}(\mathcal{S}_2)\big)$, also $B\in\mathcal{C}$ und damit $\mathcal{A}_\sigma(\mathcal{S}_2)\subseteq\mathcal{C}$. Anwendung von f^{-1} liefert schliesslich $f^{-1}\big(\mathcal{A}_\sigma(\mathcal{S}_2)\big)\subseteq f^{-1}(\mathcal{C})\subseteq\mathcal{A}_\sigma\big(f^{-1}(\mathcal{S}_2)\big)$.
(iii) Satz 2.2 kann nun wie folgt bewiesen werden: Aus $f^{-1}(E_2)\in\mathcal{A}_1$ für jedes $E_2\in\mathcal{E}_2$ folgt unter Verwendung von (ii) $f^{-1}(\mathcal{A}_2)=f^{-1}(\mathcal{A}_\sigma(\mathcal{E}_2))=\mathcal{A}_\sigma(f^{-1}(\mathcal{E}_2))\subseteq\mathcal{A}_\sigma(\mathcal{A}_1)=\mathcal{A}_1$, also die gewünschte Eigenschaft. ■

Definition 2.3 *Seien* $(\Omega_1,\mathcal{A}_1)$ *und* $(\Omega_2,\mathcal{A}_2)$ *Messräume. Eine Abbildung* $f:\Omega_1\to\Omega_2$ *heisst* $\mathcal{A}_1-\mathcal{A}_2$ messbar *genau dann wenn für jedes* $A_2\in\mathcal{A}_2$ *das vollständige Urbild*

$$f^{-1}(A_2):=\{\omega\in\Omega_1:f(\omega)\in A_2\}\in\mathcal{A}_1 \tag{2.4}$$

erfüllt. Eine Abbildung $X:(\Omega,\mathcal{A},P)\to(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ *heisst* d-dimensionaler Zufallsvektor *falls sie* $\mathcal{A}-\mathcal{B}(\mathbb{R}^d)$ *messbar ist; im Fall* $d=1$ *nennen wir ein solches* X *Zufallsvariable. Eine Zufallsvariable heisst* diskret *wenn sie höchstens abzählbar viele Werte annimmt, sie heisst* Indikatorvariable *wenn es eine Menge* $A\in\mathcal{A}$ *gibt, sodass*

$$X(\omega)=\mathbf{1}_A(\omega):=\begin{cases}1 & \text{wenn } \omega\in A\\ 0 & \text{wenn } \omega\notin A\end{cases} \tag{2.5}$$

für alle $\omega\in\Omega$ *gilt.*

Beispiel 2.4 Wir betrachten $\Omega=\{1,2,3\}$ und $\mathcal{A}=\{\emptyset,\Omega,\{1,2\},\{3\}\}$, sowie die folgenden zwei Abbildungen: $X,Y:\Omega\to\mathbb{R}$, definiert durch $X(1)=X(2)=0,X(3)=1$ und $Y(1)=1,Y(2)=2,Y(3)=3$. Dann ist X eine Zufallsvariable, Y jedoch nicht (warum ?).

Beispiel 2.5 Auch aus der Analysis bekannte 'komplizierte', nicht-Riemann-integrierbare Funktionen können Zufallsvariable sein. Betrachten wir zum Beispiel $\Omega=[0,1]$, $\mathcal{A}=\mathcal{B}([0,1])$, $P=\lambda$ und setzen $\mathbb{Q}_0=\mathbb{Q}\cap[0,1]$ sowie $X=\mathbf{1}_{\mathbb{Q}_0^c}$. Dann ist X sogar eine Indikatorvariable und wir erhalten für $\mathbb{E}(X)$

$$\mathbb{E}(X)=\,0\cdot P(X=0)+1\cdot P(X=1)=P(X=1)=\lambda(\mathbb{Q}_0^c)=1-\lambda(\mathbb{Q}_0)=1.$$

Wäre 1 nicht auch der intuitiv richtige Wert für $\int_{[0,1]}\mathbf{1}_{\mathbb{Q}_0^c}(t)\,dt$?

Nachdem es gemäß Satz 2.2 reicht, Erzeuger zu betrachten, erhalten wir sofort folgendes Resultat:

Folgerung 2.6 *Eine Abbildung $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ ist eine Zufallsvariable genau dann wenn für eines der in Lemma 1.14 erwähnten Mengensysteme $\mathcal{E}$ gilt, dass $X^{-1}(E) \in \mathcal{A}$ für jedes $E \in \mathcal{E}$.*

Bemerkung 2.7 Beachten Sie die Ähnlichkeit zwischen dem Begriff der stetigen Abbildung (Urbild jeder offenen Menge ist offen) und dem Begriff der messbaren Abbildung (Urbild jeder messbaren Menge ist messbar).

Lemma 2.8 *Jede stetige Abbildung $f : \mathbb{R}^d \to \mathbb{R}^\rho$ ist Borel-messbar.*

Beweis: Stetigkeit von f impliziert, dass das Urbild jeder offenen Menge offen ist, es gilt also $f^{-1}(V) \in \mathcal{B}(\mathbb{R}^d)$ für jedes offene $V \subseteq \mathbb{R}^\rho$. Anwendung von Satz 2.2 liefert sofort das gewünschte Resultat. ■

Das folgende einfache Resultat werden wir im Folgenden häufig verwenden:

Satz 2.9 *Seien $X_1, X_2, \ldots X_d$ Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $f : \mathbb{R}^d \to \mathbb{R}^\rho$ $\mathcal{B}(\mathbb{R}^d)-\mathcal{B}(\mathbb{R}^\rho)$-messbar (kurz: Borel-messbar). Dann ist sowohl $X = (X_1, X_2, \ldots, X_d)$ als auch $f \circ X$ ein Zufallsvektor.*

Beweis: Gemäß Bemerkung 1.15 ist $\mathcal{I}_d$ ein Erzeuger von $\mathcal{B}(\mathbb{R}^d)$. Für $\times_{i=1}^d (a_i, b_i]$ gilt aber offensichtlich

$$X^{-1}\Big(\bigtimes_{i=1}^{d}(a_i, b_i]\Big) = \bigcap_{i=1}^{d} X_i^{-1}((a_i, b_i]) \in \mathcal{A}.$$

Anwendung von Satz 2.2 liefert also sofort die Messbarkeit von X. Um die zweite Aussage zu beweisen, reicht es offensichtlich, zu zeigen, dass die Verkettung zweier messbarer Funktionen $g : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ und $f : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$ wieder messbar ist, was leicht zu zeigen ist: Für $C \in \mathcal{A}_3$ gilt

$$(f \circ g)^{-1}(C) = \big\{\omega \in \Omega_1 : f(g(\omega)) \in C\big\} = \big\{\omega \in \Omega_1 : g(\omega) \in f^{-1}(C)\big\} = g^{-1}\big(\underbrace{f^{-1}(C)}_{\in \mathcal{A}_2}\big),$$

also $(f \circ g)^{-1}(C) \in g^{-1}(\mathcal{A}_2) \subseteq \mathcal{A}_1$. ■

Nachdem, wie gesagt, jede stetige Abbildung $f : \mathbb{R}^d \to \mathbb{R}$ Borel-messbar ist erhalten wir insbesondere:

Beispiel 2.10 *Seien $X_1, X_2, \ldots X_d$ Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $a_1, a_2, \ldots, a_d \in \mathbb{R}$. Dann sind die folgenden Funktionen ebenfalls Zufallsvariable auf $(\Omega, \mathcal{A}, P)$:*

- $\sum_{i=1}^d a_i X_i$
- $\min(X_1, \ldots, X_d)$, $\max(X_1, \ldots, X_d)$
- $\prod_{i=1}^d X_i$

- $\sum_{i=1}^{d}(X_i - a_i)^2$

Frage 2.11 Gibt es überhaupt Abbildungen $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$, die keine Zufallsvariable sind?

Satz 2.12 *Jeder d-dimensionale Zufallsvektor* $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *induziert durch*

$$P^X(B) := P(X^{-1}(B)) = P\big(\{\omega \in \Omega : X(\omega) \in B\}\big) \quad \text{für } B \in \mathcal{B}(\mathbb{R}^d) \tag{2.6}$$

ein Wahrscheinlichkeitsmaß P^X *auf* $\mathcal{B}(\mathbb{R}^d)$)*. Wir nennen* P^X *im Folgenden die* Verteilung *von* X *und schreiben im Falle von* $P^X = \mu$ *auch* $X \sim \mu$*.*

Beweis: Übungsaufgabe

Definition 2.13 *Für jeden d-dimensionale Zufallsvektor* $X = (X_1, \ldots, X_d) : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *heisst die Funktion* $F_X : \mathbb{R}^d \to [0, 1]$*, definiert durch*

$$F_X(x_1, \ldots, x_d) := P\Big(\bigcap_{i=1}^{d} X_i^{-1}((-\infty, x_i]\Big) \tag{2.7}$$

Verteilungsfunktion von X.

Kapitel 3

Absolute stetige Verteilungen

In der Vorlesung Stochastische Modellbildung haben Sie neben zahlreichen diskreten Verteilungen auch schon einige absolut stetige Verteilungen kennengelernt. Der folgende Abschnitt listet einige absolut stetige Verteilungen auf und formuliert relevante Eigenschaften. Wir starten mit der stetigen Gleichverteilung $X \sim \mathcal{U}(a,b)$: Setzen wir

$$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$$

dann gilt $f : \mathbb{R} \to [0,\infty)$, $\int_{\mathbb{R}} f(x)dx = 1$, und, für $[c,d] \subseteq [a,b]$

$$P^X([c,d]) = \int_{[c,d]} \frac{1}{b-a}dx = \int_{[c,d]} f(x)dx.$$

Definition 3.1 *Eine integrierbare*[†] *Funktion* $f : \mathbb{R}^m \to [0,\infty)$ *mit*

$$\int_{\mathbb{R}^m} f(x_1, x_2, \ldots, x_m)dx_1 dx_2, \ldots dx_m = 1 \tag{3.1}$$

heisst Wahrscheinlichkeitsdichte.

Jede Wahrscheinlichkeitsdichte induziert ein Wahrscheinlichkeitsmaß - es gilt das folgende Lemma:

Lemma 3.2 *Sei* $f : \mathbb{R}^m \to [0,\infty)$ *eine Wahrscheinlichkeitsdichte, dann ist durch*

$$\mu_f\Big(\bigtimes_{i=1}^{m}(a_i, b_i]\Big) = \int_{(a_m,b_m]}\int_{(a_{m-1},b_{m-1}]}\cdots\int_{(a_1,b_1]} f(x_1, x_2, \ldots, x_m)dx_1 dx_2 \ldots dx_m \tag{3.2}$$

ein eindeutiges Wahrscheinlichkeitsmaß μ_f *auf* $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ *bestimmt.*

Beweis: Für paarweise disjunkte Rechtecke $\big(\bigtimes_{i=1}^m (a_i^n, b_i^n]\big)_{n\in\mathbb{N}}$ mit $\bigcup_{n=1}^\infty \bigtimes_{i=1}^m (a_i^n, b_i^n] \in \mathcal{I}_d$ rechnet man (mehr oder weniger) unschwer nach, dass

$$\mu_f\Big(\bigcup_{n=1}^{\infty}\bigtimes_{i=1}^{m}(a_i^n, b_i^n]\Big) = \sum_{n=1}^{\infty}\mu_f\Big(\bigtimes_{i=1}^{m}(a_i^n, b_i^n]\Big) \tag{3.3}$$

[†]integrierbar steht hier für Lebesgue integrierbar (siehe Maßtheorie/Analysis 3) und nicht für Riemann integrierbar; es reicht aber im Folgenden, an das schon aus dem Studium bekannte Riemann Integral denken; auch die erwähnten wichtigen Verteilungen haben alle Riemann integrierbare Dichten.

gilt. Unter Verwendung von Bemerkung 1.15 und Satz 1.11 folgt sofort das gewünschte Resultat.

Definition 3.3 *Ein m-dimensionaler Zufallsvektor X auf $(\Omega, \mathcal{A}, P)$ (bzw. seine Verteilung P^X) heisst* absolut stetig *wenn eine Wahrscheinlichkeitsdichte $f : \mathbb{R}^m \to [0, \infty)$ existiert, sodass $P^X = \mu_f$ gilt. In diesem Fall schreiben wir kurz $X \sim \mu_f$.*

Bemerkung 3.4 (a) Für einen m-dimensionalen Zufallsvektor X mit $X \sim \mu_f$ und $B \in \mathcal{B}(\mathbb{R}^m)$ gilt dann allgemein

$$P^X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}) = \int_B f(x_1, x_2, \ldots, x_m) dx_1 \cdots dx_m. \quad (3.4)$$

Für eine Zufallsvariable X mit $X \sim \mu_f$ ergibt sich insbesondere für die Verteilungsfunktion F_X:

$$F_X(x) = P^X((-\infty, x]) = \int_{(-\infty,x]} f(t)\, dt. \quad (3.5)$$

Den Begriff der absolut stetigen Funktion kennen Sie auch schon aus der Analysis. Insbesondere ist für $X \sim \mu_f$ also F *absolut stetig* im analytischen Sinne[†] und die Begriffsbildung ist konsistent. Wir werden in den folgenden Abschnitten noch näher auf Verteilungsfunktionen (und deren Zusammenhang mit Wahrscheinlichkeitsmaßen) eingehen, begnügen uns vorerst aber mit Gleichung (3.5) und bemerken nur, dass für 'gute' Punkte die Gleichheit $F'(x) = f(x)$ gilt.

Beispiel 3.5 Wir betrachten die Funktion

$$f_\theta(x_1, x_2) = (1 + \theta(1 - 2x_1)(1 - 2x_2))\mathbf{1}_{[0,1]^2}(x_1, x_2)$$

für $\theta \in [-1, 1]$. Offensichtlich gilt $f_\theta(x_1, x_2) \geq 0$ sowie $\int_{\mathbb{R}^2} f_\theta(x_1, x_2) dxdy = 1$, f_θ ist also eine Wahrscheinlichkeitsdichte für jedes $\theta \in [-1, 1]$. Im Folgenden sei $X = (X_1, X_2) \sim \mu_{f_\theta}$.
(i) Wir berechnen die Wahrscheinlichkeit des Ereignisses $\{X_1 \leq \frac{1}{2}\} = \{\omega \in \Omega : X_1(\omega) \leq \frac{1}{2}\}$:

$$\begin{aligned}
P(X_1 \leq \tfrac{1}{2}) &= P\big(\{\omega \in \Omega : X_1(\omega) \leq \tfrac{1}{2}\}\big) = P\big(\{\omega \in \Omega : X_1(\omega) \leq \tfrac{1}{2}, X_2(\omega) \in \mathbb{R}\}\big) \\
&= P\big(\{\omega \in \Omega : X(\omega) \in \underbrace{(-\infty, \tfrac{1}{2}] \times \mathbb{R}}_{=:B}\}\big) \\
&\overset{(3.4)}{=} \int_B f_\theta(x_1, x_2) dx_1 dx_2 = \int_{(-\infty,\frac{1}{2}]} \int_{\mathbb{R}} f_\theta(x_1, x_2) dx_2 dx_1 \\
&= \int_{[0,\frac{1}{2}]} \int_{[0,1]} 1 + \theta(1 - 2x_1)(1 - 2x_2)\, dx_2 dx_1 = \int_{[0,\frac{1}{2}]} 1 dx_1 = \frac{1}{2}
\end{aligned}$$

(ii) Wir berechnen die Wahrscheinlichkeit des Ereignisses $\{X_2 \leq X_1^2\}$ und setzen dafür $B := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq x_1^2\}$:

$$\begin{aligned}
P(X_2 \leq X_1^2) &= P\big(\{\omega \in \Omega : X_2(\omega) \leq X_1^2(\omega)\}\big) = P\big(\{\omega \in \Omega : X(\omega) \in B\}\big) \\
&\overset{(3.4)}{=} \int_B f_\theta(x_1, x_2) dx_1 dx_2 = \int_{[0,1]} \int_{[0,x_1^2]} 1 + \theta(1 - 2x_1)(1 - 2x_2)\, dx_2 dx_1 \\
&= \ldots = \frac{1}{3} - \frac{\theta}{30}
\end{aligned}$$

[†] i.e. für jedes $\varepsilon > 0$ existiert ein $\delta > 0$ sodass für paarweise disjunkte Intervalle $(a_1, b_1), \ldots, (a_n, b_n)$ mit $\sum_{i=1}^n b_i - a_i < \delta$ gilt $\sum_{i=1}^n F_X(b_i) - F_X(a_i) < \varepsilon$.

Frage 3.6 Gelte weiterhin $(X_1, X_2) \sim \mu_{f_\theta}$ mit f_θ aus Beispiel 3.5.
(iii) Wie können wir $P(X_1 = X_2)$ im Kopf berechnen?
(iv) Wie können wir $P(X_1 \leq X_2)$ im Kopf berechnen?

Erwartungswert und Varianz haben Sie schon in der Vorlesung Stochastische Modellbildung studiert - die folgenden Definitionen sind also nicht neu:

Definition 3.7 *Sei X eine absolut stetige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Dichte f. Falls $\int_{[0,\infty)} xf(x)dx < \infty$ oder $\int_{(-\infty,0]} |x|f(x)dx < \infty$ dann heisst die Größe*

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

der Erwartungswert von X.

Definition 3.8 *Sei X eine absolut stetige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Dichte f. Falls $\int_{\mathbb{R}} x^2 f(x)dx < \infty$ dann heisst die Größe*

$$\mathbb{V}(X) = \int_{\mathbb{R}} (x - \mathbb{E}(X))^2 f(x)dx$$

die Varianz von X.

Beispiel 3.9 $\mathcal{U}(a, b)$ mit $a < b$ ist, wie am Beginn dieses Abschnitts erwähnt, absolut stetig mit Dichte

$$f_{a,b}(x) := \frac{1}{b-a} \mathbf{1}_{[a,b]}(x). \tag{3.6}$$

Für $X \sim \mathcal{U}(a, b)$ erhalten wir unschwer (Übungsaufgabe)

$$\mathbb{E}(X) = \frac{a+b}{2} \quad \text{und} \quad \mathbb{V}X = \frac{(b-a)^2}{12}.$$

Beispiel 3.10 Die wohl wichtigste Verteilung (Stichwort: Zentraler Grenzwertsatz) ist die *eindimensionale Normalverteilung* $\mathcal{N}(\mu, \sigma^2)$, die die folgende Dichte hat:

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\mu \in \mathbb{R}, \sigma > 0) \tag{3.7}$$

Für $X \sim \mathcal{N}(\mu, \sigma^2)$ gilt (Übungsaufgabe)

$$\mathbb{E}(X) = \mu \quad \text{und} \quad \mathbb{V}X = \sigma^2.$$

Weiters erfüllt die Normalverteilung folgende Transformationseigenschaft:

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Longleftrightarrow \quad \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1) \tag{3.8}$$

Im Folgenden bezeichnen wir, wie allgemein üblich, die Verteilungsfunktion von $\mathcal{N}(0, 1)$ mit Φ, i.e. $\Phi(x) = \int_{(-\infty,x]} \varphi_{0,1}(t)dt$.

Beispiel 3.11 Offensichtlich ist für jedes $\theta > 0$ die Funktion f_θ, definiert durch

$$f_\theta(x) = \theta e^{-\theta x} \mathbf{1}_{[0,\infty)}(x) \tag{3.9}$$

eine Wahrscheinlichkeitsdichte. Die zugehörige Verteilung heisst *Exponentialverteilung*, wir schreiben $X \sim \mathcal{E}(\theta)$. Es lässt sich einfach nachrechnen, dass die Verteilungsfunktionen F_θ gegeben ist durch $F_\theta(x) = 1 - e^{-\theta x}$, und dass für Erwartungswert und Varianz von $X \sim \mathcal{E}(\theta)$ gilt:

$$\mathbb{E}X = \frac{1}{\theta} \quad \text{und} \quad \mathbb{V}X = \frac{1}{\theta^2}.$$

Satz 3.12 *Sei X eine nichtnegative Zufallsvariable mit stetiger Verteilungsfunktion $F(x) = P(X \leq x)$. Dann ist X exponentialverteilt (i.e. $\exists\theta$ sodass $X \sim \mathcal{E}(\theta)$) genau dann, wenn für alle $t, s \geq 0$ die folgende Eigenschaft der* Gedächtnislosigkeit *gilt*†*:*

$$P(X \geq s + t) = P(X \geq s)P(X \geq t) \tag{3.10}$$

Beweis: Übungsaufgabe

Beispiel 3.13 Die Funktion f_k, definiert durch

$$f_k(x) = \frac{1}{2\Gamma(k/2)} \left(\frac{x}{2}\right)^{k/2-1} e^{-\frac{x}{2}} \mathbf{1}_{[0,\infty)}(x) \tag{3.11}$$

ist für jedes $k \in \mathbb{N}$ eine Wahrscheinlichkeitsdichte. Die zugehörige Verteilung heisst *Chiquadratverteilung mit k Freiheitsgraden*, wir schreiben $X \sim \chi(k)$. Für Erwartungswert und Varianz gilt

$$\mathbb{E}X = k \quad \text{und} \quad \mathbb{V}X = 2k.$$

Die Chiquadratverteilung $\chi(1)$ ergibt sich in natürlicher Weise aus der Normalverteilung - es gilt das folgende einfache Resultat:

Lemma 3.14 *Ist $X \sim \mathcal{N}(0,1)$, dann gilt $X^2 \sim \chi(1)$.*

Beweis: Direktes Nachrechnen: Wir setzen $Y = X^2$, dann gilt für festes $x \in [0,\infty)$ offensichtlich

$$F_Y(x) = P(X^2 \leq x) = P^X\big([-\sqrt{x}, \sqrt{x}]\big) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1.$$

Nachdem Φ stetig differenzierbar (sogar C^∞) ist, folgt durch Differentiation für $x > 0$

$$F_Y'(x) = 2\varphi_{0,1}(\sqrt{x})\frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{x}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}.$$

Letzteres ist aber genau die Dichte von $\chi(1)$. ■

†Wir schreiben hier und im Folgenden kurz $P(X \geq s)$ für $P(\{\omega \in \Omega : X(\omega) \geq s\}) = P^X([s,\infty))$

Beispiel 3.15 Die Funktion l_{μ,σ^2}, definiert durch

$$l_{\mu,\sigma^2}(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}} \mathbf{1}_{(0,\infty)}(x) \tag{3.12}$$

ist eine Wahrscheinlichkeitsdichte für $\mu \in \mathbb{R}$ und $\sigma > 0$. Die zugehörige Verteilung heisst *Logarithmische Normalverteilung*, wir schreiben $X \sim \mathcal{L}(\mu, \sigma^2)$. Für Erwartungswert und Varianz gilt

$$\mathbb{E}X = e^{\mu+\frac{\sigma^2}{2}} \quad \text{und} \quad \mathbb{V}X = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

Der Name Logarithmische Normalverteilung ist nicht zufällig gewählt - es gilt das folgende einfache Resultat:

Lemma 3.16 $X \sim \mathcal{L}(\mu, \sigma^2)$ *dann und nur dann wenn* $\ln(X) \sim \mathcal{N}(\mu, \sigma^2)$.

Beweis: Direktes Nachrechnen.

Beispiel 3.17 Die Funktion $f_{\alpha,\beta}$, definiert durch ($B(\alpha, \beta)$ bezeichnet die Betafunktion aus der Analysis)

$$f_{\alpha,\beta}(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1}\mathbf{1}_{[0,1]}(x) \tag{3.13}$$

ist eine Wahrscheinlichkeitsdichte für $\alpha, \beta > 0$. Die zugehörige Verteilung heisst *Betaverteilung*, wir schreiben $X \sim \mathcal{B}(\alpha, \beta)$. Für Erwartungswert und Varianz gilt

$$\mathbb{E}X = \frac{\alpha}{\alpha + \beta} \quad \text{und} \quad \mathbb{V}X = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

Wir schließen diesen Abschnitt mit dem folgenden nützlichen Resultat im eindimensionalen Fall.

Satz 3.18 (Transformationssatz für Dichten) *Sei X eine absolut stetige Zufallsvariable mit Dichte f und $T : \mathbb{R} \to \mathbb{R}$ stetig differenzierbar mit $T' \neq 0$ und gelte $T(\mathbb{R}) = (a, b)^\dagger$. Dann ist $Y := T \circ X$ ebenfalls absolut stetig mit Dichte*

$$g(z) = \frac{f(T^{-1}z)}{|T'(T^{-1}(z))|}\mathbf{1}_{(a,b)}(z)$$

Beweis: Direkte Folgerung aus der Kettenregel: Ist T strikt wachsend, dann gilt

$$P(T \circ X \leq x) = P(X \leq T^{-1}x) = \int_{(-\infty,T^{-1}x]} f(z)dz,$$

und Differentiation (Kettenregel) liefert das gewünschte Resultat. Für T strikt fallend kann analog vorgegangen werden. ■

†das Intervall ist nicht notwendigerweise endlich

Kapitel 4

Verteilungsfunktionen

4.1 Eindimensionaler Fall

Wir beginnen mit dem eindimensionalen Fall, den Sie schon aus der Vorlesung Stochastische Modellbildung kennen: Gemäß Definition 2.13 ist für jede Zufallsvariable X auf $(\Omega, \mathcal{A}, P)$ die Verteilungsfunktion $F_X : \mathbb{R} \to [0,1]$ definiert durch (für $\mu = P^X$ schreiben wir auch gelegentlich F_μ statt F_X):

$$F_X(x) := P^X((-\infty, x]) = P\big(\{\omega \in \Omega : X(\omega) \leq x\}\big) \tag{4.1}$$

F_X hat die folgenden Eigenschaften:

Lemma 4.1 *Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Dann hat die Verteilungsfunktion F_X die folgenden Eigenschaften:*

(F1) F_X ist monoton wachsend und rechtsstetig.

(F2) $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$.

Beweis: Einfache Übungsaufgabe.

Definition 4.2 *$\mathcal{F}$ bezeichnet im Folgenden die Familie aller monoton wachsenden, rechtsstetigen Funktionen $F : \mathbb{R} \to [0,1]$, die $\lim_{x\to-\infty} F_X(x) = 0$ und $\lim_{x\to\infty} F_X(x) = 1$ erfüllen.*

Für jede Wahrscheinlichkeitsverteilung $\mu = P^X$ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ gilt also $F_X \in \mathcal{F}$. Umgekehrt gilt das folgende Resultat:

Lemma 4.3 *Sei $F \in \mathcal{F}$. Dann ist durch*

$$\mu_F\big((a,b]\big) := F(b) - F(a) \tag{4.2}$$

ein eindeutiges Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definiert.

Beweis: Gemäss Satz 1.11 reicht es, σ-Additivität auf $\mathcal{I}_1$ zu zeigen. Siehe Appendix.

Insgesamt erhalten wir folgendes Resultat:

Satz 4.4 *Jede Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ entspricht einem eindeutigen Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und umgekehrt. Als Konsequenz daraus nennen wir $\mathcal{F}$ die* Familie aller (eindimensionalen) Verteilungsfunktionen *und schreiben auch $X \sim F$.*

Sei X eine *diskrete Zufallsvariable* mit Wertebereich $\{\alpha_1, \alpha_2, \ldots\}$. Dann existieren paarweise disjunkte Menge $A_1, A_2, \ldots \in \mathcal{A}$ mit $\bigcup_{i=1}^{\infty} A_i = \Omega$, sodass sich X in der Form

$$X(\omega) = \sum_{i=1}^{\infty} \alpha_i \mathbf{1}_{A_i}(\omega).$$

darstellen lässt (warum?). Offensichtlich gilt $X(\omega) \leq x$ genau dann, wenn $\omega \in A_i$ für ein i mit $\alpha_i \leq x$. Für die Verteilungsfunktion F_X ergibt sich damit unmittelbar

$$F_X(x) = P^X((-\infty, x]) = P\Big(\bigcup_{i:\alpha_i \leq x} A_i\Big) = \sum_{i:\alpha_i \leq x} P(A_i) = \sum_{i=1}^{\infty} P(A_i)\mathbf{1}_{[\alpha_i,\infty)}(x). \tag{4.3}$$

Wegen $F_X(\alpha_i) - F_X(\alpha_i-)^\dagger = P(A_i) = P^X(\{\alpha_i\})$ ist also jedes α_i mit $P^X(\{\alpha_i\}) > 0$ Sprungstelle von F_X. Wir nennen $F \in \mathcal{F}$ naheliegenderweise *diskrete Verteilungsfunktion*, wenn F von der Form (4.3) ist.
Ist X hingegen *absolut stetig*, dann ist die Verteilungsfunktion F_X, wie im vorigen Abschnitt erwähnt, sogar absolut stetig, also inbesondere stetig.
Ist jede Funktion aus $\mathcal{F}$ entweder von der Form (4.3) oder absolut stetig, oder eine Kombination der beiden Typen? Die Antwort auf diese Frage ist 'nein', i.e. es gibt Verteilungsfunktionen, die weder diskret noch absolut stetig noch eine Kombination der beiden Typen sind. Diese Funktionen F tragen den Namen *singuläre Verteilungsfunktionen* und sind insbesondere vom analytischen (oder maßtheoretischen) Standpunkt, weniger aber vom rein statistischen Standpunkt aus, interessant, da sie einerseits stetig sind, aber andererseits fast überall† $F' = 0$ gilt. Die wohl bekannteste singuläre Verteilungsfunktion ist die Cantor Funktion C_∞ (oft auch 'devil's staircase' genannt);

Beispiel 4.5 (Exkursion: Konstruktion der Cantorfunktion C_∞) Wir betrachten $\mathcal{F}_0$, definiert durch

$$\mathcal{F}_0 := \{F \in \mathcal{F} : F(0) = 0, F(1) = 1 \text{ und } F \text{ stetig}\},$$

versehen mit der Maximumsmetrik d_∞ (siehe Analysis), sowie die Abbildung $\Phi : \mathcal{F}_0 \to \mathcal{F}_0$, definiert durch (eine Skizze für $F(x) = x$ hilft):

$$(\Phi(F))(x) = \begin{cases} \frac{1}{2}F(3x) & \text{für } x \in [0, \frac{1}{3}], \\ \frac{1}{2} & \text{für } x \in (\frac{1}{3}, \frac{2}{3}], \\ \frac{1}{2} + \frac{1}{2}F(3x-2) & \text{für } x \in (\frac{2}{3}, 1]. \end{cases}$$

Wir zeigen nun, dass ein $C_\infty \in \mathcal{F}_0$ existiert, sodass für jedes $F \in \mathcal{F}_0$

$$\lim_{n\to\infty} d_\infty(\Phi^n F, C_\infty) = 0 \tag{4.4}$$

gilt, und beweisen dann, dass C_∞ singulär ist.
(Schritt 1): Aus der Analysis ist bekannt, dass $(C([0,1]), d_\infty)$ vollständig (also ein Banachraum) ist. $\mathcal{F}_0$ ist eine abgeschlossene Teilmenge von $(C([0,1]), d_\infty)$, daher ist auch $(\mathcal{F}_0, d_\infty)$ vollständig. Unser Ziel ist es zu zeigen, dass Φ eine Kontraktion auf $(\mathcal{F}_0, d_\infty)$ ist, und dann

$^\dagger F_X(a-)$ bezeichnet den linksseitigen Grenzwert von F_X an der Stelle a

†im Sinne des Lebesgue'schen Maßes λ, i.e. es existiert eine Menge $A \in \mathcal{B}(\mathbb{R})$ mit $\lambda(A) = 0$ sodass $F'(x) = 0$ für alle $x \in A^c$

den Banach'schen Fixpunktsatz anzuwenden.
(Schritt 2): Seien $F, G \in \mathcal{F}_0$ beliebig. Für $x \in [0, \frac{1}{3}]$ gilt nach Konstruktion

$$|\Phi(F)(x) - \Phi(G)(x)| = \left|\frac{1}{2}F(3x) - \frac{1}{2}G(3x)\right| \leq \frac{1}{2}\, d_\infty(F, G).$$

Analog gilt für $x \in [\frac{2}{3}, 1]$

$$|\Phi(F)(x) - \Phi(G)(x)| = \left|\frac{1}{2} + \frac{1}{2}F(3x-2) - \frac{1}{2} - \frac{1}{2}G(3x-2)\right| \leq \frac{1}{2}\, d_\infty(F, G).$$

Wegen $\Phi(F) = \Phi(G)$ auf $[\frac{1}{3}, \frac{2}{3}]$ folgt also insgesamt $d_\infty(\Phi(F), \Phi(G)) \leq \frac{1}{2}\, d_\infty(F, G)$. Nachdem F und G beliebig waren, haben wir also gezeigt, dass Φ eine Kontraktion auf $(\mathcal{F}_0, d_\infty)$ (mit Kontraktionsfaktor $\frac{1}{2}$) ist.
(Schritt 3): Direkte Anwendung des Banach'schen Fixpunktsatzes liefert die Existenz einer Verteilungsfunktion $C_\infty \in \mathcal{F}_0$, für die Gleichung (4.4) gilt. Es bleibt daher nur noch die Singularität von C_∞ zu beweisen. Bezeichnen wir die (schon in der Vorlesung erwähnte) Cantor Menge mit $C^\star$, dann lässt sich leicht überprüfen, dass für jedes offene Intervall I aus $B := [0,1] \setminus C^\star$ die Funktion $\Phi^n(F)$ ab einem Index konstant ist, und zwar für jedes $F \in \mathcal{F}_0$. Daher ist auch C_∞ auf I konstant und es folgt $C'_\infty = 0$ auf I. Da für jedes $x \in B$ ein offenes Intervall I mit $x \in I \subseteq B$ existiert, folgt $C'_\infty(x) = 0$ für jedes $x \in B$. Wegen $\lambda(C^\star) = 0$ gilt $\lambda(B) = 1$ und die Konstruktion ist komplett.

Bemerkung 4.6 Die Cantorfunktion ist eine relativ 'brave' singuläre Funktion in dem Sinne, dass C_∞ zumindest nicht strikt wachsend ist. Es können aber auch stetige Verteilungsfunktionen F mit $F' = 0$ fast überall konstruiert werden, die sogar strikt wachsend sind. Das wohl bekannteste Beispiel (1904) ist Minkowski's Fragezeichenfunktion $?(x)$, die ihrerseits in engem Zusammenhang mit der Kettenbruchentwicklung reeller Zahlen steht. Generell sind Ziffernentwicklungen und Fraktale zwei der Hauptwerkzeuge um singuläre Funktionen zu konstruieren (aktives Forschungsgebiet, interessantes Thema für Bachelorarbeit).

Satz 4.7 (Zerlegungssatz) *Jede Verteilungsfunktion $F \in \mathcal{F}$ lässt sich schreiben als Konvexkombination einer diskreten, einer absolut stetigen und einer singulären Verteilungsfunktion, i.e. es existieren $a, b, c \in [0,1]$ mit $a+b+c = 1$ und eine diskrete Verteilungsfunktion F_{dis}, eine absolut stetige Verteilungsfunktion F_{abs} und eine singuläre Verteilungsfunktion F_{sing} mit $F = aF_{dis} + bF_{abs} + cF_{sing}$*

Beweis: Übersteigt leider den Rahmen dieser Vorlesung, siehe Maßtheorie oder [2, 4, 5, 7].

Bemerkung 4.8 Um für eine gegebene Verteilungsfunktion $F \in \mathcal{F}$ festzustellen, ob sie einen diskreten, absolut stetigen oder singulären Anteil hat, ist es oft hilfreich, wiefolgt vorzugehen. (i) Alle Sprungstellen der Funktion und die entsprechenden Sprunghöhen bestimmen den diskreten Anteil - hat F keine Unstetigkeitsstellen so hat F keinen diskreten Anteil (i.e. $a = 0$ für a in Satz 4.7. (ii) Die Ableitung f von F' (wo immer definiert und > 0) bestimmt den absolut stetigen Anteil F_{abs} und den Parameter b. Sind a, F_{dis}, b, F_{abs} bekannt, kann c und F_{sing} direkt aus $F = aF_{dis} + bF_{abs} + cF_{sing}$ berechnet werden.

Für jedes sample $x_1, \ldots, x_n$ ist die empirische Verteilung $\hat{\mu}_n$ und die empirische Verteilungsfunktion $\hat{F}_n$ definiert durch

$$\hat{\mu}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_A(x_i) \quad \text{für } A \in \mathcal{B}(\mathbb{R}) \tag{4.5}$$

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,x]}(x_i) \quad \text{für } x \in \mathbb{R}. \tag{4.6}$$

Beachten Sie, dass dies genau der diskreten Gleichverteilung auf $\{x_1, \ldots, x_n\}$ entspricht. Offensichtlich ist $\hat{F}_n$ eine diskrete Verteilungsfunktion.

4.2 Stichprobenerzeugung

Nicht nur für die Entwicklung und Validierung neuer Verfahren der schließenden Statistik sondern auch ganz allgemein zur Schärfung des statistischen Grundverständnisses ist es essentiell, Stichproben $x_1, \ldots, x_n$ von Zufallsvariablen $X \sim F$ für beliebiges F erzeugen zu können. Stichproben der gängigsten (diskreten oder absolut stetigen) Verteilung können unschwer in R erzeugt werden - in der Praxis treten jedoch auch andere Verteilungen, insbesondere Mischungen diskreter und absolut stetiger Verteilungen auf. Im Folgenden überlegen wir uns daher allgemein, wie Stichproben von Zufallsvariablen $X \sim F$ für beliebiges F unter Zuhilfenahme der uniformen Gleichverteilung $\mathcal{U}(0,1)$ erzeugt werden können und starten mit einem einfachen Beispiel:

Beispiel 4.9 Wir erzeugen Stichproben von $X \sim \mathcal{N}(0,1)$; wie in Kapitel 3 bezeichne Φ die Verteilungsfunktion von $\mathcal{N}(0,1)$. Sei $Z \sim \mathcal{U}(0,1)$, dann erhalten wir für die Verteilungsfunktion F_X von $X = \Phi^{-1}(Z)$ (warum ist X Zufallsvariable?).

$$\begin{aligned} F_X(x) &= P(X \leq x) = P\big(\Phi^{-1}(Z) \leq x\big) = P\big(\{\omega \in \Omega : \Phi^{-1}(Z(\omega)) \leq x\}\big) \\ &\overset{\Phi \text{ strikt wachsend}}{=} P\big(\{\omega \in \Omega : Z(\omega) \leq \underbrace{\Phi(x)}_{\in[0,1]}\}\big) = F_Z(\Phi(x)) = \Phi(x) \end{aligned}$$

Nachdem $x \in \mathbb{R}$ beliebig war, folgt sofort $X \sim \mathcal{N}(0,1)$. Stichproben $x_1, \ldots, x_n$ von $X \sim \mathcal{N}(0,1)$ können also wiefolgt erzeugt werden:

(S1) Erzeuge eine Stichprobe $z_1, \ldots, z_n$ von $Z \sim \mathcal{U}(0,1)$.

(S2) Betrachte $x_1 := \Phi^{-1}(z_1), \ldots, x_n := \Phi^{-1}(z_n)$

Die in Beispiel 4.9 verwendete Idee kann sofort auf beliebige Fälle $X \sim F$ übertragen werden, in denen F strikt wachsend ist. Verteilungsfunktion sind aber im Allgemeinen nicht streng monoton wachsend - es ist daher notwendig, mit Quasiinversen (anstatt mit Inversen) zu arbeiten:

Definition 4.10 *Für jede Verteilungsfunktion F ist die* Quasiinverse *(oft auch Pseudo- oder Linksinverse) $F^- : [0,1] \to \mathbb{R}$ definiert durch ($\inf \emptyset := \infty$)*

$$F^-(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}. \tag{4.7}$$

Der Wert $F^-(p)$ heißt p-Quantil *von F, $F^-(0.5)$ heißt* Median.

Bemerkung 4.11 Falls $F \in \mathcal{F}$ injektiv ist, dann gilt natürlich $F^- = F^{-1}$

Der folgende Satz fasst die wichtigsten Eigenschaften von F^- zusammen:

Satz 4.12 *Seien F, G Verteilungsfunktionen und ihre Quasiinverse F^-, G^- definiert durch (4.7). Dann sind F^-, G^- monoton wachsend und für $y \in (0,1)$ gelten die folgenden Aussagen:*

1. $F^-(y) = \min\{x \in \mathbb{R} : F(x) \geq y\}$.
2. $F^-(y) \leq x$ *dann und nur dann wenn* $y \leq F(x)$.
3. $F^-(F(x)) \leq x \leq F^-(F(x)+)$.
4. $F(F^-(y)-) \leq y \leq F(F^-(y))$ *[Für stetiges F gilt also $F \circ F^-(y) = y$]*
5. F^- *ist linksstetig.*
6. $F^- \circ F \circ F^- = F^-$.
7. $F \circ F^- \circ F = F$.
8. $(G \circ F)^- = F^- \circ G^-$.
9. *Es gilt* $P(F^- \circ F \circ X = X) = 1$.

Beweis: Monotonie von F^-, G^- folgt direkt aus der Definition, Rechtsstetigkeit von F impliziert Punkt 1. Die zweite Aussage lässt sich wie folgt zeigen: Ist $F^-(y) > x$, dann muss offensichtlich $y > F(x)$ gelten. Umgekehrt folgt aus $F(x) < y$ und Rechtsstetigkeit von F unmittelbar $x < F^-(y)$. Der Beweis der restlichen Punkte ist eine Übungsaufgabe. ■

Die Wichtigkeit der Pseudoinversen für die Stichprobenerzeugung wird durch folgendes Resultat unterstrichen:

Satz 4.13 *Es gelten die folgenden beiden Aussagen:*

1. *Ist $X \sim F$ und $F \in \mathcal{F}$ stetig dann gilt $F \circ X \sim \mathcal{U}(0,1)$. ($F \circ X$ wird meist als probability integral transform (PIT) bezeichnet)*
2. *Ist $Z \sim \mathcal{U}(0,1)$ und $F \in \mathcal{F}$ beliebig, dann folgt $F^- \circ Z \sim F$.*

Beweis: Als monoton wachsende Funktion ist F Borel messbar (Übungsaufgabe) und Satz 2.9 impliziert, dass $F \circ X$ eine Zufallsvariable ist. Unter Verwendung von Satz 4.12 und der Stetigkeit von F folgt für $y \in (0,1)$

$$P\big(F \circ X < y\big) \overset{4.12\,(2)}{=} P\big(X < F^-(y)\big) = P\big(X \leq F^-(y)\big) = F \circ F^-(y) \overset{4.12\,(4)}{=} y.$$

Wegen $\bigcap_{n=1}^{\infty}(-\infty, y + \frac{1}{n}) = (-\infty, y]$ und der Stetigkeit von Wahrscheinlichkeitsmaßen (Lemma 1.7) erhalten wir damit insgesamt

$$P\big(F \circ X \leq y\big) = \lim_{n \to \infty} P\big(F \circ X < y + \tfrac{1}{n}\big) = \lim_{n \to \infty} \big(y + \tfrac{1}{n}\big) = y,$$

woraus die erste Aussage folgt.
Die zweite Aussage ist eine direkte Konsequenz von Punkt 2. in Satz 4.12:

$$P\big(F^- \circ Z \leq x\big) = P\big(Z \leq F(x)\big) = F(x) \quad ■$$

Frage 4.14 Kann $F \circ X \sim \mathcal{U}(0,1)$ auch für unstetiges F gelten?

Beispiel 4.15 Wir betrachten zur Veranschaulichung von F^- das folgende sehr einfache Beispiel (weitere Beispiele in den Übungen). Sei $p \in (0,1)$ und die Verteilungsfunktion F, definiert durch

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1-p & \text{für } x \in [0,1) \\ 1 & \text{für } x \geq 1. \end{cases}$$

Dann ergibt sich für die Quasiinverse sofort

$$F^-(y) = \begin{cases} 0 & \text{für } y \in (0, 1-p] \\ 1 & \text{für } y \in (1-p, 1]. \end{cases}$$

Für $Z \sim \mathcal{U}(0,1)$ hat die Zufallsvariable $F^- \circ Z$ gemäß Satz 4.13 dann genau die Verteilungsfunktion F. Wie würden Sie Stichproben von $X \sim F$ ausgehend von $Z \sim \mathcal{U}(0,1)$ ohne Satz 4.13 erzeugen?

Beispiel 4.16 Wir betrachten die Verteilungsfunktion F der *Cauchy Verteilung*, definiert durch

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x), \quad x \in \mathbb{R}. \tag{4.8}$$

Offensichtlich ist F stetig differenzierbar und wir erhalten als Dichte

$$f(x) = F'(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

F ist strikt monoton wachsend und wir erhalten für die Quasiinverse

$$F^-(y) = F^{-1}(y) = \tan\Big(\Big(y - \frac{1}{2}\Big)\pi\Big), \quad y \in (0,1).$$

Als direkte Konsequenz von Satz 4.13 ist für $Z \sim \mathcal{U}(0,1)$ die Zufallsvariable $F^- \circ Z$ Cauchyverteilt. Beachten Sie, dass für eine Cauchyverteilte Zufallsvariable X der Erwartungswert nicht existiert (warum ?).

Beispiel 4.17 Wir betrachten die Verteilungsfunktion $F_{\alpha,\beta}$ der *Paretoverteilung* (wir schreiben $\mathcal{P}(\alpha,\beta)$), definiert durch $(\alpha, \beta > 0)$

$$F_{\alpha,\beta}(x) = \left(1 - \Big(\frac{\alpha}{x}\Big)^{\beta}\right)\mathbf{1}_{(\alpha,\infty)}(x). \tag{4.9}$$

Offensichtlich ist F stetig differenzierbar und wir erhalten als Dichte

$$f(x) = F'(x) = \beta\frac{\alpha^\beta}{x^{\beta+1}}\mathbf{1}_{(\alpha,\infty)}(x).$$

F ist strikt monoton wachsend und wir erhalten für die Quasiinverse

$$F^-(y) = F^{-1}(y) = \alpha(1-y)^{-\frac{1}{\beta}} \quad y \in (0,1).$$

Als direkte Konsequenz von Satz 4.13 ist für $Z \sim \mathcal{U}(0,1)$ die Zufallsvariable $F^- \circ Z \sim \mathcal{P}(\alpha,\beta)$.

4.3 Multivariate Verteilungsfunktionen

Wir beschränken uns der Einfachheit halber auf den zweidimensionalen Fall und beginnen analog zum vorigen Abschnitt mit dem folgenden Lemma:

Lemma 4.18 *Sei $X = (X_1, X_2)$ ein 2-dimensionaler Zufallsvektor auf $(\Omega, \mathcal{A}, P)$. Dann hat die Verteilungsfunktion (siehe Gleichung 2.7) $F_X : \mathbb{R}^2 \to [0,1]$, definiert durch*

$$F_X(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

die folgenden Eigenschaften:

(F1) F_X ist rechtsstetig, i.e. für monoton fallende Folgen $(x_1^n)_{n\in\mathbb{N}}, (x_2^n)_{n\in\mathbb{N}}$ mit Grenzwert x_1 bzw. x_2 gilt $\lim_{n\to\infty} F_X(x_1^n, x_2^n) = F_X(x_1, x_2)$.

(F2) $\lim_{x_1\to-\infty} F_X(x_1, x_2) = 0 = \lim_{x_2\to-\infty} F_X(x_1, x_2)$; $\lim_{x_1,x_2\to\infty} F_X(x_1, x_2) = 1$.

(F3) Für Invervalle $(a_1, a_2], (b_1, b_2]$ gilt

$$F_X(a_2, b_2) - F_X(a_1, b_2) - F_X(a_2, b_1) + F_X(a_1, b_1) \geq 0.$$

Beweis: Direktes Nachrechnen, Übungsaufgabe.

Definition 4.19 *$\mathcal{F}_2$ bezeichnet im Folgenden die Menge aller Funktionen $F : \mathbb{R}^2 \to [0,1]$, die die Punkte (F1),(F2),(F3) aus Lemma 4.18 erfüllen.*

Lemma 4.20 *Sei $F \in \mathcal{F}_2$. Dann ist durch*

$$\mu_F((a_1, a_2] \times (b_1, b_2])) := F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1)$$

ein eindeutiges Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}^2)$ definiert.

Beweis: Analog dem Beweis von Lemma 4.3.

Satz 4.21 *Jede Verteilungsfunktion $F : \mathbb{R}^2 \to [0,1]$ entspricht einem eindeutigen Wahrscheinlichkeitsmaß auf $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ und umgekehrt. Als Konsequenz daraus nennen wir $\mathcal{F}_2$ die* Familie aller zweidimensionalen Verteilungsfunktionen*.*

Bemerkung 4.22 Im Folgenden werden wir (um die Notation möglichst einfach zu halten) 2-dimensionale Verteilungsfunktionen auch oft mit H bezeichnen.

Beispiel 4.23 Wir betrachten $F, G \in \mathcal{F}$ und definieren

$$H_1(x, y) = \min\{F(x), G(y)\}, \quad H_2(x, y) = F(x)G(y).$$

Dann rechnet man unschwer nach, dass sowohl H_1 als auch H_2 zweidimensionale Verteilungsfunktionen sind.

Bemerkung 4.24 Beachten Sie, dass für $X = (X_1, X_2) \sim H$, die Verteilungsfunktionen F und G von X_1 und X_2 (die sogenannten *Randverteilungen*) direkt aus der gemeinsamen Verteilungsfunktion H berechnet werden können - es gilt nämlich

$$\begin{aligned} F(x) &= P(X \leq x) = \lim_{n\to\infty} P(X \leq x, Y \leq n) = \lim_{n\to\infty} H(x, n) =: H(x, \infty) \\ G(y) &= P(Y \leq y) = \lim_{n\to\infty} P(X \leq n, Y \leq y) = \lim_{n\to\infty} H(n, y) =: H(\infty, y). \end{aligned}$$

Frage 4.25 Wie vorhin gelte $(X_1, X_2) \sim H$. Offensichtlich können F und G aus H berechnet werden. Gilt dies auch umgekehrt?
Hinweis: Berechnen Sie für H_1, H_2 in Beispiel 4.23 die Randverteilungen - was ist zu erkennen?

Beispiel 4.26 Gegeben ist die Funktion $H(x,y) = (1 - e^{-x})(1 - e^{-2y})$. Dann ist H eine 2-dimensionale Verteilungsfunktion (warum?), die $H(x,y) = 0$ für $\min\{x,y\} \leq 0$ erfüllt. Zusätzlich ist H offensichtlich stetig. Wir berechnen $\mathbb{P}(Y \leq X)$ für $(X,Y) \sim H$ und bestimmen dafür zuerst die Dichte f von (X,Y) für $x,y > 0$ mittels

$$f(x,y) = \frac{\partial^2 H}{\partial x \partial y}(x,y) = 2e^{-x}e^{-2y}.$$

Insgesamt erhalten wir daher

$$f(x,y) = 2e^{-(x+2y)}\mathbf{1}_{[0,\infty)^2}(x,y).$$

Offensichtlich ist f eine 2-dimensionale Wahrscheinlichkeitsdichte. Die gesuchte Wahrscheinlichkeit kann daher sofort berechnet werden durch

$$\begin{aligned}\mathbb{P}(Y \leq X) &= \mathbb{P}^{(X,Y)}\Big(\underbrace{\{(x,y) \in [0,\infty)^2 : y \leq x\}}_{=:B}\Big) = \int_B f d\lambda_2 \\ &= \int_{[0,\infty)} \int_{[0,x]} f(t,s) d\lambda(s) d\lambda(t) = \ldots = \frac{2}{3}.\end{aligned}$$

Beispiel 4.27 (Fortsetzung von Beispiel 4.26) Gelte wiederum $(X,Y) \sim H$ mit H aus Beispiel 4.26. Als nächstes berechnen wir die Verteilungsfunktionen F von X_1 und G von X_2 mit Hilfe von Bemerkung 4.24: Für $x < 0$ gilt offensichtlich $F(x) = 0$ und für $x \geq 0$ erhalten wir

$$F(x) = \lim_{n\to\infty} H(x,n) = 1 - e^{-x}.$$

Analog ergibt sich $G(y) = 0$ für $y < 0$ sowie

$$G(y) = \lim_{n\to\infty} H(n,y) = 1 - e^{-2y}$$

für $y > 0$. Mit anderen Worten: $X \sim \mathcal{E}(1)$ und $Y \sim \mathcal{E}(2)$. Zusätzlich gilt offensichtlich auf ganz $\mathbb{R}^2$ die Gleichheit $H(x,y) = F(x)G(y)$, i.e. die gemeinsame Verteilungsfunktion ist das Produkt der Randverteilungsfunktionen. Schreiben wir $\Pi(u,v) = uv$ für $u,v \in [0,1]$, dann gilt also

$$H(x,y) = \Pi(F(x), G(y)) \tag{4.10}$$

für alle $x,y \in \mathbb{R}$. Die Verteilungsfunktion H lässt sich also zerlegen in die Randverteilungen F, G und eine 'Verknüpfung' selbiger.

M. Frechét studierte Anfang der 1950er Jahre die Frage, wie, für Zufallsvariable X_1 und X_2 mit Verteilungsfunktion F bzw. G alle Verteilungsfunktionen H des Vektors (X_1, X_2) aussehen können. Zur Beantwortung dieser Frage führte A. Sklar 1959 den folgenden Begriff ein:

Definition 4.28 *Eine* Copula *ist eine Funktion* $A : [0,1]^2 \to [0,1]$ *mit den folgenden Eigenschaften:*

1. $\forall x \in [0,1] : A(x,1) = A(1,x) = x,\ A(x,0) = A(0,x) = 0.$
2. *Für* $0 \leq x_1 \leq x_2 \leq 1$ *und* $0 \leq y_1 \leq y_2 \leq 1$ *gilt*

$$A(x_2, y_2) - A(x_1, y_2) - A(x_2, y_1) + A(x_1, y_1) \geq 0. \tag{4.11}$$

$\mathcal{C}$ *bezeichnet im Folgenden die Familie aller Copulas.*

Beispiel 4.29 Die folgenden Funktionen sind Copulas (direkt nachzurechnen):

$$M(x,y) = \min\{x,y\}, \quad \Pi(x,y) = xy, \quad W(x,y) = \max\{x+y-1, 0\}$$

Lemma 4.30 *Jede Copula* A *hat die folgenden Eigenschaften:*

1. A *ist monoton wachsend in jeder Koordinate.*
2. *Für* $x_1, x_2, y_1, y_2 \in [0,1]$ *gilt*

$$|A(x_2, y_2) - A(x_1, y_1)| \leq |x_2 - x_1| + |y_2 - y_1|. \tag{4.12}$$

Beweis: Beide Aussagen folgen aus der Eigenschaft (4.11): Betrachten wir zum Beispiel $x_1 = 0$, dann folgt wegen

$$0 \leq A(x_2, y_2) - A(0, y_2) - A(x_2, y_1) + A(0, y_1) = A(x_2, y_2) - A(x_2, y_1)$$

unmittelbar die Monotonie von A bezüglich der y-Koordinate. Monotonie bezüglich x folgt analog. Betrachten wiederum Eigenschaft (4.11) für $x = x_1$ und $x_2 = 1$ sowie $y_1 \leq y_2$, dann folgt sofort

$$A(1, y_2) - A(x, y_2) - A(1, y_1) + A(x, y_1) \geq 0$$

und damit $A(x, y_2) - A(x, y_1) \leq y_2 - y_1$. Die Ungleichung $A(x_2, y) - A(x_1, y) \leq x_2 - x_1$ für $x_1 \leq x_2$ und $y \in [0,1]$ folgt analog. Anwendung der Dreiecksungleichung liefert insgesamt (4.12). ■

Satz 4.31 (Satz von Sklar) *Gelte* $X = (X, Y) \sim H$*, sowie* $X \sim F$ *und* $Y \sim G$*. Dann existiert eine Copula* $A \in \mathcal{C}$ *sodass für alle* $x, y \in \mathbb{R}$ *die folgende Gleichheit gilt:*

$$H(x,y) = A\big(F(x), G(y)\big).$$

Falls F *und* G *stetig sind, ist die Copula* A *eindeutig*[†]*.*
Umgekehrt ist für jede Copula $A \in \mathcal{C}$ *und* $F, G \in \mathcal{F}$ *die Funktion* $H(x,y) = A\big(F(x), G(y)\big)$ *eine zweidimensionale Verteilungsfunktion.*

Beweis: Entfällt (obwohl nicht schwierig) - freiwillige Übungsaufgabe für Interessierte.

Gelte $(X, Y) \sim H$, $X \sim F, Y \sim G$ und seien T, S strikt wachsende, stetige Abbildungen auf $\mathbb{R}$. Anhang von einfachen Beispielen überlegt man sich leicht, dass im Allgemeinen weder $\tilde{X} := T \circ X \sim F$ noch $\tilde{Y} := S \circ Y \sim G$ oder $(T \circ X, S \circ Y) \sim H$ gelten muss. Was sich jedoch nicht ändert ist die zugrundeliegende Copula - es gilt folgender Satz:

[†]Wir nennen A dann die (X, Y) (bzw. die H) zugrundeliegende Copula

Satz 4.32 (Transformationsinvarianz von Copulas) *Gelte $(X,Y) \sim H$, $X \sim F, Y \sim G$ und seinen F und G stetig. Weiters seien T, S strikt wachsende, stetige Abbildungen auf $\mathbb{R}$. Dann ist die (X,Y) und die $(T \circ X, S \circ Y)$ zugrundeliegende Copula gleich.*

Beweis: Übungsaufgabe. Hinweis: Es genügt, die Verteilungsfunktion von $(\tilde{X}, \tilde{Y})$ sowie die entsprechenden Randverteilungsfunktionen zu berechnen und dann einen Blick auf Satz 4.31 zu werfen.

Bemerkung 4.33 Copulas sind auch in der Praxis (insbesondere, aber nicht nur, im Finanzsektor, Stichwort Risk Management) sehr wichtig, da sich Abhängigkeiten von Zufallsvariablen über Copulas vollständig modellieren und analysieren lassen $\rightarrow$ aktives Forschungsgebiet mit vielen offenen Problemstellungen, siehe Copulas_Anwendungsbeispiel.pdf auf www.trutschnig.net/courses.

Kapitel 5

Erwartungswerte

Wir haben bisher Erwartungswerte nur für diskrete und für absolut stetige Zufallsvariable definiert und berechnet. Ziel dieses Abschnitts ist die Erweiterung des Begriffs des Erwartungswerts auf beliebige Zufallsvariable[†]. Diese Erweiterung wird einerseits das 'Rechnen' mit Erwartungswerten (und höheren Momenten) erleichtern und andererseits sicherstellen, dass alle in den folgenden Kapiteln verwendeten Konzepte wohlfundiert sind.
Wir starten mit einem ersten Resultat, das zeigt, dass Grenzwerte von Folgen von Zufallsvariablen wieder Zufallsvariable sind und zeigen nachfolgend, dass beliebige Zufallsvariable durch sehr einfache (schon bekannte) Zufallsvariable approximiert werden können.

Satz 5.1 *Seien* $X_1, X_2, X_3, \ldots (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *Zufallsvariable und die Funktionen* $\underline{Y}, \overline{Y}, \underline{X}, \overline{X} : \Omega \to [-\infty, \infty]$ *definiert durch*

$$\begin{aligned} \underline{Y}(\omega) &:= \inf_{i\in\mathbb{N}} X_i(\omega), \quad \overline{Y}(\omega) := \sup_{i\in\mathbb{N}} X_i(\omega) \\ \underline{X}(\omega) &= \liminf_{n\to\infty} X_n(\omega) := \sup_{n\in\mathbb{N}} \inf_{k\geq n} X_k(\omega) \\ \overline{X}(\omega) &= \limsup_{n\to\infty} X_n(\omega) := \inf_{n\in\mathbb{N}} \sup_{k\geq n} X_k(\omega). \end{aligned} \tag{5.1}$$

Falls $\underline{Y}$ *nur endliche Werte annimmt, dann ist* $\underline{Y}$ *ebenfalls eine Zufallsvariable. Selbige Implikation gilt für* $\overline{Y}$, $\underline{X}$ *und* $\overline{X}$.
Im Falle, dass die Folge $(X_n)_{n\in\mathbb{N}}(\omega)$ *für jedes* $\omega \in \Omega$ *gegen* $X(\omega) \in \mathbb{R}$ *konvergiert, ist auch der Grenzwert* X *eine Zufallsvariable.*

Beweis: Falls $\underline{Y}(\omega) \in \mathbb{R}$ für jedes $\omega \in \Omega$ dann gilt für jedes $b \in \mathbb{R}$ offensichtlich $\underline{Y}^{-1}((-\infty, b)) = \bigcup_{n=1}^{\infty} X_n^{-1}((-\infty, b)) \in \mathcal{A}$. Anwendung von Lemma 1.14 liefert unmittelbar die Messbarkeit von $\underline{Y}$. Die Beweise für $\overline{Y}$, $\underline{X}$ und $\overline{X}$ verlaufen vollkommen analog.
Falls nun $(X_n)_{n\in\mathbb{N}}$ punktweise gegen X konvergiert und X nur endliche Werte annimmt, dann gilt offensichtlich $X = \limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n$, woraus die Messbarkeit von X unmittelbar folgt. ■

Bemerkung 5.2 Um sich von der (inbesondere bei Grenzübergängen relevanten) Endlichkeitsbedingung zu lösen werden in vielen Lehrbüchern oft direkt Zufallsvariable mit Werten

[†]Ohne Verwendung der in der Analysis nicht behandelten Stieltjes Integrale.

in $\overline{\mathbb{R}} := [-\infty, \infty]$ betrachtet, i.e. Zufallsvariable, die auch die Werte $\{-\infty, \infty\}$ annehmen können. Es lässt sich unschwer nachweisen, dass $\mathcal{B}(\overline{\mathbb{R}})$, definiert durch

$$\mathcal{B}(\overline{\mathbb{R}}) := \big\{A \cup E : A \in \mathcal{B}(\mathbb{R}), E \subseteq \{-\infty, \infty\}\big\}$$

wieder eine σ-Algebra ist. Messbarkeit einer Abbildung $X : (\Omega, \mathcal{A}, P) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ ist dann in gewohnter Weise definiert, i.e. für jedes $B \in \mathcal{B}(\overline{\mathbb{R}})$ muss $X^{-1}(B) \in \mathcal{A}$ gelten. Wir werden in der Folge die Fälle $\mathbb{R}$ und $\overline{\mathbb{R}}$ nicht explizit unterschieden (da die Resultate, insbesondere Satz 5.1, auch für den größeren Messraum $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ gelten) und der Einfachheit halber in beiden Fällen nur von 'Zufallsvariablen' sprechen.
Konsistenterweise wird meist auch $\lim_{n\to\infty} x_n = \infty$ geschrieben falls für jedes $M > 0$ ein Index $n_0 = n_0(\varepsilon)$ existiert, sodass $x_n > M$ für alle $n \geq n_0$ gilt.

Definition 5.3 *Eine Zufallsvariable* $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *heisst* einfach *genau dann, wenn paarweise verschiedene Werte* $\alpha_1, \alpha_2, \ldots, \alpha_n \in [0, \infty)$ *und paarweise disjunkte Mengen* $A_1, A_2, \ldots, A_n \in \mathcal{A}$ *mit* $\bigcup_{i=1}^n A_i = \Omega$ *existieren, sodass*

$$X(\omega) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(\omega) \tag{5.2}$$

für alle $\omega \in \Omega$ *gilt. Die Familie aller einfachen Zufallsvariablen wird im Folgenden mit* $\mathcal{S} = \mathcal{S}(\Omega, \mathcal{A})$ *bezeichnet.*

Einfache Zufallsvariable sind also inbesondere diskret und nicht-negativ.

Bemerkung 5.4 Sie wissen schon aus der Vorlesung 'Stochastische Modellbildung', dass für Zufallsvariable in der Form (5.2) der Erwartungswert (wir schreiben $\mathbb{E}X$ oder $\mathbb{E}(X)$) definiert ist durch

$$\mathbb{E}X = \sum_{i=1}^n \alpha_i \, P^X(\{\alpha_i\}) = \sum_{i=1}^n \alpha_i P(A_i) =: \int_\Omega X(\omega)\, dP(\omega). \tag{5.3}$$

Damit ist der im Folgenden beschriebene Zugang (der Rückführung auf den diskreten Fall) in keinster Weise überraschend.

Das folgende einfache Resultat hat weitreichende Konsequenzen und ist der Schlüssel für die Verallgemeinerung des Erwartungswerts von diskreten auf beliebige Zufallsvariable:

Satz 5.5 *Sei* X *eine nicht-negative Zufallsvariable auf* $(\Omega, \mathcal{A}, P)$. *Dann existiert eine Folge* $X_1, X_2, X_3 \ldots \in \mathcal{S}(\Omega, \mathcal{A})$ *mit*

(a) $0 \leq X_1 \leq X_2 \leq \cdots \leq X$ *und*

(b) $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ *für jedes* $\omega \in \Omega$.

Mit anderen Worten: Jede nicht-negative Zufallsvariable ist der Grenzwert einer monoton wachsenden Folge von einfachen Zufallsvariablen.

Beweis: Für festes $n \in \mathbb{N}$ seien die Mengen $(A_{n,i})_{i=0}^{n2^n}$ definiert durch

$$A_{n,i} := \begin{cases} X^{-1}\big([\frac{i}{2^n}, \frac{i+1}{2^n})\big) & \text{für } i \in \{0, \dots, n2^n - 1\} \\ X^{-1}([n, \infty)) & \text{für } i = n2^n. \end{cases} \tag{5.4}$$

Die Messbarkeit von X impliziert $A_{n,i} \in \mathcal{A}$ für jedes $i \in \{0, \dots, n2^n\}$. Setzen wir also

$$X_n(\omega) := \sum_{i=0}^{n2^n-1} \frac{i}{2^n} \mathbf{1}_{A_{n,i}}(\omega) \, + \, n \mathbf{1}_{A_{n,n2^n}}(\omega) \tag{5.5}$$

für jedes $n \in \mathbb{N}$, dann gilt offensichtlich $X_n \in \mathcal{S}(\Omega, \mathcal{A})$. Dass die Folge $(X_n)_{n\in\mathbb{N}}$ Punkt (a) in Satz 5.5 erfüllt, kann leicht nachgeprüft werden. Zum Beweis von (b) betrachten wir ein beliebiges, aber festes $\omega_0 \in \Omega$. Wegen $X(\omega_0) < \infty$ existiert ein Index $n_0 \in \mathbb{N}$ sodass $X(\omega_0) < n$ und damit $\omega_0 \in \bigcup_{i=0}^{2^n-1} A_{n,i}$ für jedes $n \geq n_0$. Auf jedem $A_{n,i}$ mit $i < n2^n$ gilt aber nach Konstruktion $|X_n - X| \leq \frac{1}{2^n}$, also insbesondere $\lim_{n\to\infty} X_n(\omega_0) = X(\omega_0)$. ■

Ausgehend von (5.3) und Satz 5.5 ist die folgende Definition naheliegend:

Definition 5.6 *Sei X eine nicht-negative Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Dann ist der* Erwartungswert $\mathbb{E}X$ von X *definiert durch*

$$\mathbb{E}X := \int_\Omega X dP := \sup_{n\in\mathbb{N}} \mathbb{E}(X_n) = \lim_{n\to\infty} \mathbb{E}(X_n) \tag{5.6}$$

wobei $X_1, X_2, X_3, \dots$ eine (gemäß Satz 5.5 existierende) monoton wachsende Folge einfacher Zufallsvariable ist, die punktweise gegen X konvergiert.

Bemerkung 5.7 A priori ist nicht klar, ob der Erwartungswert gemäß Definition 5.6 wohldefiniert ist, d.h. ob bei Wahl einer anderen monoton wachsenden Folge einfacher Zufallsvariable $(X_n^*)_{n\in\mathbb{N}}$, die punktweise gegen X konvergiert, auch tatsächlich $\sup_{n\in\mathbb{N}} \mathbb{E}X_n = \sup_{n\in\mathbb{N}} \mathbb{E}X_n^*$ gilt. Diese Gleichheit kann aber unschwer bewiesen werden und wir können daher alternativ den Erwartungswert auch schreiben als

$$\mathbb{E}X = \sup \Big\{ \mathbb{E}Z : Z \in \mathcal{S}(\Omega, \mathcal{A}) \text{ und } 0 \leq Z \leq X \Big\}. \tag{5.7}$$

Beispiel 5.8 Wir berechnen für $X \sim \mathcal{U}(0,1)$ den Erwartungswert $\mathbb{E}(X)$ über die in Satz 5.5 verwendete Approximation (5.5) und erhalten ($P(A_{n,i}) = 0$ für $i \geq 2^n$)

$$\begin{aligned}
\mathbb{E}(X_n) &= \sum_{i=0}^{n2^n-1} \frac{i}{2^n} P(A_{n,i}) \, + \, nP(A_{n,n2^n}) = \sum_{i=0}^{2^n-1} \frac{i}{2^n} P(A_{n,i}) \\
&= \sum_{i=0}^{2^n-1} \frac{i}{2^n} P^X\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)\right) = \sum_{i=0}^{2^n-1} \frac{i}{2^n} \frac{1}{2^n} = \frac{1}{4^n} \sum_{i=0}^{2^n-1} i \\
&= \frac{1}{4^n} \frac{(2^n-1)2^n}{2}.
\end{aligned}$$

Damit folgt sofort das schon aus den vorigen Kapitel bekannte Resultat

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} \frac{1}{4^n} \frac{(2^n-1)2^n}{2} = \frac{1}{2}.$$

Weitere Beispiele folgen in den Übungen.

Als letzter Schritt erweitern wir die Definition des Erwartungswerts auf allgemeine Zufallsvariable. Jede Zufallsvariable X auf $(\Omega, \mathcal{A}, P)$ kann als Differenz zweier nicht-negativer Zufallsvariable geschrieben werden - setzen wir (Skizze!)

$$X^+(\omega) := \max(X(\omega), 0) \quad \text{und } X^-(\omega) := \max(-X(\omega), 0) = -\min(X(\omega), 0)$$

für jedes $\omega \in \Omega$, dann sind X^+, X^- nicht-negative Zufallsvariable und es gilt $X = X^+ - X^-$ sowie $|X| = X^+ + X^-$. Damit macht folgende Definition Sinn:

Definition 5.9 *Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Ist $\mathbb{E}(X^+) < \infty$ oder $\mathbb{E}(X^-) < \infty$ dann ist der* Erwartungswert $\mathbb{E}X$ von X *definiert durch*

$$\mathbb{E}(X) := \int_\Omega X dP = \mathbb{E}(X^+) - \mathbb{E}(X^-). \tag{5.8}$$

Wir nennen X integrierbar *wenn $\mathbb{E}(X^+), \mathbb{E}(X^-) < \infty$ (oder, äquivalent dazu, wenn $\mathbb{E}(|X|) < \infty$).*

Bemerkung 5.10 Alle bisher betrachteten Zufallsvariable X, mit Ausnahme Cauchy-verteilter X, sind integrierbar. Weiters ist offensichtlich jede beschränkte Zufallsvariable X (i.e. $|X| \leq M$ für ein $M > 0$) integrierbar.

Bemerkung 5.11 Für integrierbare X gilt wegen $|\mathbb{E}(X^+) - \mathbb{E}(X^-)| \leq \mathbb{E}(X^+) + \mathbb{E}(X^-)$ auch

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) < \infty. \tag{5.9}$$

Bemerkung 5.12 Es lässt sich unschwer zeigen, dass $\mathbb{E}(X)$ konsistent definiert ist, i.e. für diskrete und absolut stetige Zufallsvariable ergeben sich die schon bekannten Ausdrücke (siehe Übungen). Weiters gilt im Falle einer absolut stetigen Zufallsvariable X mit Dichte f und einer (messbaren) Funktion $\Phi : \mathbb{R} \to \mathbb{R}$ falls $\mathbb{E}(\Phi(X)^+) < \infty$ oder $\mathbb{E}(\Phi(X)^-) < \infty$ auch

$$\mathbb{E}(\Phi(X)) = \int_\mathbb{R} \Phi(x) f(x) dx.$$

Selbige Aussage gilt auch für absolut stetige Zufallsvektoren, i.e. falls $(X_1, \ldots, X_m)$ absolut stetig mit Dichte f und $\Phi : \mathbb{R}^m \to \mathbb{R}$ eine messbare Funktion mit $\mathbb{E}(\Phi(X_1, \ldots, X_m)^+) < \infty$ oder $\mathbb{E}(\Phi(X_1, \ldots, X_m)^-) < \infty$, dann gilt auch

$$\mathbb{E}(\Phi(X_1, \ldots, X_m)) = \int_{\mathbb{R}^m} \Phi(x_1 \ldots, x_m) f(x_1, \ldots, x_m) dx_1 \ldots dx_m.$$

Beispiel 5.13 Sei $X \sim \mathcal{U}(a, b)$. Die Zufallsvariable $Y = X^2$ ist beschränkt und damit integrierbar. Wir berechnen $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \int_\mathbb{R} x^2 \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_{[a,b]} x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

Der Erwartungswert ist linear und monoton - es gilt das folgende Resultat (vergleiche mit dem diskreten Fall in der Vorlesung Stochastische Modellbildung):

Satz 5.14 *Seien X, Y integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $a, b \in \mathbb{R}$. Dann ist auch $aX + bY$ eine integrierbare Zufallsvariable und es gilt*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y). \tag{5.10}$$

Gilt weiters $X \leq Y$ dann folgt $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Beweis: (Schritt 1): Wir starten analog dem bisherigen Zugang und beweisen die Linearität zuerst für nicht-negative integrierbare Zufallsvariable X, Y und $a, b \in [0, \infty)$: Sei $(X_n)_{n\in\mathbb{N}}$ eine monoton wachsende Folge von einfachen Zufallsvariablen, die gegen X konvergiert. Dann ist offensichtlich $(a\, X_n)_{n\in\mathbb{N}}$ eine nicht-fallende Folge von einfachen Zufallsvariablen, die gegen aX konvergiert. Da X nach Voraussetzung integrierbar ist und der Erwartungswert auf $\mathcal{S}(\Omega, \mathcal{A})$ monoton ist erhalten wir damit unmittelbar

$$\mathbb{E}(aX) = \lim_{n\to\infty} \mathbb{E}(aX_n) = a \lim_{n\to\infty} \mathbb{E}(X_n) = a\mathbb{E}(X).$$

Um $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ zu zeigen gehen wir vollkommen analog vor und betrachten wachsende Folgen $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ einfacher Zufallsvariable die gegen X bzw. Y konvergieren:

$$\mathbb{E}(X + Y) = \lim_{n\to\infty} \mathbb{E}(X_n + Y_n) = \lim_{n\to\infty} \mathbb{E}(X_n) + \lim_{n\to\infty} \mathbb{E}(Y_n) = \mathbb{E}(X) + \mathbb{E}(Y)$$

(Schritt 2): Als zweiten Schritt zeigen wir $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ für integrierbare Zufallsvariable X, Y. Setzen wir $Z := X + Y$, dann ist Z integrierbar und es gilt $Z^+ - Z^- = X^+ - X^- + Y^+ - Y^-$. Schritt 1 impliziert also

$$\mathbb{E}(Z^+) + \mathbb{E}(X^-) + \mathbb{E}(Y^-) = \mathbb{E}(Z^-) + \mathbb{E}(X^+) + \mathbb{E}(Y^+)$$

woraus sich unmittelbar $\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y)$ ergibt.
(Schritt 3:) Zu zeigen, dass $\mathbb{E}(aX) = a\mathbb{E}(X)$ für integrierbares X und $a \in \mathbb{R}$ gilt, ist eine einfache Übungsaufgabe.
(Schritt 4:) Für nicht-negative Zufallsvariable X, Y mit $X \leq Y$ folgt aus (5.7) sofort $\mathbb{E}(X) \leq \mathbb{E}(Y)$. Gilt nun $X \leq Y$ für integrierbare Zufallsvariable X, Y, dann folgt offensichtlich (Skizze!) $X^+ \leq Y^+$ sowie $X^- \geq Y^-$, woraus sich mit dem Vorhergehenden unmittelbar

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \leq \mathbb{E}(Y^+) - \mathbb{E}(Y^-) = \mathbb{E}(Y)$$

ergibt. ■

Eine oft sehr nützliche, alternative Berechnungsmöglichkeit von $\mathbb{E}(X)$, insbesondere für den Fall nicht rein absolut stetiger oder rein diskreter Verteilungen, liefert das folgende Resultat:

Satz 5.15 *Sei X integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Verteilungsfunktion F, dann gilt:*

$$\mathbb{E}(X) = \int_{(0,\infty)} (1 - F(t))dt - \int_{(-\infty,0)} F(t)dt \tag{5.11}$$

Beweis: Übungsaufgabe

Erwartungswerte sind sehr flexibel gegenüber Grenzwertbildung - im Appendix finden Sie drei standard Resultate, die vielseitig anwendbar sind und das 'Rechnen' mit Erwartungswerten signifikant erleichtern.

In der Vorlesung Stochastische Modellbildung haben Sie schon die Varianz diskreter Zufallsvariable betrachtet, in Definition 3.8 die Varianz für den absolut stetigen Fall. Mit Hilfe des soeben einführten allgemeinen Begriffs des Erwartungswerts definieren wir nun:

Definition 5.16 *Sei X eine integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Dann ist die* Varianz $\mathbb{V}(X)$ von X *definiert durch*

$$\mathbb{V}(X) := \mathbb{E}\big(X - \mathbb{E}X\big)^2 \tag{5.12}$$

Die Größe $\sigma := \sqrt{\mathbb{V}(X)}$ heisst Standardabweichung *von X.*

Nachdem $\mathbb{V}(X)$ als Erwartungswert der Zufallsvariable $(X - \mathbb{E}(X))^2$ definiert ist, folgt mit Hilfe von Bemerkung 5.12, dass sich für diskrete oder absolut stetige Zufallsvariable wieder die schon bekannten Ausdrücke ergeben. Die Varianz kann auch für integrierbare Zufallsvariable $+\infty$ sein. Aufgrund der Linearität des Erwartungswerts gilt für integrierbares X

$$\mathbb{V}(X) = \mathbb{E}\big(X - \mathbb{E}X\big)^2 = \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2, \tag{5.13}$$

sowie

$$\mathbb{V}(aX + b) = \mathbb{E}\big(aX + b - a\mathbb{E}(X) - b\big)^2 = a^2\mathbb{E}(X - \mathbb{E}(X)) = a^2\mathbb{V}(X). \tag{5.14}$$

Die Varianz ist ein Maß für die Variabilität - gilt zum Beispiel $\mathbb{V}(X) = 0$, dann folgt die Existenz einer Menge $A \in \mathcal{A}$ mit $P(A) = 1$, auf der X konstant ist (Übungsaufgabe).

Beispiel 5.17 Wir berechnen $\mathbb{V}(X)$ für $X \sim \mathcal{U}(a, b)$. Wegen $X \sim \mathcal{U}(a, b)$ genau dann wenn $Z := \frac{X-a}{b-a} \sim \mathcal{U}(0, 1)$, betrachten wir zuerst Z und erhalten

$$\mathbb{V}(Z) = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Wegen $X = a + (b - a)Z$ ergibt sich unter Verwendung von (5.14) sofort $\mathbb{V}(X) = \frac{(b-a)^2}{12}$.

Definition 5.18 *Eine Zufallsvariable X auf $(\Omega, \mathcal{A}, P)$ heisst* quadratisch integrierbar *genau dann wenn $\mathbb{E}(X^2) < \infty$ (i.e. wenn X^2 integrierbar ist).*

Jede quadratisch integrierbare Zufallsvariable ist auch integrierbar und es gilt (siehe Übungen)

$$\mathbb{E}(|X|) \leq \big(\mathbb{E}(X^2)\big)^{1/2}. \tag{5.15}$$

Allgemein gilt die Hölder'sche Ungleichung:

Satz 5.19 *Seien X, Y Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $p, q \in (1, \infty)$ konjugiert (i.e. $\frac{1}{p} + \frac{1}{q} = 1$), dann gilt*

$$\mathbb{E}(|XY|) \leq \big(\mathbb{E}(|X|^p)\big)^{1/p} \big(\mathbb{E}(|Y|^q)\big)^{1/q}. \tag{5.16}$$

Im Falle, dass die rechte Seite von (5.16) endlich ist gilt Gleichheit genau dann, wenn Konstanten $\alpha, \beta \geq 0$ (nicht beide gleich 0) und eine Menge $A \in \mathcal{A}$ mit $P(A) = 1$ existieren, sodass $\alpha|X|^p(\omega) = \beta|Y|^q(\omega)$ für jedes $\omega \in A$ gilt.

Beweis: erfolgt in wesentlich allgemeinerer Form in der Maßtheorie.

Der Erwartungswert minimiert den quadratischen Abstand - es gilt folgendes einfache Resultat:

Satz 5.20 (Steiner'scher Verschiebungssatz) *Sei X eine quadratisch integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$, dann gilt:*

$$\mathbb{E}(X-a)^2 = \mathbb{V}(X) + (\mathbb{E}X - a)^2$$

Insbesondere ist also $\mathbb{E}(X-a)^2$ genau dann minimal wenn $a = \mathbb{E}(X)$.

Beweis: Unter Verwendung der Linearität des Erwartungswerts ergibt sich

$$\begin{aligned}
\mathbb{E}(X-a)^2 &= \mathbb{E}\big(X - a + \mathbb{E}X - \mathbb{E}X\big)^2 \\
&= \mathbb{E}(X-\mathbb{E}X)^2 + \mathbb{E}(a-\mathbb{E}X)^2 + 2\mathbb{E}\big((X-\mathbb{E}X)(\mathbb{E}X-a)\big) \\
&= \mathbb{E}(X-\mathbb{E}X)^2 + \mathbb{E}(a-\mathbb{E}X)^2 + 2(\mathbb{E}X-a)\underbrace{\mathbb{E}(X-\mathbb{E}X)}_{=0} \\
&= \mathbb{V}(X) + (\mathbb{E}X-a)^2.
\end{aligned}$$

Die behauptete Minimierungseigenschaft von $\mathbb{E}(X)$ folgt nun unmittelbar. ■

Sei $x_1, \ldots, x_n$ eine Stichprobe der Zufallsvariable X. Dann sind Stichprobenmittel $\overline{x}_n$ und Stichprobenvarianz s_n^2 definiert durch

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i, \quad s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x}_n)^2. \tag{5.17}$$

Den Grund dafür, in der Definition von s_n^2 durch $n-1$ und nicht durch n zu dividieren werden wir später noch genauer diskutieren - alternativ kann dem aber auch schon jetzt mit Hilfe von Simulationen (siehe Übungsaufgaben) vorgegriffen werden.

Frage 5.21 Beachten Sie, dass die Ausdrücke in (5.17) genau $\mathbb{E}(Z)$ und $\frac{n-1}{n}\mathbb{V}(Z)$ einer speziellen diskreten Zufallsvariable Z entsprechen - welche Zufallsvariable ist dies?

Die Varianz einer Zufallsvariable kann auch zur Abschätzung der Wahrscheinlichkeit dafür, dass $|X - \mathbb{E}X| \geq \varepsilon$ ist, verwendet werden:

Satz 5.22 (Tschebyscheff und Markov Ungleichung) *Sei X eine quadratisch integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$, dann gilt für jedes $\varepsilon > 0$*

$$P(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}. \tag{5.18}$$

Allgemeiner gilt für jede Zufallsvariable Y, jede (nicht notwendigerweise streng) monoton wachsende Abbildung $\varphi : [0,\infty) \to [0,\infty)$ und jedes $\varepsilon > 0$ mit $\varphi(\varepsilon) > 0$

$$P(|Y| \geq \varepsilon) \leq \frac{\mathbb{E}(\varphi \circ |Y|)}{\varphi(\varepsilon)}. \tag{5.19}$$

Beweis: Wir beweisen zuerst die allgemeinere zweite Aussage und setzen $Z := |Y|$. Für jedes $\omega \in \Omega$ gilt offensichtlich

$$\varphi(Z(\omega)) \geq \varphi(Z(\omega))\,\mathbf{1}_{[\varphi(\varepsilon),\infty)}(\varphi(Z(\omega))) \geq \varphi(\varepsilon)\mathbf{1}_{[\varepsilon,\infty)}(Z(\omega)) = \varphi(\varepsilon)\mathbf{1}_{Z^{-1}([\varepsilon,\infty))}(\omega).$$

Unter Verwendung der Monotonie und Linearität des Erwartungswerts ergibt sich damit sofort die gewünschte Markov Ungleichung

$$\mathbb{E}(\varphi \circ Z) \geq \varphi(\varepsilon)\mathbb{E}(\mathbf{1}_{Z^{-1}([\varepsilon,\infty))}) = \varphi(\varepsilon)P(Z \geq \varepsilon).$$

Um die Tschebyscheff Ungleichung zu beweisen, setzen wir einfach $\varphi(t) = t^2$ und betrachten die Zufallsvariable $Y = X - \mathbb{E}X$. ■

Seien X, Y quadratisch integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Dann gilt offensichtlich

$$\begin{aligned} \mathbb{V}(X+Y) &= \mathbb{E}\big(X+Y-\mathbb{E}(X+Y)\big)^2 = \mathbb{E}\big(X-\mathbb{E}X+Y-\mathbb{E}Y)\big)^2 \\ &= \mathbb{V}(X)+\mathbb{V}(Y)+2\mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y)). \end{aligned}$$

Definition 5.23 *Seien X, Y quadratisch integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Dann ist die* Kovarianz von (X, Y) definiert durch

$$Cov(X,Y) := \mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y))$$

Im Falle von $Cov(X,Y) = 0$ nennen wir X, Y unkorreliert. *Gilt weiters $\mathbb{V}(X), \mathbb{V}(Y) > 0$ dann heisst*

$$\rho(X,Y) := \frac{Cov(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

Korrelationskoeffizient von (X, Y).

Definition 5.24 *Wir sagen im Folgenden, eine Aussage/Eigenschaft gilt P-fast sicher (und schreiben $[P]$), genau dann wenn eine Menge $A \in \mathcal{A}$ mit $P(A) = 1$ existiert, sodass die Aussage/Eigenschaft für jedes $\omega \in A$ gilt.*

Beispiel 5.25 Wie in den Übungen gezeigt gilt für jede quadratisch integrierbare Zufallsvariable X, dass $\mathbb{V}(X) = 0$ genau dann wenn eine Konstante $a \in \mathbb{R}$ existiert, sodass $X = a$ $[P]$.

$\rho(X, Y)$ misst die *lineare Abhängigkeit der Zufallsvariablen X, Y*. Der folgende Satz fasst die wichtigsten Eigenschaften von Kovarianz und Korrelationskoeffizienten zusammen:

Satz 5.26 *Seien X, Y quadratisch integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$, dann gilt:*

1. *$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.*

2. *Im Falle $\mathbb{V}(X), \mathbb{V}(Y) > 0$ gilt $\rho(X,Y) \in [-1, 1]$.*

3. *Falls $\rho(X,Y) = 1$ dann existiert eine Konstante $a > 0$ sodass*

$$Y - \mathbb{E}(Y) = a(X - \mathbb{E}(X))\ [P].$$

4. *Falls* $\rho(X,Y) = -1$ *dann existiert eine Konstante* $a < 0$ *sodass*

$$Y - \mathbb{E}(Y) = a(X - \mathbb{E}(X))\ [P].$$

5. *Für* $a, c > 0$ *und* $b, d \in \mathbb{R}$ *gilt* $\rho(aX + b, cY + d) = \rho(X,Y)$.

Beweis: Der erste Punkt ist klar (Linearität des Erwartungswerts).
Der zweite Punkte ist eine direkte Folgerung aus der Hölderschen Ungleichung für den Fall $p = q = 2$. Zum Beweis des dritten Punkts kann wie folgt vorgegangen werden: Wegen $\mathbb{V}(X), \mathbb{V}(Y) > 0$ sind die Zufallsvariablen $X^\star = \frac{X-\mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}}$ und $Y^\star = \frac{Y-\mathbb{E}(Y)}{\sqrt{\mathbb{V}(Y)}}$ wohldefiniert und es folgt (Linearität des Erwartungswerts)

$$0 \leq \mathbb{E}(X^\star - Y^\star)^2 = \mathbb{V}(X^\star - Y^\star) = \underbrace{\mathbb{V}(X^\star)}_{=1} + \underbrace{\mathbb{V}(Y^\star)}_{=1} - 2\underbrace{Cov(X^\star, Y^\star)}_{=\rho(X,Y)}.$$

Gilt nun $\rho(X,Y) = 1$ dann folgt sofort $0 \leq \mathbb{E}(X^\star - Y^\star)^2 = 1 + 1 - 2 = 0$, also $X^\star - Y^\star = 0\,[P]$ und damit wie behauptet $\frac{X-\mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}} = \frac{Y-\mathbb{E}(Y)}{\sqrt{\mathbb{V}(Y)}}\ [P]$.
Der Fall $\rho = -1$ kann analog behandelt werden. Der letzte Punkt ist eine unmittelbare Folgerung aus der Linearität des Erwartungswerts und Gleichung (5.14). ■

Gegeben sei eine Strichprobe $(x_1, y_1), \ldots (x_n, y_n)$ von (X,Y). Dann ist der *empirische Korrelationskoeffizient* ρ_n definiert durch

$$\rho_n = \frac{\sum_{i=1}^n (x_i - \overline{x}_n)(y_i - \overline{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \overline{x}_n)^2}\sqrt{\sum_{i=1}^n (y_i - \overline{y}_n)^2}}. \tag{5.20}$$

Beachten Sie, dass dies genau dem Korrelationskoeffizienten des diskret gleichverteilten Zufallsvektors (X,Y) auf $\{(x_1, y_1), \ldots (x_n, y_n)\}$ entspricht.

Für unkorrelierte, quadratisch integrierbare Zufallsvariable $X_1, \ldots, X_n$ ist die Varianz additiv - es gilt folgendes Resultat:

Satz 5.27 *Seien* $X_1, \ldots, X_n$ *quadratisch integrierbare Zufallsvariable und* $(a_1, \ldots, a_n) \in \mathbb{R}^n$. *Setzen wir* $\sigma_{ij} := Cov(X_i, X_j)$ *für jedes Paar* $(i,j) \in \{1, \ldots, n\}^2$, *dann ist die Matrix* $\sigma = (\sigma_{ij})$ *positiv semidefinit.*
Falls $\sigma_{ij} = 0$ *für* $i \neq j$, *dann gilt die folgende Gleichheit:*

$$\mathbb{V}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2\, \mathbb{V}(X_i)$$

Mit anderen Worten: Wenn $X_1, \ldots, X_n$ *paarweise unkorreliert sind, dann ist die Varianz additiv*

Beweis: Die Linearität des Erwartungswerts impliziert

$$\begin{aligned} 0 \leq \mathbb{V}\left(\sum_{i=1}^n a_i X_i\right) &= \mathbb{E}\left(\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i \mathbb{E}(X_i)\right)^2 = \mathbb{E}\left(\sum_{i=1}^n a_i (X_i - \mathbb{E}(X_i))\right)^2 \\ &= \mathbb{E}\left(\sum_{i,j=1}^n a_i\, a_j (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) = \sum_{i,j=1}^n a_i\, a_j\, \sigma_{ij} \end{aligned}$$

woraus sich beide Behauptungen unmittelbar ergeben. ■

Mit Hilfe der bisher behandelten Werkzeuge können wir schon die folgende Version des *schwachen Gesetzes der Großen Zahlen* beweisen:

Satz 5.28 (WLLN) *Sei $X_1, X_2, \ldots$ eine Folge quadratisch integrierbarer, paarweise unkorrelierter Zufallsvariable mit*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}(X_i) = 0.$$

Dann gilt für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i)\Big| \geq \varepsilon\Big) = 0. \tag{5.21}$$

Folgerung 5.29 *Sei $X_1, X_2, \ldots$ eine Folge quadratisch integrierbarer, paarweise unkorrelierter, identisch verteilter Zufallsvariable. Dann gilt für jedes $\varepsilon > 0$*

$$\lim_{n\to\infty} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X)\Big| \geq \varepsilon\Big) = 0.$$

Beweis von Satz 5.28: Wir definieren eine neue Zufallsvariable Y_n durch

$$Y_n := \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)).$$

Dann gilt $\mathbb{E}(Y_n) = 0$ und es folgt mit Satz 5.27

$$\mathbb{V}(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i).$$

Damit liefert die Tschebysheff'sche Ungleichung für $\varepsilon > 0$ sofort

$$P\big(|Y_n| \geq \varepsilon\big) \leq \frac{1}{\varepsilon^2}\mathbb{V}(Y_n) = \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) \longrightarrow 0 \quad \text{für } n \to \infty. \; \blacksquare$$

Beispiel 5.30 Wir berechnen Kovarianz und Korrelationskoeffizient eines absolut stetigen Vektors $X = (X_1, X_2)$ mit Dichte (stetige Gleichverteilung auf $[0,1]^2$)

$$f(x,y) = \mathbf{1}_{[0,1]^2}(x,y).$$

Sei $x, y \in [0,1]$. Aus Beispiel 3.5 (setze $\theta = 0$) wissen wir schon gilt $X_1, X_2 \sim \mathcal{U}(0,1)$. Insbesondere ergibt sich also $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \frac{1}{2}$ sowie $\mathbb{V}(X_1) = \mathbb{V}(X_2) = \frac{1}{12}$. Für $Cov(X_1, X_2)$ erhalten wir schließlich

$$\begin{aligned} Cov(X_1, X_2) &= \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \int_{\mathbb{R}^2} x_1x_2 f(x_1,x_2)dx_1dx_2 - \frac{1}{4} \\ &= \int_{[0,1]^2} x_1x_2 dx_1dx_2 - \frac{1}{4} = 0. \end{aligned}$$

X_1 und X_2 sind also unkorreliert und wir erhalten $\rho(X_1, X_2) = 0$.

Wie vorhin bemerkt, misst der Korrelationskoeffizient nur lineare Abhängigkeit - er kann aber nicht als generelles (Un)Abhängigkeitsmaß verwendet werden. Wir beschließen diesen Abschnitt mit einem Beispiel, das genau diese Tatsache veranschaulicht.

Beispiel 5.31 Wir betrachten $a \in [0,1]$ und die Funktion T_a definiert durch

$$T_a(x) = \begin{cases} a - x & \text{für } x \leq a, \\ x & \text{für } x > a. \end{cases}$$

Weiters sei $X \sim \mathcal{U}(0,1)$ und $Y = T_a \circ X$, i.e. Y ist eine deterministische Funktion von X (genaues Gegenteil von Unabhängigkeit - wenn X bekannt dann auch Y). Es lässt sich leicht nachrechnen, dass $T_a \circ X \sim \mathcal{U}(0,1)$ und wir erhalten

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}(X(T_a \circ X)) = \int_{\mathbb{R}} xT_a(x)f(x)dx = \int_{[0,1]} xT_a(x)dx \\ &= \int_{[0,a]} x(a-x)dx + \int_{[a,1]} x^2 dx = \frac{1}{3} - \frac{a^3}{6}. \end{aligned}$$

Insgesamt ergibt sich also $Cov(X,Y) = \frac{1}{3} - \frac{a^3}{6} - \frac{1}{4}$. Für den Fall $a = \frac{1}{\sqrt[3]{2}}$ sind die Zufallsvariablen X, Y (trotz der extremen Abhängigkeit) sogar unkorreliert.

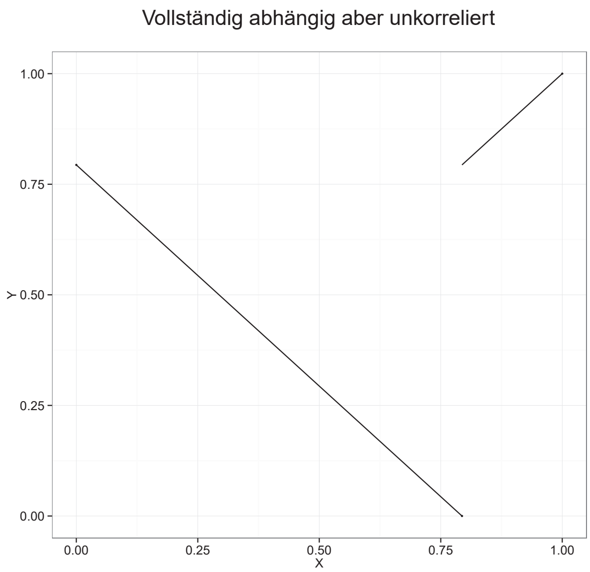

Abbildung 5.1: Transformation T_a von Beispiel 5.31 für $a = \frac{1}{\sqrt[3]{2}}$.

Kapitel 6

Unabhängigkeit

6.1 Unabhängige Mengensysteme und Zufallsvariable

Sie haben in der Vorlesung 'Stochastische Modellbildung' schon den Begriff der Unabängigkeit von Ereignissen kennengelernt: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$. Dann sind A und B unabhängig (per definitionem) genau dann, wenn

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

Im Falle von $P(B) > 0$ bedeutet dies

$$P(A|B) := \frac{P(A \cap B)}{P(B)} = P(A),$$

das Eintreten von B ändert also nicht die Eintrittsswahrscheinlichkeit von A.
Allgemein heißt eine Familie $(A_i)_{i \in I}$ (I beliebige Indexmenge) mit $A_i \in \mathcal{A}$ für jedes $i \in I$ *unabhängig* genau dann wenn für jede endliche (nichtleere) Teilmenge $\{i_1, \ldots, i_n\} \subseteq I$ die folgende Gleichheit gilt:

$$P\Big(\bigcap_{k=1}^{n} A_{i_k}\Big) = \prod_{k=1}^{n} P(A_{i_k}) \tag{6.1}$$

Wie das folgende einfache Beispiel zeigt, folgt aus der paarweisen Unabhängigkeit von Mengen $A_1, \ldots, A_n \in \mathcal{A}$ im Allgemeinen nicht die Unabhängigkeit der Familie $(A_i)_{i=1}^n$:

Beispiel 6.1 Wir betrachten den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit

$$\Omega := \{1, 2, \ldots, 6\}^2, \quad \mathcal{A} := \mathfrak{p}(\Omega), \quad P(\{(i,j)\}) = \frac{1}{36} \text{ für jedes Paar } (i,j) \in \Omega$$

als Modell für zweimaliges Würfeln mit einem fairen Würfel und die folgenden drei Ereignisse:

$$\begin{aligned} A_1 &= \{(i,j) \in \Omega : i \text{ ungerade}\}, \quad A_2 = \{(i,j) \in \Omega : j \text{ ungerade}\}, \\ A_3 &= \{(i,j) \in \Omega : i + j \text{ ungerade}\} \end{aligned}$$

Dann gilt offensichtlich $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$ sowie $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}$, die Familie $(A_i)_{i=1}^3$ ist also paarweise unabhängig. Wegen $A_1 \cap A_2 \cap A_3 = \emptyset$ und $P(A_1)P(A_2)P(A_3) = \frac{1}{8} > 0$ ist sie aber nicht unabhängig.

Wir erweitern nun den Begriff der Unabhängigkeit in naheliegender Weise auf Mengensysteme:

Definition 6.2 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge und $\mathcal{E}_i \subseteq \mathcal{A}$ ein System von Mengen in $\mathcal{A}$ für jedes $i \in I$. Dann heisst die Familie $(\mathcal{E}_i)_{i\in I}$* unabhängig *genau dann wenn für jede nichtleere endliche Teilmenge $\{i_1, \ldots, i_n\} \subseteq I$ und beliebige Mengen $E_{i_1} \in \mathcal{E}_{i_1}, E_{i_2} \in \mathcal{E}_{i_2}, \ldots, E_{i_n} \in \mathcal{E}_{i_n}$ die Gleichheit*

$$P\Big(\bigcap_{k=1}^{n} E_{i_k}\Big) = \prod_{k=1}^{n} P(E_{i_k}) \tag{6.2}$$

gilt.

Beachten Sie, dass diese Definition für den Fall $\mathcal{E}_i = \{A_i\}$ für alle $i \in I$ genau der vorhin erwähnten Definition für Familien $(A_i)_{i\in I}$ entspricht.

Aus der Unabängigkeit von $A, B \in \mathcal{A}$ folgt sofort die Unabängigkeit von A^c, B^c (und umgekehrt) - es gilt nämlich

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) = 1 - P(A \cup B) = 1 - \big(P(A) + P(B) - P(A \cap B)\big) \\ &= (1 - P(A))(1 - P(B)) = P(A^c)P(B^c). \end{aligned}$$

Unabhängigkeit von A^c, B sowie Unabhängigkeit von A, B^c lässt sich analog nachrechnen. Insbesondere impliziert also die Unabhängigkeit von A und B auch die Unabhängigkeit der erzeugten σ-Algebren $\mathcal{A}_\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$ und $\mathcal{A}_\sigma(\{B\}) = \{\emptyset, \Omega, B, B^c\}$ gemäß Definition 6.2. Damit ist das folgende wichtige Resultat nicht ganz überraschend (stellen Sie sich der Einfachheit halber den Fall $I = \mathbb{N}$ oder I endlich vor):

Satz 6.3 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge, jedes $\mathcal{E}_i \subseteq \mathcal{A}$ durchschnittsstabil, und die Familie $(\mathcal{E}_i)_{i\in I}$ unabhängig. Dann ist auch die Familie $(\mathcal{A}_\sigma(\mathcal{E}_i))_{i\in I}$ unabhängig.*

Beweis: Mit Methoden der Maßtheorie.

Das folgende Resultat wird sich im Zusammenhang mit der Unabhängigkeit von Zufallsvariablen als sehr nützlich erweisen (stellen Sie sich der Einfachheit halber $I = \mathbb{N}$ oder I endlich vor):

Satz 6.4 (Zusammenfassen unabhängiger σ-Algebren) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge, jedes $\mathcal{E}_i \subseteq \mathcal{A}$ durchschnittsstabil, und die Familie $(\mathcal{E}_i)_{i\in I}$ unabhängig. Weiters sei $(I_j)_{j\in J}$ eine Zerlegung von I in paarweise disjunkte Mengen I_j und für jedes $j \in J$ definiert durch*

$$\mathcal{A}_j := \mathcal{A}_\sigma\Big(\bigcup_{i\in I_j} \mathcal{E}_i\Big).$$

Dann ist auch die Familie $(\mathcal{A}_j)_{j\in J}$ unabhängig.

Beweis: Für jedes $j \in J$ setzen wir

$$\tilde{\mathcal{E}}_j := \Big\{ \bigcap_{i \in K} A_i : K \subseteq I_j \text{ endlich und } A_i \in \mathcal{E}_i \ \forall i \in K \Big\}.$$

Dann ist $\tilde{\mathcal{E}}_j$ durchschnittsstabil und wegen $\bigcup_{i \in I_j} \mathcal{E}_i \subseteq \tilde{\mathcal{E}}_j \subseteq \mathcal{A}_j$ folgt

$$\mathcal{A}_j := \mathcal{A}_\sigma\Big(\bigcup_{i \in I_j} \mathcal{E}_i \Big) \subseteq \mathcal{A}_\sigma(\tilde{\mathcal{E}}_j) \subseteq \mathcal{A}_\sigma(\mathcal{A}_j) = \mathcal{A}_j$$

und damit $\mathcal{A}_\sigma(\tilde{\mathcal{E}}_j) = \mathcal{A}_j$. Nach Konstruktion ist die Familie $(\tilde{\mathcal{E}}_j)_{j \in J}$ unabhängig, Anwendung von Satz 6.3 liefert das gewünschte Resultat. ■

Wir erweitern nun den Begriff der Unabhängigkeit auf Zufallsvariable. Naheliegenderweise nennen wir (als ersten Schritt) Zufallsvariable $X_1, \ldots, X_n$ auf $(\Omega, \mathcal{A}, P)$ unabhängig, wenn für beliebige Borel Mengen $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$ die folgende Gleichheit gilt

$$P\Big(\bigcap_{k=1}^{n} \underbrace{X_i^{-1}(B_i)}_{=:A_i \in \mathcal{A}} \Big) = \prod_{k=1}^{n} P(\underbrace{X_i^{-1}(B_i)}_{=:A_i \in \mathcal{A}}), \tag{6.3}$$

i.e. wenn also die entsprechenden vollständigen Urbilder A_i unabhängig sind. Beachten Sie, dass wir obige Gleichung mit $X := (X_1, \ldots, X_n)$ auch wie folgt schreiben können:

$$P^X\Big(\bigtimes_{i=1}^{n} B_i \Big) = \prod_{k=1}^{n} P^{X_i}(B_i)$$

Vom Beweis von Satz 2.2 (Gleichung 2.3) wissen wir das folgende: Ist X eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ dann ist auch $\mathcal{A}_\sigma(X) := X^{-1}(\mathcal{B}(\mathbb{R})) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ eine σ-Algebra und es gilt $X^{-1}(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{A}$. Wir nennen $\mathcal{A}_\sigma(X)$ im Folgenden auch *die von X erzeugte σ-Algebra* und machen folgende Definition.

Definition 6.5 *Eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ heißt* unabängig *genau dann, wenn die Familie $(\mathcal{A}_\sigma(X_i))_{i \in I}$ der erzeugten σ-Algebren unabhängig ist.*

Um Unabhängigkeit von $(X_i)_{i \in I}$ zu zeigen reicht es aus, Gleichung (6.3) für Mengen aus durchschnittsstabilen Erzeugern zu überprüfen - durchschnittsstabile Erzeuger von $\mathcal{B}(\mathbb{R})$ haben wir schon in Lemma 1.14 kennengelernt (welche der 12 Erzeuger sind wirklich durchschnittsstabil ?).

Satz 6.6 *Seien $X_1, \ldots, X_n$ Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $\mathcal{E}$ ein durchschnittsstabiler Erzeuger von $\mathcal{B}(\mathbb{R})$. Dann ist die Familie $(X_i)_{i=1}^n$ unabhängig genau dann wenn*

$$P\Big(\bigcap_{k=1}^{n} X_i^{-1}(B_i) \Big) = \prod_{k=1}^{n} P(X_i^{-1}(B_i)) \tag{6.4}$$

für jede Auswahl von Mengen $B_1, \ldots, B_n \in \mathcal{E}$ gilt.

Beweis: Gemäß Gleichung (2.3) gilt $\mathcal{A}_\sigma(X^{-1}(\mathcal{E})) = X^{-1}(\mathcal{A}_\sigma(\mathcal{E}))$ für jede Zufallsvariable X. Ist also $\mathcal{E}$ durchschnittsstabiler Erzeuger von $\mathcal{B}(\mathbb{R})$ dann ist $X^{-1}(\mathcal{E})$ durchschnittsstabiler Erzeuger von $\mathcal{A}_\sigma(X)$. Setzen wir $\mathcal{A}_i := \mathcal{A}_\sigma(X_i)$ für $i \in \{1,\dots,n\}$ dann ist also $\mathcal{E}_i = X_i^{-1}(\mathcal{B}(\mathbb{R}))$ durchschnittsstabiler Erzeuger von $\mathcal{A}_i$. Anwendung von Satz 6.3 liefert die gewünschte Behauptung. ■

Folgerung 6.7 *Seien $X_1,\dots,X_n$ Zufallsvariable auf $(\Omega,\mathcal{A},P)$ und $X = (X_1,\dots,X_n)$. Dann gilt: $(X_i)_{i=1}^n$ ist unabhängig genau dann wenn die gemeinsame Verteilungsfunktion das Produkt der einzelnen Verteilungsfunktionen ist, i.e. wenn für alle $x_1,\dots,x_n \in \mathbb{R}$ gilt:*

$$F_X(x_1,\dots,x_n) = \prod_{k=1}^{n} F_{X_k}(x_k) \tag{6.5}$$

Beweis: Wir betrachten $\mathcal{E} := \mathcal{E}_4 = \{(-\infty,b] : b \in \mathbb{R}\}$. Dann ist $\mathcal{E}$ offensichtlich ein durchschnittsstabiler Erzeuger von $\mathcal{B}(\mathbb{R})$ und das Resultat folgt sofort aus Satz 6.6. ■

Beispiel 6.8 Unabhängigkeit kann auch in etwas überraschenden Situationen auftreten. Wir betrachten den Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P) = ([0,1),\mathcal{B}([0,1)]),\lambda)$. Aus der Analysis wissen Sie, dass alle bis auf abzählbar viele $w \in [0,1)$ eine eindeutige Dezimaldarstellung

$$\omega = 0.x_1x_2x_3\dots$$

haben - für die abzählbar vielen Ausnahmen wählen wir jene Darstellung ohne Periode 9. Mit dieser Zuordnung bezeichne $X_i(\omega) \in \{0,\dots,9\}$ die i-te Stelle der Dezimalentwicklung. Es gilt dann für jedes $n \in \mathbb{N}$ und $j \in \{0,\dots,9\}$

$$X_n^{-1}(\{j\}) = \bigcup_{i_1,i_2,\dots i_{n-1}=0}^{9} \Big[\sum_{k=1}^{n-1}\frac{i_k}{10^k} + \frac{j}{10^n}, \sum_{k=1}^{n-1}\frac{i_k}{10^k} + \frac{j+1}{10^n}\Big).$$

Insbesondere ist also jedes X_i Zufallsvariable und es lässt sich zeigen, dass $(X_i)_{i\in\mathbb{N}}$ unabhängig ist (Übungsaufgabe).

Für den absolut stetigen Fall ist auch eine einfache Charakterisierung der Abhängigkeit direkt über die Wahrscheinlichkeitsdichten möglich - wir beginnen mit folgender Beobachtung: Angenommen $X = (X_1,\dots,X_n)$ ist absolut stetig mit Dichte f. Sei B_1 eine Borel Menge, dann gilt offensichtlich

$$\begin{aligned} P^{X_1}(B_1) = P^X(B_1 \times \mathbb{R} \times \ldots \times \mathbb{R}) &= \int_{B_1\times\mathbb{R}\times\ldots\times\mathbb{R}} f(x_1,\dots x_n)\,dx_1\dots dx_n \\ &= \int_{B_1} \underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1,\dots x_n)\,dx_2\dots dx_n}_{:=f_1(x_1)}\,dx_1, \end{aligned}$$

die Zufallsvariable X_1 ist also auch absolut stetig mit (Rand-) Dichte f_1. Für $X_2,\dots X_n$ kann analog vorgegangen werden um zu zeigen, dass jedes X_i ebenfalls absolut stetig mit Dichte f_i ist. Damit lässt sich folgendes Resultat zeigen:

Satz 6.9 *Sei* $X = (X_1, \dots, X_n)$ *absolut stetig mit Dichte* $f : \mathbb{R}^n \to [0, \infty)$ *und bezeichne* $f_i : \mathbb{R} \to [0, \infty)$ *die (Rand-) Dichte von* X_i*. Dann gilt:* $(X_i)_{i=1}^n$ *ist unabhängig genau dann wenn die gemeinsame Dichte das Produkt der Randdichten ist, i.e. wenn für ein* $A \in \mathcal{B}(\mathbb{R}^n)$ *mit* $\lambda_n(A) = 0$ *und alle* $(x_1, \dots, x_n) \in A^c$ *gilt:*

$$f(x_1, \dots, x_n) = \prod_{k=1}^{n} f_k(x_k) \tag{6.6}$$

Beweis: Für einen sauberen Beweis benötigten wir einige zusätzliche Tatsachen aus der Maßtheorie. Wichtig ist aber zumindest zu verstehen, warum die Gleichheit nicht überall bestehen muss/kann - was passiert wenn eine Wahrscheinlichkeitsdichte f in genau endlich vielen Punkte (oder in einer λ-Nullmenge) ändern - ändert das auch das Wahrscheinlichkeitsmaß und die Verteilung der Zufallsvariable ?

Bemerkung 6.10 Für die Praxis wichtig ist folgende Situation: Sei $f_i : \mathbb{R} \to [0, \infty)$ eine Wahrscheinlichkeitsdichte für jedes $i \in \{1, \dots, n\}$. Dann ist f, definiert durch

$$f(x_1, \dots, x_n) := \prod_{k=1}^{n} f_k(x_k)$$

eine Wahrscheinlichkeitsdichte auf $\mathbb{R}^n$ und für $X = (X_1, \dots, X_n) \sim \mu_f$ ist $(X_i)_{i=1}^n$ unabhängig.

Beispiel 6.11 Wir betrachten die in Beispiel 5.30 erwähnte Wahrscheinlichkeitsdichte f, definiert durch

$$f(x, y) = \mathbf{1}_{[0,1]^2}(x, y)$$

und $(X, Y) \sim \mu_f$. Wie in Beispiel 5.30 gezeigt gilt $X \sim \mathcal{U}(0, 1)$ und $Y \sim \mathcal{U}(0, 1)$. Die Randdichten f_1, f_2 sind also gegeben durch $f_1(x) = f_2(x) = \mathbf{1}_{[0,1]}(x)$ und wir erhalten $f(x, y) = f_1(x) f_2(y)$ für alle $x, y \in \mathbb{R}$. Die Zufallsvariable X und Y sind also unabhängig.

Nachdem unabhängige σ-Algebren ohne Verlust der Unabhängigkeit zusammengefasst werden können (siehe Satz 6.4) und Unabhängigkeit von Zufallsvariablen über die Unabhängigkeit der erzeugten σ-Algebren definiert ist, ist folgendes Resultat wenig überraschend (stellen Sie sich der Einfachheit halber wieder den Fall $I = \mathbb{N}$ oder I endlich vor):

Satz 6.12 *Sei* I *eine beliebige Indexmenge und* $(X_i)_{i \in I}$ *eine unabhängige Familie von Zufallsvariablen auf* $(\Omega, \mathcal{A}, P)$*. Weiters sei* $(I_j)_{j \in J}$ *eine Zerlegung von* I *in paarweise disjunkte Mengen* $I_j := \{i_1^j, i_2^j, \dots i_{m_j}^j\}$ *und* $g_j : \mathbb{R}^{m_j} \to \mathbb{R}$ *Borel messbar. Dann ist auch die Familie* $(g_j(X_{i_1^j}, X_{i_2^j}, , \dots X_{i_{m_j}^j}))_{j \in J}$ *unabhängig.*

Beweisskizze: Anwenden von Satz 6.4.

Beispiel 6.13 Seien X_1, X_2, X_3 unabhängige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Dann sind beispielsweise auch $Y := X_1 + X_2$ und $Z := X_3^2$ unabhängig. Selbiges gilt für $Y = g_1(X_1, X_3)$ und $Z = g_2(X_2)$ falls g_1, g_2 Borel messbar (also insbesondere für stetige g_1, g_2).

Wir beschließen diesen Abschnitt mit dem folgenden wichtigen Resultat:

Satz 6.14 *Seien X, Y unabhängige, integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Dann gilt*

$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y), \tag{6.7}$$

also inbesondere $Cov(X, Y) = 0$.
Allgemeiner gilt für unabhängige, integrierbare Zufallsvariable $(X_i)_{i=1}^n$ die Gleichheit

$$\mathbb{E}\Big(\prod_{i=1}^n X_i\Big) = \prod_{i=1}^n \mathbb{E}(X_i). \tag{6.8}$$

Beweis: In gewohnter Manier: zuerst für einfache Zufallsvariablen, dann erweitern auf den allgemeinen Fall (Übungsaufgabe).

6.2 Faltung von Zufallsvariablen

Wir betrachten in diesem kurzen Abschnitt, wie sich die Verteilung von $X + Y$ berechnen lässt wenn X und Y unabhängig sind, und starten mit einem einfachen Beispiel:

Beispiel 6.15 Sei $X \sim Pois(\lambda), Y \sim Pois(\mu)$ und X und Y unabhängig. Dann gilt offensichtlich für jedes $n \in \mathbb{N}_0$

$$\begin{aligned} P(X+Y=n) &= \sum_{k\in\mathbb{N}_0} P(X+Y=n,\, X=k) = \sum_{k=0}^n P(Y=n-k,\, X=k) \\ &= \sum_{k=0}^n P(Y=n-k)\,P(X=k) = \sum_{k=0}^n \frac{\mu^{n-k}}{(n-k)!}e^{-\mu}\,\frac{\lambda^k}{k!}e^{-\lambda} \\ &= e^{-(\mu+\lambda)}\frac{1}{n!}\sum_{k=0}^n \binom{n}{k}\mu^{n-k}\,\lambda^k = \frac{1}{n!}(\mu+\lambda)^n e^{-(\lambda+\mu)}. \end{aligned}$$

Die Zufallsvariable $X + Y$ ist also wieder Poisson-verteilt und zwar mit Parameter $\lambda + \mu$.

Zum Beweis des folgendes Resultats kann vollkommen analog vorgegangen werden.

Satz 6.16 *Seien X, Y unabhängige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ mit Werten in $\mathbb{Z}$. Setzen wir $p_n := P(X = n) = P^X(\{n\})$, $q_n = P(Y = n) = P^Y(\{n\})$ und $r_n = P(X + Y = n) = P^{X+Y}(\{n\})$ für jedes $n \in \mathbb{Z}$, dann gilt:*

$$r_n = \sum_{k\in\mathbb{Z}} p_k\, q_{n-k} = \sum_{k\in\mathbb{Z}} p_{n-k}\, q_k \tag{6.9}$$

Beispiel 6.17 Für $X_1, \dots, X_n \sim Bin(1, p)$ und unabhängig lässt sich unschwer nachrechnen, dass $\sum_{i=1}^n X_i \sim Bin(n, p)$ gilt (einfache Übungsaufgabe). Im einfachsten Fall $n = 2$ erhalten wir zum Beispiel

$$r_0 = p_0\, q_0 = (1-p)^2,\ r_1 = \sum_{k=0}^1 p_k\, q_{1-k} = 2p(1-p),\ r_2 = p_1\, q_1 = p^2.$$

Wir wechseln nun zum absolut stetigen Fall:

Satz 6.18 *Seien X, Y absolut stetige, unabhängige Zufallsvariable mit Dichten f bzw. g. Dann ist auch $X+Y$ absolut stetig mit Dichte $f * g$, wobei*

$$(f * g)(t) = \int_{\mathbb{R}} f(x)g(t-x)\,dx = \int_{\mathbb{R}} f(t-x)g(x)\,dx. \tag{6.10}$$

Wir nennen im Folgenden $f * g$ auch die Faltung von f und g.

Beweis: Nachdem X und Y unabhängig sind, erfüllt die Dichte h von (X, Y) die Gleichung $h(x, y) = f(x)g(y)$. Weiters gilt offensichtlich für jedes $t \in \mathbb{R}$ unter Verwendung von $T(x, y) := x + y$

$$\begin{aligned} F_{X+Y}(t) &= P(X+Y \le t) = \mathbb{E}\big(\mathbf{1}_{(-\infty,t]}(T(X,Y))\big) = \int_{\mathbb{R}^2} \mathbf{1}_{(-\infty,t]}(x+y)h(x,y)dy\,dx \\ &= \int_{\mathbb{R}} f(x) \int_{(-\infty,\, t-x]} g(y)dy\,dx = \int_{\mathbb{R}} f(x) \int_{(-\infty,\, t]} g(y-x)dy\,dx = \\ &= \int_{(-\infty,\, t]} \int_{\mathbb{R}} f(x)g(y-x)dx\,dy, \end{aligned}$$

woraus das gewünschte Resultat unmittelbar folgt. ■

Beispiel 6.19 Seien $X, Y \sim \mathcal{U}(0,1)$ und unabhängig. Für die Dichte von $X+Y$ ergibt sich

$$(f * g)(t) = \int_{\mathbb{R}} f(x)g(t-x)\,dx = \int_{[0,1]} \mathbf{1}_{[0,1]}(t-x)\,dx,$$

und damit

$$(f * g)(t) = \begin{cases} t & \text{für } t \in [0,1] \\ 2-t & \text{für } t \in [1,2]. \end{cases}$$

Wir beschliessen den Abschnitt mit einem wichtigen Resultat betreffend die Faltung von Normalverteilungen (weitere Beispiele werden in den Übungen betrachtet):

Satz 6.20 *Sei $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ und X, Y unabhängig. Dann ist auch $X+Y$ normalverteilt und es gilt $X+Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.*

Beweis: Übungsaufgabe.

Folgerung 6.21 *Seien $X_1, X_2, \ldots X_n \sim \mathcal{N}(\mu, \sigma^2)$ und unabhängig. Dann gilt:*

1. $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
2. $\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$

Beweis: Direkte Folgerung aus Satz 6.20 und den Rechenregel für Erwartungswert und Varianz.

Bemerkung 6.22 Für $P^X = \mu, P^Y = \nu$ und X, Y unabhängig, nennt man die Verteilung von $X+Y$ *Faltung von μ und ν* und schreibt $P^{X+Y} = \mu * \nu$.