Notizen zu

Einführung in die Stochastik

bei Prof. Kaiser im WS 17/18

author	Maximilian Reif <reifmaxi@fim.uni-passau.de>
last change	9. Feb 2018
version	0.8.4
github	https://github.com/lordreif/stochastik

Disclaimer

Dies ist eine Kurzzusammenfassung des Stoffes der Vorlesung *Einführung in die Stochastik* im Wintersemester 17/18 an der Universität Passau.

Das Dokument erhebt nicht den Anspruch, vollständig oder in sich geschlossen zu sein, und verzichtet gänzlich auf Beweise.

Falls Sie Fehler in Layout und Sprache finden, oder sich falsche Behauptungen eingeschlichen haben, teilen Sie dies bitte dem Autor mit.

Inhaltsverzeichnis

1 Basics

- *Ergebnis:* Element aus Ω, *Ereignis:* Element aus $\mathcal{A}$ (Teilmenge von Ω)
- *σ-Algebra* $\mathcal{A} \subset \mathfrak{P}(\Omega)$
  1. $\Omega \in \mathcal{A}$
  2. $A \in \mathcal{A} \Rightarrow A^{\mathrm{C}} \in \mathcal{A}$
  3. $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
- *Wahrscheinlichkeitsmaß:* $\mathrm{P} : \mathcal{A} \to [0,1]$
  1. $\mathrm{P}(\Omega) = 1$
  2. $A_1, A_2, \ldots \in \mathcal{A}$ p.d. $\Rightarrow \mathrm{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)$ *(σ-Additivität)*
- $A, B \in \mathcal{A}$:
  1. $A \subset B \Rightarrow \mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(B \backslash A)$
  2. $A \subset B \Rightarrow \mathrm{P}(A) \leq \mathrm{P}(B)$ *(Monotonie)*
  3. $\mathrm{P}(A^{\mathrm{C}}) = 1 - \mathrm{P}(A)$
  4. $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
- $A_1, A_2, \ldots \in \mathcal{A}$
  1. $\mathrm{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathrm{P}(A_i)$ *(σ-Subadditivität)*
  2. $A_1 \subset A_2 \subset \ldots \Rightarrow \mathrm{P}(\bigcup_{i=1}^{\infty} A_i) = \lim\limits_{i \to \infty} \mathrm{P}(A_i)$ *(σ-Stetigkeit von unten)*
  3. $A_1 \supset A_2 \supset \ldots \Rightarrow \mathrm{P}(\bigcap_{i=1}^{\infty} A_i) = \lim\limits_{i \to \infty} \mathrm{P}(A_i)$ *(σ-Stetigkeit von oben)*
- $\mathrm{P}(B) > 0$, dann $\mathrm{P}(\,\cdot \mid B) := \frac{\mathrm{P}(\,\cdot\, \cap B)}{\mathrm{P}(B)}$ W'Maß auf $\mathcal{A}$ mit $\mathrm{P}(B \mid B) = 1$
- **Formel der totalen Wahrscheinlichkeit**
  Für $B_1, B_2, \ldots, B_n \in \mathcal{A}$ p.d. mit $\mathrm{P}(B_i) > 0$ für alle $i \in \{1, \ldots, n\}$ und $\bigcup_{i=1}^{n} B_i = \Omega$ gilt:
  $$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(A \mid B_i) \cdot \mathrm{P}(B_i) \text{ für alle } A \in \mathcal{A}$$
- **Formel von Bayes**
  Für $B_1, B_2, \ldots, B_n \in \mathcal{A}$ p.d. mit $\mathrm{P}(B_i) > 0$ für alle $i \in \{1, \ldots, n\}$ und $\bigcup_{i=1}^{n} B_i = \Omega$ gilt:
  $$\mathrm{P}(B_i \mid A) = \frac{\mathrm{P}(A \mid B_i) \cdot \mathrm{P}(B_i)}{\sum\limits_{j=1}^{n} \mathrm{P}(A \mid B_j) \cdot \mathrm{P}(B_j)} \text{ für alle } A \in \mathcal{A} \text{ mit } \mathrm{P}(A) > 0$$
- $\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A|B) \cdot \mathrm{P}(B)}{\mathrm{P}(A)}$ für $A, B \in \mathcal{A}$ mit $\mathrm{P}(A), \mathrm{P}(B) > 0$

- $\mathrm{P}(A \cap B) = \mathrm{P}(A \mid B) \cdot \mathrm{P}(B)$ für $B \in \mathcal{A}$ mit $\mathrm{P}(B) > 0$
- $A_1, \ldots, A_n \in \mathcal{A}$ mit $\mathrm{P}(\bigcap_{i=1}^{n-1} A_i) > 0$, dann
  $\mathrm{P}(\bigcap_{i=1}^{n} A_i) = P(A_1) \cdot \mathrm{P}(A_2 \mid A_1) \cdots \mathrm{P}(A_1 \cap \ldots \cap A_n) \cdot \mathrm{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1})$
- A, B unabhängig $:\Leftrightarrow \mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B) \Leftrightarrow P(A \mid B) = P(A)$ falls $\mathrm{P}(B) > 0$
- $A_1, \ldots, A_n$ unabhängig $\Leftrightarrow \mathrm{P}(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i) \Rightarrow$ paarweise Unabhängigkeit
- $\binom{n}{k} := \frac{n!}{k!(n-k)!}$, $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$, $\binom{n}{k} = \binom{n}{n-k}$
- $|N_1 \times N_2 \times \cdots \times N_k| = |N_1| \cdot |N_2| \cdots |N_k|$, speziell $|N^k| = |N|^k$
- $n = |N|, k \leq n$, dann $|\{(\omega_1, \ldots, \omega_k) \in N^k \mid \omega_i \neq \omega_j \text{ für } i \neq j\}| = \frac{n!}{(n-k)!}$
- $n = |N|, k \leq n$, dann $|\{K \subset N \mid |K| = k\}| = \binom{n}{k} = |\{(\omega_1, \ldots, \omega_k) \in N^k \mid 1 \leq \omega_1 < \ldots < \omega_k \leq n\}|$
- $k, n \in \mathbb{N}$, dann $|\{(\omega_1, \ldots, \omega_k) \in N^k \mid 1 \leq \omega_1 \leq \ldots \leq \omega_k \leq n\}| = \binom{n+k-1}{k}$
- Insgesamt:

	mit Berücksichtigung der Reihenfolge	ohne Berücksichtigung der Reihenfolge
mit Zurücklegen	$\Omega = N^k$, $|\Omega| = n^k$	$\Omega = \{\omega \in N^k \mid 1 \leq \omega_1 \leq \ldots \leq \omega_n \leq n\}$, $|\Omega| = \binom{n+k-1}{k}$
ohne Zurücklegen	$\Omega = \{\omega \in N^k \mid \omega_i \neq \omega_j \text{ für } i \neq j\}$ $|\Omega| = \frac{n!}{(n-k)!}$	$\Omega = \{\omega \in N^k \mid 1 \leq \omega_1 < \ldots < \omega_k \leq n\}$, $|\Omega| = \binom{n}{k}$

2 Integration

Lebesgue-Maß $\lambda_d : \mathfrak{B}_d \to [0,\infty]$

- Für $B_1, B_2, \ldots \in \mathfrak{B}_d$ $\underline{\text{p.d.}}$: $\lambda_d(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \lambda_d(B_i)$.
- Für $a_1, b_1, \ldots, a_d, b_d \in \mathbb{R}$ mit $a_i \leq b_i$: $\quad \lambda_d([a_1,b_1] \times \cdots \times [a_d,b_d]) = \prod_{i=1}^{d}(b_i - a_i)$.
- Für $a \in \mathbb{R}^d$, $Q \in \mathrm{O}(d)$, $A \in \mathfrak{B}_d$: $\lambda_d(a + Q(A)) = \lambda_d(A)$
- $\forall x \in \mathbb{R}^d : \lambda_d(\{x\}) = 0$
- λ_d ist σ-subadditiv, σ-stetig von unten (nicht von oben!), monoton

Borel-messbare Funktionen

- $f : \mathbb{R}^d \to \mathbb{R}$ *Borel-messbar* $:\Leftrightarrow \forall B \in \mathfrak{B}_1 : \{f \in B\} \in \mathfrak{B}_d$
- $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ *Borel-messbar* $:\Leftrightarrow \forall B \in \mathfrak{B}_1 : \{f \in B\} \in \mathfrak{B}_d$ und $\{f = \infty\} \in \mathfrak{B}_d$
- $\mathcal{F}_d := \{f : \mathbb{R}^d \to \overline{\mathbb{R}} \mid f \text{ ist Borel-messbar}\}$, $\mathcal{F}_d^+ := \{f \in \mathcal{F}_d \mid f \geq 0\}$
- $\forall B \in \mathfrak{B}_d : \mathbb{1}_B \in \mathcal{F}_d^+$
- f stetig $\Rightarrow\ f \in \mathcal{F}_d$
- $f, g \in \mathcal{F}_d \Rightarrow a \cdot f + g, f \cdot g, f/g, \max(f,g) \in \mathcal{F}_d$ falls wohldefiniert ($a \in \overline{\mathbb{R}}$)

Elementare Funktionen

- Abbildung $e : \mathbb{R}^d \to \overline{\mathbb{R}}$ *elementar* $:\Leftrightarrow$
  $\exists b_1, \ldots, b_n \in \overline{\mathbb{R}} \land \exists B_1, \ldots, B_n \in \mathfrak{B}_d$ $\underline{\text{p.d.}}$ mit $e(x) = \sum_{i=1}^{n} b_i \cdot \mathbb{1}_{B_i}(x)$ für alle $x \in \mathbb{R}^d$
- $\mathcal{E}_d := \{f : \mathbb{R}^d \to \overline{\mathbb{R}} \mid f \text{ ist elementar}\}$, $\mathcal{E}_d^+ := \{f \in \mathcal{E}_d \mid f \geq 0\}$
- $\mathcal{E}_d \subset \mathcal{F}_d$

Lesbegue-Integral

- Für $f \in \mathcal{E}_d^+$: $\int f \, d\lambda_d := \sum_{i=1}^{n} b_i \cdot \lambda_d(B_i) \in [0,\infty]$
- Definiert für alle $f \in \mathcal{F}_d^+$
- $\int_B f \ \lambda_d := \int (f \cdot \mathbb{1}_B) d\lambda_d$ für $B \in \mathfrak{B}_d$
- Lesbegue-Integral ist monoton und linear
- Für $f : \mathbb{R} \to \mathbb{R}$ Borel-messbar und beschränkt auf $[a,b]$ mit $\lambda(\{x \in [a,b] \mid f \text{ unstetig in } x\}) = 0$ gilt:
$$\int_{[a,b]} f \ d\lambda = \underbrace{\int_a^b f(x) dx}_{\text{Riemann-Integral}}$$
- $\forall A \in \mathfrak{B}_d : \int_A f \ d\lambda_d = \int_A g \ d\lambda_d \Leftrightarrow \lambda_d(\{x \in \mathbb{R}^d \mid f(x) \neq g(x)\}) = 0$

Satz von der monotonen Konvergenz
Für jede monoton wachsende Folge $0 \leq f_1 \leq f_2 \leq \ldots$ in $\mathcal{F}_d^+$ gilt

$$\lim_{n\to\infty} \int f_n \, d\lambda_d = \int \lim_{n\to\infty} f_n \, d\lambda_d \in [0,\infty].$$

Satz von der dominierten Konvergenz
Für $f, f_1, f_2, \ldots \in \mathcal{F}_d$ mit $\lim_{n\to\infty} f_n = f$ und $g \in \mathcal{F}_d$ integrierbar mit $|f_n| \leq g$ für alle $n \in \mathbb{N}$ gilt

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

Insbesondere für $f \in \mathcal{F}_1$ integrierbar oder $f \in \mathcal{F}_1^+$:

$$\int_{\mathbb{R}} f \, d\lambda = \lim_{n\to\infty} \int_{[n,n]} f \, d\lambda.$$

Satz von Fubini
Für $f \in \mathcal{F}_d$ integrierbar oder $f \in \mathcal{F}_d^+$ und $B = B_1 \times \cdots \times B_d \subset \mathfrak{B}_1^d$ gilt

$$\int_B f \, d\lambda_d = \int_{B_{i_1}} \cdots \int_{B_{i_d}} f(x_1, \ldots, x_d) d\lambda(x_{i_d}) \cdots d\lambda(x_{i_1})$$

für jede Permutation $(i_1, \ldots, i_d)$ von $(1, \ldots, d)$.

Substitutionsregel (Riemann-Integration)
Für $a, b \in \mathbb{R}$ mit $a < b$, $g : [a,b] \to I$ stetig differenzierbar, $f : I \to \mathbb{R}$ stetig gilt

$$\int_a^b f(g(y)) \cdot g'(y) dy = \int_{g(a)}^{g(b)} f(x) dx.$$

Partielle Integration (Riemann-Integration)
Für $a, b \in \mathbb{R}$ mit $a < b$ und $f, g : [a,b] \to \mathbb{R}$ stetig differenzierbar gilt

$$\int_a^b f'(x) \cdot g(x) dx = f(x) \cdot g(x) \Big|_a^b - \int_a^b f(x) \cdot g'(x) dx.$$

3 Zufallsvariablen

Reellwertige Zufallsvariable
$X : \Omega \to \mathbb{R}$ ist *reellwertige Zufallsvariable* auf $(\Omega, \mathcal{A}, \mathrm{P})$, wenn

$$\forall I \text{ Intervall} : \{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{A}.$$

Gilt $\mathcal{A} = \mathfrak{P}(\Omega)$, dann ist jedes $X : \Omega \to \mathbb{R}$ Zufallsvariable.

Verteilungsfunktion
Die *Verteilungsfunktion*

$$\mathrm{F}_X : \mathbb{R} \to [0,1],\ x \mapsto \mathrm{P}(\{X \leq x\})$$

einer Zufallsvariablen ist monoton wachsend, rechtsseitig stetig und es gilt
$\lim\limits_{x\to\infty} \mathrm{P}(x) = 1$, $\lim\limits_{x\to-\infty} \mathrm{P}(x) = 0$, $\mathrm{F}_X(x) - \mathrm{F}_X(x\text{-}) = \mathrm{P}(\{X = x\})$.

Reellwertiger Zufallsvektor
$X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ ist *reellwertiger Zufallsvektor*, wenn jede Komponente X_i reellwertige Zufallsvariable ist. Es gilt:

$$X \text{ ist Zufallsvektor } \Leftrightarrow X \text{ ist Borel-messbar.}$$

Verteilung
Die *Verteilung* eines d-dimensionalen Zufallsvektors X:

$$\mathrm{P}_X : \mathfrak{B}_d \to [0,1],\ B \mapsto \mathrm{P}(\{X \in B\})$$

ist ein Wahrscheinlichkeitsmaß auf $\mathfrak{B}_d$.

Randverteilung
Zu einem Zufallsvektor $X = (X_1, \ldots, X_d)$ heißen $(i \in \{1, \ldots, d\})$

$$\mathrm{P}_{X_i} : \mathfrak{B}_1 \to [0,1],\ B \mapsto \mathrm{P}(\{X_i \in B\})$$

die *(eindimensionalen) Randverteilungen* von X.

Wahrscheinlichkeitsfunktion
Sei Ω eine abzählbare Menge. Eine Funktion $f : \Omega \to [0, \infty)$ heißt *Wahrscheinlichkeitsfunktion* falls

$$\sum_{\omega\in\Omega} f(\omega) = 1.$$

Wahrscheinlichkeitsdichte
Eine Funktion $f \in \mathcal{F}_d^+$ heißt *Wahrscheinlichkeitsdichte* falls

$$\int_{\mathbb{R}^d} f(x)\ d\lambda(x) = 1.$$

Bedeutung von W'funktionen/-dichten
Wahrscheinlichkeitsfunktionen/-dichten definieren Wahrscheinlichkeitsmaß auf $\mathcal{A}$ durch

$$\mathrm{P}(A) = \sum_{\omega \in A} f(\omega) \text{ bzw. } \mathrm{P}(A) = \int_A f(x)\ d\lambda(x)$$

für $A \in \mathcal{A}$.

Zu jeder Wahrscheinlichkeitsfunktion/-dichte f gibt es $(\Omega, \mathcal{A}, \mathrm{P})$ sodass darauf eine (diskrete/absolut stetige) Zufallsvariable X mit $f_X = f$ existiert.

Diskrete Zufallsvariablen
Eine Zufallsvariable heißt *diskret*, wenn $\mathrm{P}(\{X \in D\}) = 1$ für ein abzählbares $D \subset \mathbb{R}$ gilt.
Das kleinste solche $D =: D_X$ heißt *Träger* der Zufallsvariablen X.
Es gilt: $\underbrace{(\Omega, \mathcal{A}, \mathrm{P}) \text{ diskret}}_{\text{also } \Omega \text{ abzählbar}} \Rightarrow X(\Omega)$ abzählbar $\Rightarrow X$ diskret.

Diskrete Zufallsvektoren
Ein d-dimensionaler Zufallsvektor $X = (X_1, \ldots, X_d)$ heißt *diskret*, wenn es ein abzählbares $D \subset \mathbb{R}^d$ mit $\mathrm{P}(\{X \in D\}) = 1$ gibt.

Dann ist $D_X = \{x \in \mathbb{R}^d \mid \mathrm{P}(\{X = x\}) > 0\} \overset{\text{i.A.}}{\neq} D_{X_1} \times \cdots \times D_{X_d}$ der Träger von X.
Es gilt: X diskret $\Leftrightarrow \forall i \in \{1, \ldots, d\} : X_i$ diskret.
Für $h : \mathbb{R}^d \to \mathbb{R}$ Borel-messbar ist $h(X)$ diskret mit $D_{h(X)} = h(D_X)$.

Absolut stetige Zufallsvektoren
Ein Zufallsvektor X heißt *absolut stetig verteilt* falls die Verteilung P_X eine Dichte f_X besitzt.

Unabhängigkeit von Zufallsvariablen
Eine Folge $(X_i)_{i \in I}$ von Zufallsvariablen heißt *unabhängig* wenn für jede Folge $(J_i)_{i \in I}$ von Intervallen die Folge der Ereignisse $(\{X_i \in J_i\})_{i \in I}$ unabhängig ist.
Wegen:

$$\forall x \in \mathbb{R} : \{X \leq x\} \in \mathcal{A} \Leftrightarrow \forall I \text{ Intervall} : \{X \in I\} \in \mathcal{A}$$

ist $(X_i)_{i \in I}$ genau dann unabhängig, wenn für alle endlichen Mengen $\emptyset \neq \tilde{I} \subset I$ und $(x_i)_{i \in \tilde{I}} \subset \mathbb{R}$ gilt:

$$\mathrm{P}(\bigcap_{i \in \tilde{I}} \{X_i \leq x_i\}) = \prod_{i \in \tilde{I}} \mathrm{P}(\{X_i \leq x_i\}).$$

Die Unabhängigkeit einer Folge $(X_i)_{i \in I}$ impliziert die paarweise Unabhängigkeit von X_i, X_j für $i, j \in I$ mit $i \neq j$.

Beachte auch (**??**) und (**??**).

Es gilt: X, X unabhängig $\Leftrightarrow \exists c \in \mathbb{R} : \mathrm{P}(\{X = c\}) = 1$ (Übung)

Identische Verteilung
Zwei Zufallsvariablen X, Y heißen *identisch verteilt*, wenn $F_X = F_Y$.
Zwei Zufallsvektoren $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_d), \tilde{Y} = (\tilde{Y}_1, \ldots, \tilde{Y}_d)$ heißen *identisch verteilt*, wenn $P_{\tilde{X}} = P_{\tilde{Y}}$.
Es gilt $\tilde{X}, \tilde{Y}$ identisch verteilt $\Rightarrow \forall i \in \{1, \ldots, d\} : \tilde{X}_i, \tilde{Y}_i$ identisch verteilt

Vergleich

diskret	**absolut stetig**
W'funktion/Dichte	
Es existiert W'funktion: $f_X : D_X \to [0,1],\ x \mapsto \mathrm{P}(\{X = x\})$	P_X besitzt Dichte f_X
Verteilungsfunktion (eindim.!)	
$\mathrm{F}_X(x) = \sum\limits_{y \in D_X \cap (-\infty, x]} \mathrm{P}(\{X = y\})$	$\mathrm{F}_X(x) = \int_{(-\infty, x]} f_X(y) d\lambda(y)$
Verteilung	
$\mathrm{P}_X(A) = \sum\limits_{x \in D_X \cap A} \mathrm{P}(\{X = x\})$	$\mathrm{P}_X(A) = \int_A f_X(x) d\lambda(x)$
Randverteilung/Randdichte $X = (X_1 \ldots, X_d)$	
$P_{X_i}(\{x\}) =$ $\mathrm{P}(\{X_i = x\} \cap \{(X_1, \ldots, X_d) \in D_X\})$	$f_{X_i}(x) =$ $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d) d\lambda(x_1) \cdots d\lambda_{x_d}$ (alles integrieren außer i-te Koordinate)
$X_1, \ldots, X_d$ *unabhängig*	
$\forall x_1 \ldots, x_d \in \mathbb{R}:$ $\mathrm{P}(\bigcap_{i=1}^d \{X_i = x_i\}) = \prod_{i=1}^d \mathrm{P}(\{X_i = x_i\})$	$\forall B_1, \ldots, B_d \in \mathfrak{B}_1:$ $\mathrm{P}(\bigcap_{i=1}^d \{X_i \in B_i\}) = \prod_{i=1}^d \mathrm{P}(\{X_i \in B_i\})$ $\Leftrightarrow$ $\forall B_1, \ldots, B_d \in \mathfrak{B}_1:$ $\mathrm{P}_X(B_1 \times \cdots \times B_d) = \prod_{i=1}^d \mathrm{P}_{X_i}(B_i)$ Gemeinsame Verteilung P_X von $X_1, \ldots, X_d$ ist also Produkt der Randverteilungen $P_{X_{i \in \{1, \ldots, d\}}}$
X, Y *identisch verteilt (eindim.!)*	
$D_X = D_Y$ und $f_X = f_Y$ $\forall z \in \mathbb{R}: \mathrm{P}(\{X = z\}) = \mathrm{P}'(\{Y = z\})$	$P_X = P_Y$ $\forall A \in \mathfrak{B}_1: \mathrm{P}(\{X \in A\}) = \mathrm{P}'(\{Y \in A\})$

4 Erwartungswert & Varianz

Integriebare Zufallsvariablen

- Diskretes X mit Träger D_X *integrierbar* $:\Leftrightarrow \sum_{x \in D_X} |x| \cdot \mathrm{P}(\{X = x\}) < \infty$
  Diskrete Zufallsvariablen mit <u>endlichem</u> Träger sind immer integrierbar!
- Absolut stetige Zufallsvariable X mit Dichte f_X *integrierbar* $:\Leftrightarrow \int_{\mathbb{R}} |x| \cdot f_X(x) d\lambda(x) < \infty$
- $\mathfrak{L}_1 := \{X \text{ integrierbar}\}$ ist Vektorraum
- *Erwartungswert* $\mathrm{E}(X)$ für diskretes $X \in \mathfrak{L}_1$: $\mathrm{E}(X) := \sum_{x \in D_X} x \cdot \mathrm{P}(\{X = x\}) \in \mathbb{R}$
- *Erwartungswert* $\mathrm{E}(X)$ für absolut stetiges $X \in \mathfrak{L}_1$: $\mathrm{E}(X) := \int_{\mathbb{R}} x \cdot f_X \; d\lambda(x) \in \mathbb{R}$
- Der Erwartungswert ist linear und monoton.
- **Transformationssatz**: Für X d-dimensionaler Zufallsvektor diskret/ absolut stetig und $h : \mathbb{R}^d \to \mathbb{R}$ Borel-messbar gilt

  $$h(X) \in \mathfrak{L}_1 \Leftrightarrow \begin{cases} \sum_{x \in D_X} |h(x)| \cdot \mathrm{P}(\{X = x\}) < \infty & X \text{ diskret} \\ \int_{\mathbb{R}^d} |h(x)| \cdot f_X(x) \; d\lambda_d(x) < \infty & X \text{ absolut stetig} \end{cases}.$$

  Gegebenenfalls

  $$\mathrm{E}(h(X)) = \begin{cases} \sum_{x \in D_X} h(x) \cdot \mathrm{P}(\{X = x\}) < \infty & X \text{ diskret} \\ \int_{\mathbb{R}^d} h(x) \cdot f_X(x) \; d\lambda_d(x) < \infty & X \text{ absolut stetig} \end{cases}.$$

- $X, Y \in \mathfrak{L}_1$ unabhängig $\Rightarrow X \cdot Y \in \mathfrak{L}_1$ mit $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$ (1)

Quadratisch integrierbare Zufallsvariablen

- X *quadratisch integrierbar* $:\Leftrightarrow X^2 \in \mathfrak{L}_1$
- $\mathfrak{L}_2 := \{X \text{ quadratisch integrierbar}\}$ ist Untervektorraum von $\mathfrak{L}_1$
- Für X diskret mit Träger D_X bzw. absolut stetig mit Dichte f_X gilt:

  $$X \in \mathfrak{L}_2 \Leftrightarrow \begin{cases} \sum_{x \in D_X} x^2 \cdot \mathrm{P}(\{X = x\}) < \infty & \text{X diskret} \\ \int_{\mathbb{R}^d} x^2 \cdot f_X(x) \; d\lambda_d(x) < \infty & \text{X absolut stetig} \end{cases}.$$

  Gegebenenfalls ist $\mathrm{E}(X^2)$ durch obige(s) Summe/Integral gegeben.
- Für $X \in \mathfrak{L}_2$ ist $\mathrm{Var}(X) := \mathrm{E}((X - \mathrm{E}(X))^2)$ die *Varianz* von X.
- $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2, \quad \mathrm{Var}(\alpha \cdot X + \beta) = \alpha^2 \, \mathrm{Var}(X)$
- **Tschebyschev-Ungleichung**. Für $X \in \mathfrak{L}_2, \varepsilon > 0$ gilt:

  $$\mathrm{P}(|X - \mathrm{E}(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}(X)$$

  sowie

  $$\mathrm{Var}(X) = 0 \Leftrightarrow \mathrm{P}(\{X = \mathrm{E}(x)\}) = 1.$$

- **Formel von Bienaymé**. Für $X_1, \ldots, X_n \in \mathfrak{L}_2$ unabhängig gilt:

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

- $X, Y \in \mathfrak{L}_2 \Rightarrow X \cdot Y \in \mathfrak{L}_1$
- $\overline{\mathfrak{L}_2}$ ist Hilbertraum mit $\langle X, Y \rangle := \mathrm{E}(X \cdot Y)$

Kovarianz

- Für $X, Y \in \mathfrak{L}_2$ ist die *Kovarianz* definiert durch $\mathrm{Cov}(X,Y) := \mathrm{E}\left((X - \mathrm{E}(X) \cdot (Y - \mathrm{E}(Y))\right)$
- Für $X, Y \in \mathfrak{L}_2$ gilt $\mathrm{Cov}(X,Y) = \mathrm{E}(X \cdot Y) - \mathrm{E}(X) \cdot \mathrm{E}(Y)$
- X, Y heißen *unkorreliert* wenn $\mathrm{Cov}(X,Y) = 0$. Es gilt: X, Y unabhängig $\Rightarrow X, Y$ unkorreliert (2)

- Für X, Y mit $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$ ist

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}} = \cos(\sphericalangle(X,Y))$$

  der *Korrelationskoeffizient* von X und Y.
- **Cauchy-Schwarz**: $X, Y \in \mathfrak{L}_2 \Rightarrow |\,\mathrm{E}(X \cdot Y)| \leq \sqrt{\mathrm{E}(X^2) \cdot \mathrm{E}(Y^2)}$

5 Verteilungen

Bernoulli-Verteilung $X \sim \mathbf{B}(1,p)$ mit $p \in [0,1]$

- $\mathrm{P}(\{X=1\}) = p,\ \mathrm{P}(\{X=0\}) = 1-p$
- diskret mit $D_X = \begin{cases} \{0,1\} & \text{falls } p \in (0,1) \\ \{0\} & \text{falls } p = 0 \\ \{1\} & \text{falls } p = 1 \end{cases}$
- $\mathrm{F}_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1-p & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}$
- $\mathrm{E}(X) = p$, $\mathrm{Var}(X) = p - p^2 = p(1-p)$

Binomial-Verteilung $X \sim \mathbf{B}(n,p)$ mit $p \in [0,1]$

- $\forall k \in \{0,\dots,n\} : \mathrm{P}(\{X=k\}) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$
- diskret mit $D_X = \begin{cases} \{0,\dots,n\} & \text{falls } p \in (0,1) \\ \{0\} & \text{falls } p = 0 \\ \{n\} & \text{falls } p = 1 \end{cases}$
- $\mathrm{E}(X) = n \cdot p$, $\mathrm{Var}(X) = n \cdot p \cdot (1-p)$
- Anwendung: Zählen der Erfolge von n unabhängigen, hintereinander ausgeführten Experimenten mit Erfolgswahrscheinlichkeit p_n

Hypergeometrische Verteilung $X \sim \mathbf{H}(N, N_0, n)$

- $\forall l \in D_X : \mathrm{P}(\{X=l\}) = \frac{\binom{N_0}{l} \cdot \binom{N-N_0}{n-l}}{\binom{N}{n}}$
- diskret mit $D_X = \{\max(0, n-(N-N_0)), \dots, \min(N_0, n)\}$
- Anwendung: unter N Objekten finden sich N_0 markierte und es werden n entnommen

Poisson-Verteilung $X \sim \mathbf{P}(\lambda)$ für $\lambda > 0$

- $\forall k \in \mathbb{N}_0 : \mathrm{P}(\{X=k\}) = \exp(-\lambda) \cdot \frac{\lambda^k}{k!}$
- diskret mit $D_X = \mathbb{N}_0$
- $X \sim \mathbf{P}(\lambda_2), Y \sim \mathbf{P}(\lambda_2) \Rightarrow X + Y \sim \mathbf{P}(\lambda_2 + \lambda_2)$ `(Übung)`
- $\mathrm{E}(X) = \lambda = \mathrm{Var}(X)$, $\mathrm{E}(X^2) = \lambda^2 + \lambda$
- Anwendung: Approximation von $\mathbf{B}(n,p)$ durch $\mathbf{P}(\lambda)$ mit $\lambda = n \cdot p$ für 'große' n und 'kleine' p, also Eintreffen eines seltenen Ereignis bei großer Anzahl an Wiederholungen

Geometrische Verteilung $X \sim \mathbf{G}(p)$ mit $p \in (0,1]$

- $\forall k \in \mathbb{N} : \mathrm{P}(\{X = k\}) = p \cdot (1-p)^{k-1}$
- diskret mit $D_X = \mathbb{N}$
- $\mathrm{E}(X) = \frac{1}{p}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$, $\mathrm{E}(X^2) = \frac{2-p}{p^2}$
- Gedächtnislos:
  $\forall k_1, k_2 \in \mathbb{N}$ mit $k_1 < k_2 : \mathrm{P}(\{X > k_2\} \mid \{X > k_1\}) = \mathrm{P}(\{X > k_2 - k_1\})$
- Anwendung: diskretes Warten bis zum ersten Eintritt eines Ereignisses

Gleichverteilung $X \sim \mathbf{U}([a,b])$ für $-\infty < a < b < \infty$

- absolut stetig mit $f_X(x) = \begin{cases} \frac{1}{b-a} & \text{falls } x \in [a,b] \\ 0 & \text{sonst} \end{cases}$
- $\mathrm{F}_X(x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } x \in [a,b] \\ 1 & \text{sonst} \end{cases}$
- $\mathrm{E}(X) = \frac{a+b}{2}$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$, $\mathrm{E}(X^2) = \frac{b^3-a^3}{3(b-a)}$

Einpunktverteilung $X \sim \mathbf{U}(\{c\})$ für ein $c \in \mathbb{R}$

- $P(\{X = c\}) = c$
- diskret mit $D_X = \{c\}$
- $\mathrm{F}_X(x) = \begin{cases} 0 & \text{falls } x < c \\ 1 & \text{falls } x \geq c \end{cases}$
- $\mathrm{E}(X) = c$, $\mathrm{Var}(X) = 0$

Exponentialverteilung $X \sim \mathbf{Exp}(\lambda)$ für $\lambda > 0$

- absolut stetig mit $f_X : \mathbb{R} \to [0,\infty],\ x \mapsto \begin{cases} \lambda \cdot \exp(-\lambda x) & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$
- $\mathrm{F}_X(x) = 1 - \exp(-\lambda x)$ für $x > 0$; $F_X(x) = 0$ für $x \leq 0$.
- $\mathrm{E}(X) = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$, $\mathrm{E}(X^2) = \frac{2}{\lambda^2}$
- Gedächtnislos: $\forall s,t > 0 : \mathrm{P}(\{X > t+s\} \mid \{X > t\}) = \mathrm{P}(\{X > s\})$
- Anwendung: Warten bis zum ersten Eintritt eines Ereignisses (zB Lebensdauer, radioaktiver Zerfall,...)

Standard-Normalverteilung $X \sim \mathbf{N}(0,1)$

- absolut stetig mit $f_X : \mathbb{R} \to [0,\infty],\ x \mapsto \frac{1}{\sqrt{2\pi}} \cdot \exp(-\frac{x^2}{2})$
- $\Phi(x) := \mathrm{F}_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{y^2}{2}) d\lambda(y)$ (keine explizite Formel)

- $\forall z \in \mathbb{R} : \Phi(z) + \Phi(-z) = 1$ (Übung)
- $\mathrm{E}(X) = 0$, $\mathrm{Var}(X) = 1$

Normalverteilung $X \sim \mathbf{N}(\mu, \sigma^2)$ für $\mu \in \mathbb{R},\ \sigma > 0$

- absstetig mit $f_X : \mathbb{R} \to [0, \infty],\ x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp(-\frac{(x-\mu)^2)}{2\sigma^2})$
- $X \sim \mathbf{N}(\mu, \sigma^2) \Rightarrow a \cdot X + b \sim \mathbf{N}(a \cdot \mu + b, a^2 \cdot \sigma^2)$
- Also: $\frac{X-\mu}{\sigma} \sim \mathbf{N}(0, 1)$
- $\mathrm{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$

Gleichverteilung (mehrdimensional) $X \sim \mathbf{U}(S)$ für $S \subset \mathbb{R}^n$ <u>endlich</u>

- $\forall x \in S : \mathrm{P}(\{X = x\}) = \frac{1}{|S|}$
- diskret mit $D_X = S$

Gleichverteilung (mehrdimensional) $X \sim \mathbf{U}(G)$ für $G \in \mathfrak{B}_d$

- $\mathrm{P}_X(A) = \frac{\lambda_d(A \cap G)}{\lambda_d(G)} = \int_A \frac{1}{\lambda_d(G) \cdot 1_G(x)} d\lambda(x)$
- absolut stetig mit $f_X = \frac{1}{\lambda_d(G)} \cdot \mathbb{1}_G$

Standard-Normalverteilung (mehrdimensional)

- absolut stetig mit $f_X : \mathbb{R}^d \to [0, \infty),\ x \mapsto (2\pi)^{-d/2} \cdot \exp(-\frac{1}{2} \sum_{i=1}^{d} x_i^2)$
- $X = (X_1, \ldots, X_d)$ ist standard-normalverteilt $\Leftrightarrow X_1, \ldots, X_d$ i. i. d $\sim \mathbf{N}(0, 1)$

6 Sonstiges

Exponentialfunktion:

- $\exp : \mathbb{C} \to \mathbb{C},\ x \mapsto \sum_{k=0}^{\infty} \frac{x^k}{k!}$
- $\exp(a \cdot b) = \exp(a) + \exp(b)$
- $\exp$ stetig, $\exp > 0$
- $\exp' = \exp$

Natürlicher Logarithmus

- $\ln : (0, \infty) \to \mathbb{R} :\ x \mapsto \exp^{-1}(x)$
- $\ln(a \cdot b) = \ln(a) + \ln(b)$
- $\ln'(x) = \frac{1}{x}$

Trigonometrische Funktionen

- $\cos : \mathbb{R} \to \mathbb{R} :\ x \mapsto \sum_{k=0}^{\infty} (-1)^k \cdot \frac{x^{2k}}{(2k)!}$
- $\sin : \mathbb{R} \to \mathbb{R} :\ x \mapsto \sum_{k=0}^{\infty} (-1)^k \cdot \frac{x^{2k+1}}{(2k+1)!}$
- $\cos' = -\sin$, $\sin' = \cos$

Reihen

- **Harmonische Reihe**: $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$
- **Geometrische Reihe**: $\sum_{n=0}^{\infty} a \cdot x^n = \frac{a}{1-x}$ für $a, x \in \mathbb{R}$ mit $|x| < 1$, divergent für alle anderen $x \in \mathbb{R}$
- **Alternierende harmonische Reihe**: $\sum_{n=1}^{\infty} (-1)^n \cdot \frac{1}{n} = -\ln(2)$
- $\sum_{k=0}^{\infty} \exp(-\lambda) \cdot \frac{\lambda^k}{k!} = 1$ für $\lambda > 0$

Binomischer Lehrsatz:
Für $x, y \in \mathbb{R},\ n \in \mathbb{N}_0$ gilt:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$