

Vorlesungsskript: Wahrscheinlichkeitstheorie für Physiker von
Dr. Nagel

Simon Stützer

Letzte Änderung: 11. September 2008

Inhaltsverzeichnis

1	Wahrscheinlichkeitsräume	3
1.1	Wahrscheinlichkeitsraum als Grundmodell	3
1.1.1	Wahrscheinlichkeitsraum: $[\Omega, \mathcal{A}, P]$	3
1.1.2	Schritte zur Modellierung: Ω	3
1.1.3	Schritte zur Modellierung: \mathcal{A}	4
1.1.4	Schritte zur Modellierung: P	6
1.2	Beschreibungsmöglichkeiten(Bestimmungsstücke) für Wahrscheinlichkeitsmaße . . .	7
1.2.1	Diskrete Wahrscheinlichkeitsräume	7
1.2.2	Kontinuierliche Wahrscheinlichkeitsräume $[\mathbb{R}, \mathcal{R}, P]$	8
1.3	Einige spezielle Wahrscheinlichkeitsräume	9
1.3.1	diskrete Gleichverteilung	9
1.3.2	ka Verteilungsverteilung	9
1.3.3	Bernoulli-Verteilung	9
1.3.4	würfeln mit nichtunterscheidbaren Würfeln	10
1.3.5	Physik: Maxwell-Boltzman, Fermi-Dirac, Bose-Einstein Verteilung	10
1.3.6	geometrische Verteilung	12
1.3.7	Poisson-Verteilung	12
1.3.8	Gleichverteilung	12
1.3.9	Normalverteilung	13
1.3.10	Cauchy-Verteilung (Lorenz-Verteilung, Breit-Wiegner-Verteilung)	13
1.4	Bedingte Wahrscheinlichkeit	14
1.4.1	Anschauliche Vorstellung:	14
1.4.2	Schritte zur Formalisierung	14
1.5	stochastische Unabhängigkeit von Ereignissen	16
1.5.1	Anschauliche Vorstellung	16
1.5.2	Definition	17
1.5.3	Beispiele	17
2	Zufällige Variablen, Zufallsgrößen, zufällige Vektoren	19
2.1	Zufällige Variablen: meßbare Abbildungen zwischen wahrscheinlichkeitsräumen . .	19
2.1.1	Formalisierung	20
2.2	Zufallsgrößen: reellwertige zufällige Variablen	20
2.3	Unabhängigkeit von Zufallsgrößen	21
2.4	Diskrete Zufallsgrößen	23
2.4.1	Formalisierung	23
2.4.2	Beispiele	24
2.5	Stetige Zufallsgrößen	26
2.5.1	Formalisierung	26
2.5.2	Wichtige Spezialfälle	27
2.6	Zufällige Vektoren	29
2.6.1	Formalisierung	29
2.6.2	Wichtige Spezialfälle	33

3	Weitere Verteilungsgesetze von transformierten zufälligen Vektoren	38
3.1	Transformation von eindimensionalen Zufallsgrößen	38
3.2	Summe zweier Zufallsgrößen	39
3.2.1	diskreter Fall	39
3.2.2	stetiger Fall	40
3.3	Produkt und Quotient zweier Zufallsgrößen	42
3.4	Injektive differenzierbare Transformationen von zufälligen Vektoren	43
4	Erwartungswert, Varianz und Kovarianz	45
4.1	Vorbemerkung	45
4.2	Erwartungswert einer Zufallsgröße	46
4.3	Varianz einer Zufallsgröße	52
4.4	Die Kovarianz zweier Zufallsgrößen / die Kovarianz-Matrix eines zufälligen Vektors	54
5	Ungleichungen und Grenzwertsätze	57
5.1	Einführung	57
5.1.1	Vorbetrachtung	57
5.2	Markovsche und Tschebyschersche Ungleichung	58
5.3	Gesetze der Großen Zahlen	59
5.4	Der zentrale Grenzwertsatz	61

Kapitel 1

Wahrscheinlichkeitsräume

1.1 Wahrscheinlichkeitsraum als Grundmodell

1.1.1 Wahrscheinlichkeitsraum: $[\Omega, \mathcal{A}, P]$

- Ω ... nichtleere Menge der Elementarereignisse / nichtleere Menge der möglichen Beobachtungsereignisse
- $\mathcal{A} \subseteq \mathcal{P}(\Omega)$... Ereignisalgebra, σ -Algebra
- $\mathcal{P}(\Omega)$... Potenzmenge von Ω / Menge aller Teilmengen einschließlich Ω und \emptyset
- $P: \mathcal{A} \rightarrow [0,1]$... Wahrscheinlichkeitsmaß

1.1.2 Schritte zur Modellierung: Ω

1. Ω ... hängt auch davon ab, was beobachtet wird
 ω ... Ereignis/Beobachtung: $\omega \in \Omega$

Beispiele

- (a) Einmaliger Münzwurf:

$$\Omega = \{0, 1\} / (0 \dots \text{Wappen}, 1 \dots \text{Zahl})$$

- (b) n-maliger Münzwurf mit Berücksichtigung der Reihenfolge:

$$\begin{aligned}\Omega &= \{(a_1, \dots, a_n) : a_i \in \{0, 1\}, i = 1, \dots, n\} \\ &= \underbrace{\{0, 1\} \times \{ \dots \} \times \dots \times \{0, 1\}}_{n \text{ mal}} = \{0, 1\}^n\end{aligned}$$

- (c) n-maliger Münzwurf, Beobachtung Anzahl der "1":

$$\Omega = \{0, 1, 2, \dots, n\}$$

(d) 4-maliges würfeln mit einem Würfel unter Berücksichtigung der Reihenfolge:

$$\begin{aligned}\Omega &= \{(a_1, a_2, a_3, a_4) : a_i \in \{1, \dots, 6\}, i = 1, \dots, 4\} \\ &= \{1, \dots, 6\}^4\end{aligned}$$

(e) gleichzeitiges würfeln mit 4 nichtunterscheidbaren Würfeln:

$$\begin{aligned}\Omega &= \{(a_1, \dots, a_6) : a_i \in \{0, \dots, 4\}, i = 1, \dots, 4, \sum_{i=1}^6 a_i = 4\} \\ &= \{1, \dots, 6\}^4\end{aligned}$$

(f) Zustand eines Teilchens: 3 Orts- und 3 Impulskoordinaten:

$$\Omega = \mathbb{R}^6$$

(g) Zustand einer Gesamtheit von N Teilchen:

$$\Omega = \mathbb{R}^{6N} \text{ bzw. } \mathbb{R}^{6N} \text{ (Gamma - Raum)}$$

1.1.3 Schritte zur Modellierung: \mathcal{U}

2. \mathcal{U} ... Menge von Ereignissen

Unter einem Ereignis versteht man eine Teilmenge von Ω

Beispiele

zu (b) n-maliger Münzwurf, $n \geq 2$ Ereignis: A ... Wurfresultat des ersten und zweiten Wurfes stimmen überein

$$\begin{aligned}A &= \{(a_1, \dots, a_n) \in \{0, 1\}^n : a_1 = a_2\} \\ &= \{0\}^2 \times \{0, 1\}^{n-2} \cup \{1\}^2 \times \{0, 1\}^{n-2}\end{aligned}$$

Ereignis B: B ... es fällt mindestens einmal "1"

$$\begin{aligned}B &= \{0, 1\}^n \setminus \{0, \dots, 0\} \\ &= \{0, 1\}^n \setminus \{0\}^n\end{aligned}$$

zu (c) B ... es fällt mindestens einmal "1"

$$B = \{1, \dots, n\} = \{0, 1, \dots, n\} \setminus \{0\}$$

Wiederholung

Wahrscheinlichkeitsraum... $[\Omega, \mathcal{U}, P]$

Ω ... Menge der möglichen Ereignisse

$\mathcal{U} \subseteq \mathcal{P}(\Omega)$, $A \in \mathcal{U}$, A ... Ereignis

Beobachtung: $\omega \in \Omega$

Sprechweisen

Für gegebene Ereignisse $A, B, C \dots$ sagt man:

- "A ist eingetreten" $\Leftrightarrow \omega \in A$
- "A ist nicht eingetreten" $\Leftrightarrow \omega \in A^c$
- "A und B sind eingetreten" $\Leftrightarrow \omega \in A \cap B$
- "A oder B sind eingetreten" $\Leftrightarrow \omega \in A \cup B$
- "A ist eingetreten und B nicht" $\Leftrightarrow \omega \in A \cap B^c = A \setminus B$
- "wenigstens eines der Ereignisse A_1, A_2, \dots ist eingetreten" $\Leftrightarrow \omega \in \bigcup_{i=1}^{\infty} A_i$
- "alle Ereignisse A_1, A_2, \dots sind eingetreten" $\Leftrightarrow \omega \in \bigcap_{i=1}^{\infty} A_i$

Außerdem gilt:

- Ω tritt stets ein, Ω heißt sicheres Ereignis
- \emptyset tritt nie ein, \emptyset heißt unmögliches Ereignis

Anforderungen an das Mengensystem \mathcal{A} als System von Ereignissen

- $\mathcal{A} \subseteq \mathcal{P}(\Omega)$
- interessierende Ereignisse sollen in \mathcal{A} liegen (\mathcal{A} also nicht zu klein)
- \mathcal{A} abgeschlossen gegenüber Komplementbildung und gegenüber abzählbar unendlicher Vereinigungs- und Durchschnittsbildung
- jedem Ereignis $A \in \mathcal{A}$ soll eine Wahrscheinlichkeit $P(A)$ zugeordnet werden können, d.h. \mathcal{A} muss als Definitionsbereich eines Wahrscheinlichkeitsmaßes geeignet sein. (!!! deshalb kann nicht immer $\mathcal{A} = \mathcal{P}(\Omega)$ gewählt werden)

Definition 1.1.1 (σ -Algebra) Es sei Ω eine nichtleere Menge. Das Mengensystem $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ -Algebra über Ω , wenn gilt:

- (a) $\Omega \in \mathcal{A}$
- (b) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ für alle $A \subseteq \Omega$
- (c) Für alle $A_1, A_2, \dots \in \mathcal{A}$ ist auch $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Folgerung 1.1.1 Wenn \mathcal{A} eine σ -Algebra über Ω ist, dann gilt:

- (a) $\emptyset \in \mathcal{A}$
- (b) für alle $A_1, A_2, \dots \in \mathcal{A}$ ist auch $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

Beweise:

(a) Aus Definition 1.1.1 (a) und (b) folgt $\emptyset = \Omega^c \in \mathcal{A}$

(b) Seien $A_1, A_2, \dots \in \mathcal{A}$. Dann folgt aus (b) und (c)

$$\bigcap_{i=1}^{\infty} A_i = \left(\left(\bigcap_{i=1}^{\infty} A_i \right)^c \right)^c = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathcal{A}$$

(c) $A_1, \dots, A_n \in \mathcal{A}$, $\emptyset = A_{n+1} = A_{n+2} = \dots$

$$\bigcup_{i=1}^n A_i = \bigcup_{i=1}^{\infty} A_i, \quad \Omega = A_{n+1} = A_{n+2} = \dots$$

$$\bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$$

Beispiele

(a) Ω beliebig nichtleere Menge

$\mathcal{A} = \{\emptyset, \Omega\}$ ist eine σ -Algebra über Ω

(b) Ω beliebig nichtleere

$\mathcal{A} = \mathcal{P}(\Omega)$ ist eine σ -Algebra über Ω

Wir verwenden hier immer folgende Standard- σ -Algebren:

i. Falls Ω endlich oder abzählbar unendlich ist: $\mathcal{A} = \mathcal{P}(\Omega)$

ii. Falls $\Omega = \mathbb{R}$, $\mathcal{A} \dots$ Borelsche σ -Algebra

$$\mathcal{A} = \mathcal{B} = \sigma(\{(a, b) : -\infty < a < b < \infty\}) = \sigma(\{(-\infty, a] : a \in \mathbb{R}\})$$

Wobei für $\mathcal{C} \subseteq \mathcal{P}(\Omega) : \sigma(\mathcal{C}) \dots$ kleinste σ -Algebra über Ω , die \mathcal{C} enthält

$$\sigma(\mathcal{C}) = \bigcap \gamma, \quad \gamma \text{ ist } \sigma\text{-Algebra über } \Omega, \text{ die } \mathcal{C} \text{ enthält}$$

iii. Falls $\Omega = \mathbb{R}^n$

Borelsche σ -Algebra:

$$\begin{aligned} \mathcal{A} = \mathcal{B}_n &= \sigma(\{(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_n, b_n) : -\infty < a_i < b_i < \infty, i = 1, \dots, n\}) \\ &= \sigma(\{(-\infty, a_1] \times (-\infty, a_2] \times \dots \times (-\infty, a_n] : (a_1, \dots, a_n) \in \mathbb{R}^n\}) \end{aligned}$$

1.1.4 Schritte zur Modellierung: P

3. Wahrscheinlichkeitsmaß P

anschauliche Vorstellung: "Gesamtmasse 1" bzw. "100%" ist auf Ω "verteilt"

Definition 1.1.2 (Axiomensystem von Kolmogorov) Ein Wahrscheinlichkeitsraum ist ein Tripel $[\Omega, \mathcal{A}, P]$ wobei Ω eine nichtleere Menge, \mathcal{A} eine σ -Algebra über Ω ist und $P: \mathcal{A} \rightarrow [0, 1]$ mit

1. $P(\Omega) = 1$

2. Für die Folgen $A_1, A_2, \dots \in \mathcal{A}$ mit $A_i \cap A_j = \emptyset$, falls $j \neq i$ gilt:

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

(σ -Additivität) Die Mengenfunktion P heißt Wahrscheinlichkeitsraummaß

Folgerung 1.1.2 Es sei $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum, und es seien $A, B, C, \dots \subset \mathcal{A}$ Ereignisse, dann gilt:

1. $P(\emptyset) = 0$
2. $\forall n \in \mathbb{N} : A_i \cap A_j = \emptyset \text{ für } i \neq j \Rightarrow P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$
3. $P(A^c) = 1 - P(A)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. $A \subseteq B \Rightarrow P(A) \leq P(B)$ (Monotonie)
6. $A \subseteq B \Rightarrow P(B \setminus A) = P(B) - P(A)$
7. $A_1 \subseteq A_2 \subseteq \dots \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$
8. $A_1 \supseteq A_2 \supseteq \dots \Rightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$
9. $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$ (σ -Subadditivität von P)

1.2 Beschreibungsmöglichkeiten (Bestimmungstücke) für Wahrscheinlichkeitsmaße

$P : \mathcal{A} \rightarrow [0, 1]$

Frage: Teilsystem $\mathcal{L} \subseteq \mathcal{A}$ suchen, dass möglichst klein und übersichtlich ist und für das gilt:

Falls für Wahrscheinlichkeitsmaße P_1, P_2 gilt:

$P_1(A) = P_2(A)$ für alle $A \subseteq \mathcal{L}$ dann $P_1(A) = P_2(A)$ für alle $A \subseteq \mathcal{A}$, d.h. $P_1 = P_2$

Betrachte nur die wichtigsten Standardfälle:

1.2.1 Diskrete Wahrscheinlichkeitsräume

Lemma 1.2.1 Es sei $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum wobei Ω endlich oder abzählbar unendlich und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. Ein Wahrscheinlichkeitsmaß ist durch die Werte $P(\{\omega\}), \omega \in \Omega$, eindeutig festgelegt.

Beweis

$A \in \mathcal{P}(\Omega)$, damit $A \subseteq \Omega$. $\rightarrow A$ ist endlich oder abzählbar unendlich

schreibe $A = \bigcup_{\omega \in A} \{\omega\}$ für $\omega \neq \omega'$ gilt $\{\omega\} \cap \{\omega'\} = \emptyset$

Wegen der endlichen bzw. σ -Additivität von P gilt:

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\})$$

Also: Formel für Wahrscheinlichkeitsmaße auf endlichen oder abzählbar unendlichen Grundmengen.

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

Andererseits: Zu jeder Wahl von Werten $P(\{\omega\})$, $\omega \in \Omega$, mit $P(\{\omega\}) \in [0, 1]$ und $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$ existiert ein Wahrscheinlichkeitsmaß P auf $[\Omega, \mathcal{P}(\Omega)]$

1.2.2 Kontinuierliche Wahrscheinlichkeitsräume $[\mathbb{R}, \mathcal{R}, P]$

Lemma 1.2.2 Ein Wahrscheinlichkeitsmaß P auf $[\mathbb{R}, \mathcal{R}]$ ist durch die Werte $P((-\infty, x])$, $x \in \mathbb{R}$ eindeutig festgelegt.

Definition 1.2.1 (Verteilungsfunktion) Es sei $[\mathbb{R}, \mathcal{R}, P]$ ein Wahrscheinlichkeitsraum. Die Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit $F(x) = P((-\infty, x])$, $x \in \mathbb{R}$ heißt Verteilungsfunktion des Wahrscheinlichkeitsmaßes P .

Satz 1.2.1 (Charakterisierung von Verteilungsfunktionen)

1. Es sei $[\mathbb{R}, \mathcal{R}, P]$ ein Wahrscheinlichkeitsraum und F die Verteilungsfunktion von P . Dann gilt:
 - (a) $x_1 < x_2 \rightarrow F(x_1) \leq F(x_2)$, $x_1, x_2 \in \mathbb{R}$ (Monotonie)
 - (b) $\forall x \in \mathbb{R} : \lim_{h \rightarrow 0} F(x+h) = F(x)$ (rechtsseitige Stetigkeit)
 - (c) $\lim_{x \rightarrow \infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$
2. Zu jeder Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit den Eigenschaften (a), (b) und (c) existiert ein Wahrscheinlichkeitsmaß P auf $[\mathbb{R}, \mathcal{R}]$, so dass F die Verteilungsfunktion von P ist.

Beweis

zu a Sei $x_1 < x_2 \rightarrow (-\infty, x_1] \subseteq (-\infty, x_2]$ daraus folgt mit Definition 1.1.1
 $F(x_1) = P((-\infty, x_1]) \leq P((-\infty, x_2]) = F(x_2)$

zu b Sei $x \in \mathbb{R}$, $(h_n)_{n \in \mathbb{N}}$ ist $h_1 \geq h_2 \geq \dots$, $\lim_{n \rightarrow \infty} h_n = 0$

$$\rightarrow (-\infty, x] = \bigcap_{n=1}^{\infty} (-\infty, x+h_n] \text{ und es folgt}$$

$$F(x) = P((-\infty, x]) = P\left(\bigcap_{n=1}^{\infty} (-\infty, x+h_n]\right) = \lim_{n \rightarrow \infty} P((-\infty, x+h_n]) = \lim_{n \rightarrow \infty} F(x+h_n)$$

zu c Sei $(x_n)_{n \in \mathbb{N}}$ eine Folge mit $x_1 \geq x_2 \geq \dots$ und $\lim_{n \rightarrow \infty} x_n = -\infty$

$$\rightarrow \bigcup_{n=1}^{\infty} (-\infty, x_n] = \mathbb{R} \text{ und es folgt}$$

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P((-\infty, x_n]) = P\left(\bigcup_{n=1}^{\infty} (-\infty, x_n]\right) = P(\mathbb{R}) = 1$$

andere Grenzwert analog

Definition 1.2.2 (Verteilungsdichte) Es sei $[\mathbb{R}, \mathcal{R}, P]$ ein Wahrscheinlichkeitsraum und F die Verteilungsfunktion von P . Falls eine integrierbare Funktion $f : \mathbb{R} \rightarrow [0, \infty)$ existiert, sodass.

$$F(x) = \int_{-\infty}^x f(t) dt \text{ für alle } x \in \mathbb{R}$$

dann heißt f Verteilungsdichte von P . Offenbar ist $\int_{-\infty}^{\infty} f(t) dt = 1$

Bemerkung

Nicht zu jeder Verteilung P auf $[\mathbb{R}, \mathcal{R}]$ existiert eine Dichte. Aber: Zu jeder integrierbaren Funktion $f : \mathbb{R} \rightarrow [0, \infty)$ mit $\int_{-\infty}^{\infty} f(t) dt = 1$ existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P , dass die Verteilungsdichte f besitzt.

1.3 Einige spezielle Wahrscheinlichkeitsräume

1.3.1 diskrete Gleichverteilung

$[\Omega, \mathcal{P}(\Omega), P]$, wobei Ω endlich und $P(\{\omega\}) = \frac{1}{|\Omega|}$, $|\Omega|$... Anzahl der Elemente von Ω

P nennt man die diskrete Gleichverteilung auf Ω . ("klassische Definition der Wahrscheinlichkeit")

Es folgt für beliebige $A \subseteq \Omega$

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}$$

$[\Omega, \mathcal{P}(\Omega), P]$... Laplace'scher Raum der Ordnung

Spezialfälle

- Münzwurf
- Würfeln mit regulärem Würfel
- Lottoziehung

1.3.2 ka Verteilungverteilung

$[\{0, 1\}, \mathcal{P}(\{0, 1\}), P]$ mit $P(\{1\}) = 1 - P(\{0\}) = p$

$0 \leq p \leq 1$, falls $p = \frac{1}{2}$ Spezialfall von 1.

1.3.3 Bernoulli-Verteilung

$[\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n), P]$

$$P(\{(a_1, a_2, \dots, a_n)\}) = p^k (1-p)^{n-k} \text{ für } a_i \in \{0, 1\}, \sum_{i=1}^n a_i = k, 0 \leq p \leq 1$$

Falls $p = \frac{1}{2}$ Spezialfall von (1), Modell für das n-Gledrige Bernoulli-Schema

P ... Bernoulli-Vertielung

p ... "Erfolgswahrscheinlichkeit"

1.3.4 würfeln mit nichtunterscheidbaren Würfeln

$[\Omega, \mathcal{A}, P]$ mit $\Omega = \{(r_1, \dots, r_n); r_i \in \mathbb{N}_0, \sum_{i=1}^n r_i = r\}$, r fest, n fest, $\mathcal{A} = \mathcal{P}(\Omega)$

mögliche Vorstellung: - Würfeln mit nicht unterscheidbaren Würfeln

1.3.5 Physik: Maxwell-Boltzman, Fermi-Dirac, Bose-Einstein Verteilung

- r Teilchen
- für jedes Teilchen wird der Raum der möglichen Zustände diskretisiert in n (Mikro-)Zustände (z.B. Energieniveaus)
- der Zustand des Gesamtsystems wird beschrieben durch ein n -Tupel (r_1, \dots, r_n) der Besetzungszahlen der Zustände

Drei verschiedene Wahrscheinlichkeitsmaße werden in der physikalischen Statistik betrachtet

1. Maxwell-Boltzmann Verteilung

$$P_{MB}(\{r_1, \dots, r_n\}) = \frac{1}{n^r} \cdot \frac{r!}{r_1! \cdot \dots \cdot r_n!}$$

(spezielle Multinomialverteilung, Polynomialverteilung)

2. Fermi-Dirac Verteilung Voraussetzung: $n \geq r$

$$P_{FD}(\{r_1, \dots, r_n\}) = \begin{cases} \binom{n}{r}^{-1}, & \text{falls } r_i \in \{0, 1\} \\ 0, & \text{sonst} \end{cases}$$

3. Bose-Einstein Verteilung

$$P_{BE}(\{r_1, \dots, r_n\}) = \binom{n+r-1}{r}^{-1}$$

also die diskrete Gleichverteilung auf Ω

Beispiele:

MB-Verteilung: zur Modellierung von "klassischen Teilchen, ohne Quanteneffekte" z.B. Molekülen

FD-Verteilung: zur Modellierung von Fermi-Gasen oder Ensembles von Fermionen, halbzahlgiger Spin (Elektronen, Neutronen, protonen, Quarks)

BE-Verteilung: zur Modellierung von Bose-Gasen oder Ensembles von Bosonen, ganzzahliger Spin (Photonen, Kerne, Atome, Mesonen)

Zahlenbeispiel:

n=3, r=2

$\omega \in \Omega$	P_{MB}	P_{FD}	P_{BE}
(2, 0, 0)	$\frac{1}{9}$	0	$\frac{1}{6}$
(0, 2, 0)	$\frac{1}{9}$	0	$\frac{1}{6}$
(0, 0, 2)	$\frac{1}{9}$	0	$\frac{1}{6}$
(1, 1, 0)	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{1}{6}$
(1, 0, 1)	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{1}{6}$
(0, 1, 1)	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{1}{6}$

zu a) Maxwell-Boltzmann ist offenbar keine diskrete Gleichverteilung

Vorstellung: Die r Teilchen werden nacheinander und unabhängig, d.h. ohne rücksicht auf die Zustände der anderen Teilchen, den Zuständen zugeordnet, wobei der gewählte Zustand jeweils gleichverteilt auf der Menge $\{1, \dots, n\}$ ist.

Zur Betrachtung der Wahrscheinlichkeit für das Auftreten einer Konfiguration (r_1, \dots, r_n) , r Teilchen, wird durchnummeriert. Es gibt n^r Möglichkeiten der Zuordnung der Teilchen zu den Zuständen ($\Omega' = \{1, \dots, n\}^r$). Man geht davon aus, dass alle diese Möglichkeiten dieselbe Wahrscheinlichkeit besitzen. Nämlich $\frac{1}{n^r}$.

Anzahl der Möglichkeiten, die auf ein vorgegebenes Tupel (r_1, \dots, r_n) führen.

$$(a_1, \dots, a_n) \in \{1, \dots, n\}^r \mapsto (r_1, \dots, r_n) \in \Omega$$

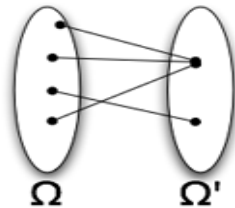


Abbildung 1.1: Tupel addieren sich

Anzahl:

$$\begin{aligned}
 &= \binom{r}{r_1} \cdot \binom{r-r_1}{r_2} \cdot \binom{r-r_1-r_2}{r_3} \cdot \dots \cdot \binom{r-r_1-r_2-\dots-r_{n-1}}{r_n} \\
 &= \frac{r!}{r_1!(r-r_1)!} \cdot \frac{(r-r_1)!}{r_2!(r-r_1-r_2)!} \cdot \dots \cdot \frac{(r-r_1-r_2-\dots-r_{n-1})!}{r_n \cdot 0!} \\
 &= \frac{r!}{r_1! \cdot \dots \cdot r_n!}
 \end{aligned}$$

$$P_{MB}(\{r_1, \dots, r_n\}) = \frac{1}{n^r} \cdot \frac{r!}{r_1! \cdot \dots \cdot r_n!}$$

zu b) Fermi-Dirac

Annahme: pro Zustand höchstens ein Teilchen (Pauli-Prinzip)

$|\{(r_1, \dots, r_n) \in \Omega : r_i \in \{0, 1\}, i = 1, \dots, n\}| = \binom{n}{r}$ | · | ... Anzahl

$\binom{n}{k}$... Anzahl der k-elementigen Teilmenge einer n-elementigen Menge ... $0 \leq k \leq n$

”ziehen ohne Zurücklegen, ohne Berücksichtigung der Reihenfolge”

Fermi-Dirac Verteilung ist die diskrete Gleichverteilung auf der Teilmenge von Ω der ”zulässigen” Konfigurationen

zu c) Bose-einstein Verteilung ist Gleichverteilung auf Ω

Probe dafür, dass die Definition korrekt ist:

zur Bestimmung von Anzahl der Elemente von $\Omega \hat{=} |\Omega|$

Bijektion:

$$h : \Omega \rightarrow \{(b_1, \dots, b_{n+r-1}) : b_i \in \{0, 1\}, i = 1, \dots, n+r-1, \sum_{i=1}^{n+r-1} b_i = 1\}$$

mit $h((r_1, \dots, r_n)) = (b_1, \dots, b_{n+r-1})$

wobei

$$b_i \begin{cases} 0, & \text{für } i = r_1 + 1 \text{ oder } i = r_1 + r_2 + 2 \dots \text{ oder } i = r_1 + r_2 + \dots + r_{n-1} + n - 1 \\ 1, & \text{sonst} \end{cases}$$

Mit:

$$|\Omega| = |\{(h_1, \dots, h_{n+r-1}) : b_i \in \{0, 1\}, \sum_{i=1}^{n+r-1} b_i = r\}| = \binom{n+r-1}{n}$$

1.3.6 geometrische Verteilung

$[\mathbb{N}_0, \mathbb{N}, P]$, $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ Es sei $0 < p < 1$ und $k = 0, 1, 2, \dots$

$$P(\{k\}) = (1-p)^k \cdot p$$

(Nebenrechnung: $\sum_{k=0}^{\infty} (1-p)^k = \frac{1}{p}$)

P heißt geometrische Verteilung mit Parameter p.

1.3.7 Poisson-Verteilung

$[\mathbb{N}_0, \mathbb{N}, \Pi_\lambda]$, $\lambda > 0$

$$\Pi(\{k\}) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

(Nebenrechnung: $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda$)

Π_λ heißt Poisson-Verteilung mit Parameter $\lambda > 0$.

1.3.8 Gleichverteilung

$[\mathbb{R}, \mathcal{R}, U_{(a,b)}]$ mit $-\infty < a < b < \infty$

Wobei $U_{(a,b)}$ gegeben ist durch die Verteilungsdichte:

$$f(t) = \frac{1}{b-a} \cdot 1_{(a,b)} = \begin{cases} \frac{1}{b-a} & \text{für } t \in (a, b) \\ 0 & \text{sonst} \end{cases}$$

Verteilungsfunktion:

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{für } x \leq a \\ \frac{x-a}{b-a} & \text{für } a < x < b \\ 1 & \text{für } x \geq b \end{cases}, x \in \mathbb{R}$$

$U_{(a,b)}$ heißt stetige Gleichverteilung auf dem Intervall $[a, b]$. Für $a < c < d < b$ gilt:

$$U_{(a,b)}((c, d)) = F(d) - F(c) = \frac{d-c}{b-a}$$

d.h. die Wahrscheinlichkeitsverteilung hängt nur von der Länge des Intervalls ab, nicht aber von dessen Lage (Invarianz gegen Translation). Besonders wichtiger Spezialfall $U_{(0,1)}$ wird verwendet bei Zufallsgeneratoren.

1.3.9 Normalverteilung

$[\mathbb{R}, \mathcal{R}, \mathcal{N}_{\mu, \sigma^2}]$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ μ ... Mittelwert, σ^2 ... Varianz
 $\mathcal{N}_{\mu, \sigma^2}$ gegeben durch Verteilungsdichte:

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

Verteilungsfunktion:

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$\mathcal{N}_{\mu, \sigma^2}$ heißt Gaußverteilung (Normalverteilung mit Erwartungswert μ und Varianz σ^2).

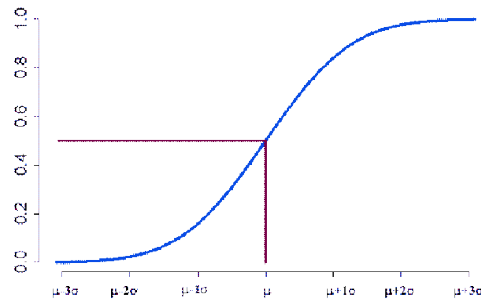


Abbildung 1.2: Verteilungsfunktion der Normalverteilung

1.3.10 Cauchy-Verteilung (Lorenz-Verteilung, Breit-Wiegner-Verteilung)

$[\mathbb{R}, \mathcal{R}, P]$ und P sei gegeben durch die Verteilungsdichte:

$$f(t) = \frac{1}{\pi} \frac{1}{t^2 + 1}, t \in \mathbb{R}$$

mit der Verteilungsfunktion:

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

P heißt (Standard-) Cauchy-/Lorenz-/Breit-Wiegner-Verteilung

Frage:

Wie findet man bei der Modellierung eine geeignete Wahrscheinlichkeitsverteilung auf einem Raum?

- quantitative Eigenschaften z.B.
 - Symmetrie (Münzwurf, Würfel) und auch andere Invarianzeigenschaften
 - Unabhängigkeit (z.B. Bernoulli-Schema)
 - "Gedächtnislosigkeit" (Exponentialverteilung, Markov-Eigenschaften)
- Grenzwertsätze:
 - Zentraler Grenzwertsatz
 - Poissonscher Grenzwertsatz
- Abbildung von Wahrscheinlichkeitsräumen (z.B. Maxwell-Boltzmann Verteilung)
Zusammenhänge zwischen Verteilungen z.B.
 - Bernoulli-Verteilung \rightarrow Binomialverteilung
 - Poisson-Verteilung \rightarrow Exponentialverteilung
- Statistische Methoden zur Prüfung von Modellannahmen (Schätzungen, Tests, insbesondere Anpassungstests)

1.4 Bedingte Wahrscheinlichkeit

Allgemeiner Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P]$, Ereignisse $A, B \in \mathcal{A}$ mit Wahrscheinlichkeit $P(A), P(B)$

1.4.1 Anschauliche Vorstellung:

Unter der Bedingung, dass bekannt ist, dass B eingetreten ist, ändert sich möglicher Weise die Wahrscheinlichkeit für das Eintreten von A.

Beispiel:

Würfeln mit Würfel $[\{1, \dots, 6\}, \mathcal{P}(\{1, \dots, 6\}), P]$, wobei P die diskrete Gleichverteilung $\{1, \dots, 6\}$ ist. $A = \{2\}$, $P(A) = \frac{1}{6}$ und $B = \{2, 4, 6\}$, $P(B) = \frac{1}{2}$

Vorstellung: Wahrscheinlichkeit für A unter der Bedingung, dass B eingetreten ist.

$$\frac{|A \cap B|}{|B|} = \frac{1}{3} = \frac{|A \cap B|}{|\Omega|} \cdot \frac{|\Omega|}{|B|}$$

oder Wahrscheinlichkeit für A unter der Bedingung, dass B nicht eingetreten ist.

$$\frac{|A \cap B^c|}{|B|} = \frac{0}{3} = 0$$

1.4.2 Schritte zur Formalisierung

- "... unter der Bedingung, dass B eingetreten ist"
 - \rightarrow B wird als Ereignis betrachtet, das mit Wahrscheinlichkeit 1 eintritt.
 - $P(B|B) = 1; P(B^c|B) = 0$
- Ereignis A kann nur unter der Bedingung eintreten, dass $A \cap B$ eintritt.

Definition 1.4.1 (bedingte Wahrscheinlichkeit) Es sei $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum, $A, B \in \mathcal{A}$, und $P(B) > 0$. Die bedingte Wahrscheinlichkeit von A unter der Bedingung B ist definiert durch:

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

Es gilt: $P(\cdot|B)$ ist ein Wahrscheinlichkeitsmaß auf $[\Omega, \mathcal{A}]$ ($P(\cdot|B) : \mathcal{A} \rightarrow [0, 1]$).

Nachweis:

- $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(\Omega)}{P(B)} = 1$
- $0 \leq P(A \cap B) \leq P(B) \leq 1 \implies 0 \leq \frac{P(A \cap B)}{P(B)} \leq 1$
- Nachweis der σ -Additivität $P(\cdot|B)$: Es seien $A_1, A_2, \dots \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ für $i \neq j$

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i | B\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} \\ &= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i | B) \end{aligned}$$

Damit:

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

→ Übergang zu einem neuen Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P(\cdot|B)]$

Insbesondere gilt für $A_1, A_2 \in \mathcal{A}$, $A_1 \cap A_2 = \emptyset$:

$$P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$$

Wenn $P(A) > 0$, $P(B) > 0$ dann gilt:

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$\longrightarrow P(B|A) = P(A|B) \cdot \frac{P(B)}{P(A)}$$

Satz 1.4.1 (totale Wahrscheinlichkeit/Bayessche Formel) Es sei $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum und $B_1, \dots, B_n \in \mathcal{A}$ eine Partition (Zerlegung) von Ω (d.h. $\bigcup_{i=1}^n B_i = \Omega$, $B_i \cap B_j = \emptyset, i \neq j$) mit $P(B_i) > 0$ für $i = 1, \dots, n$. Dann gilt für alle $A \in \mathcal{A}$:

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i) \quad \text{Formel der totalen Wahrscheinlichkeit}$$

und falls $P(A) > 0$ auch:

$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{P(A)} = \frac{P(A|B_j) \cdot P(B_j)}{\sum_{i=1}^n P(A|B_i) \cdot P(B_i)}, \forall j = 1, \dots, n \quad \text{Bayessche Formel}$$

Ansatz:

$P(A) = P(A|B) = P(A|B^c)$, falls $P(B) > 0$ und $P(B^c) > 0$

$$\begin{aligned} \rightarrow P(A) \cdot P(B) &= P(A \cap B) \\ P(A) \cdot P(B^c) &= P(A \cap B^c) \end{aligned}$$

Analoge Vorstellung:

B stochastisch unabhängig von A, wenn $P(B) = P(B|A) = P(B|A^c)$, falls $P(A) > 0$ und $P(A^c) > 0$

$$\begin{aligned} \rightarrow P(A) \cdot P(B) &= P(A \cap B) \\ P(A^c) \cdot P(B) &= P(A^c \cap B) \end{aligned}$$

1.5.2 Definition

Definition 1.5.1 (stochastische Unabhängigkeit) Sei $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum

1. Die Ereignisse $A, B \in \mathcal{A}$ heißen stochastisch unabhängig wenn:

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

2. Die Ereignisse $A_1, \dots, A_n \in \mathcal{A}$ heißen stochastisch vollständig unabhängig, wenn für alle $I \subseteq \{1, \dots, n\}, I \neq \emptyset$, gilt:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Bemerkung:

- Begriffe "unvereinbar" und "unabhängig" nicht verwechseln, A, B unvereinbar $\Leftrightarrow A \cap B = \emptyset$. Unabhängigkeit von A und B hängt auch von P ab, z.B. können A, B in $[\Omega, \mathcal{A}, P_1]$ abhängig aber in $[\Omega, \mathcal{A}, P_2]$ unabhängig sein.
- Bedingung in 2.) kann i.A. nicht reduziert werden, insbesondere muss aus Paarweise Unabhängigkeit nicht vollständige Unabhängigkeit folgen. Aus $P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot \dots \cdot P(A_n)$ muss auch nicht die vollständige Unabhängigkeit folgen.

1.5.3 Beispiele

1. $[\{1, \dots, 6\}, \mathcal{P}(\{1, \dots, 6\}), P]$ P diskrete Gleichverteilung
 $A = \{1, 2\}, B = \{2, 4, 6\}, A \cap B = \{2\}$
 $P(A) = \frac{1}{3}, P(B) = \frac{1}{2}, P(A \cap B) = \frac{1}{6}$

$$P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6} = P(A \cap B)$$

$\rightarrow A, B$ stochastisch unabhängig

Merke: $P(A) = \frac{1}{3}$ nun fällt gerade Zahl $\rightarrow P(A) = \frac{1}{3}, \rightarrow P(A)$ ändert sich nicht.

2. n-Gliedriges Bernoulli Schema, $0 \leq p \leq 1$
 $[\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n), P]$ mit

$$P(\{(a_1, \dots, a_n)\}) = p^{\sum_{i=1}^n a_i} \cdot (1-p)^{n-\sum_{i=1}^n a_i}$$

Speziell:

A ... beim 1. Versuch erscheint eine "0", B ... beim 2. Versuch erscheint eine "1"
 $A = \{0\} \times \{0, 1\}^{n-1}$, $B = \{0, 1\} \times \{1\} \times \{0, 1\}^{n-2}$, $A \cap B = \{0\} \times \{1\} \times \{0, 1\}^{n-2}$

$$\begin{aligned} P(A) &= \sum_{(a_1, \dots, a_n) \in A} P(\{(a_1, \dots, a_n)\}) = \sum_{k=0}^{n-1} \sum_{(a_1, \dots, a_n) \in A} P(\{(a_1, \dots, a_n)\}) : \sum_{i=1}^n a_i = k \\ &= \sum_{k=0}^{n-1} \sum_{(a_1, \dots, a_n) \in A} p^k \cdot (1-p)^{n-k} = \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &= (1-p) \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot (1-p)^{n-1-k} = (1-p) \cdot (p + 1-p)^{n-1} \\ &= (1-p) \end{aligned}$$

$\rightarrow P(A) \cdot P(B) = P(A \cap B)$ und damit A,B unabhängig!

Satz 1.5.1 (Bemerkungen zur unabhängigkeit) Es sei $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum und $A, B, A_1, \dots, A_n \in \mathcal{A}$. Dann gilt:

- a) A und \emptyset sind unabhängig
- b) A und Ω sind unabhängig
- c) Wenn A und B unabhängig sind, dann sind auch A und B^c , A^c und B , A^c und B^c unabhängig
- d) Wenn $A \cap B = \emptyset \Leftrightarrow A$ und B unvereinbar, dann gilt: A, B sind unabhängig $\Leftrightarrow P(A) = 0$ oder $P(B) = 0$
- e) Wenn A_1, \dots, A_n vollständig unabhängig sind, dann sind für $m < n$ A_m, \dots, A_n

Kapitel 2

Zufällige Variablen, Zufallsgrößen, zufällige Vektoren

2.1 Zufällige Variablen: meßbare Abbildungen zwischen Wahrscheinlichkeitsräumen

$[\Omega, \mathcal{A}, P]$ und $[\Omega', \mathcal{A}', P']$ mit $\Omega \xrightarrow{g} \Omega'$

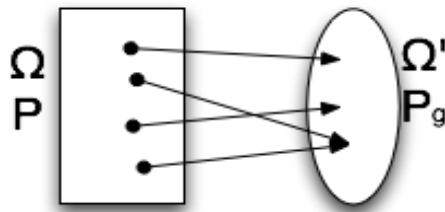


Abbildung 2.1: Abbildung zwischen Wahrscheinlichkeitsräumen

Beispiele:

- zweimaliges Würfeln $(a_1, a_2) \longrightarrow$ vernachlässigen der Reihenfolge (r_1, \dots, r_n)
- zweimaliges Würfeln $(a_1, a_2) \longrightarrow$ Summe der Augenzahlen $a_1 + a_2$
- Betrachtung der Bose-Einstein Verteilung
- Werfen eines Punktes auf eine Fläche: rechtwinklige Koordinaten $(x, y) \longrightarrow$ Polarkoordinaten (r, ϕ)

2.1.1 Formalisierung

Sei $g : \Omega \rightarrow \Omega'$

Ubildfunktion zu g: $g^{-1} : \mathcal{P}(\Omega') \rightarrow \mathcal{P}(\Omega)$ mit $g^{-1}(A') = \{\omega \in \Omega : g(\omega) \in A'\}$ für $A' \subseteq \Omega'$

Definition 2.1.1 (zufällige Variable/ induziertes Wahrscheinlichkeitsmaß) Gegeben seien ein Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P]$ und ein meßbarer Raum $[\Omega', \mathcal{A}', P']$.

- Die Abbildung $g : \Omega \rightarrow \Omega'$ heißt $(\mathcal{A}, \mathcal{A}')$ meßbar oder zufällige Variable, falls $g^{-1}(A') \in \mathcal{A}$ für alle $A' \in \mathcal{A}'$
- Es sei eine $(\mathcal{A}, \mathcal{A}')$ meßbar Abbildung. Dann heißt das Wahrscheinlichkeitsmaß P_g auf $[\Omega, \Omega']$, das gegeben ist durch:

$$P_g(A') = P(g^{-1}(A')) = \text{für alle } A' \in \mathcal{A}'$$

das durch g induzierte Wahrscheinlichkeitsmaß.

Besipiel

Summe der Augenzahl bei zweimaligem Würfeln

$[\{1, \dots, 6\}, \mathcal{P}(\{1, \dots, 6\}), P]$ P diskrete Gleichverteilung

$g((a_1, a_2)) = a_1 + a_2 \Rightarrow [\{2, \dots, 12\}, \mathcal{P}(\{2, \dots, 12\}), P_g]$

$$P_g(\{k\}) = P(\{(a_1, a_2) \in \{1, \dots, 6\}^2 : a_1 + a_2 = k\}) = \frac{|\{(a_1, a_2) : a_1 + a_2 = k\}|}{36}$$

2.2 Zufallsgrößen: reellwertige zufällige Variablen

$[\Omega, \mathcal{A}, P]$ und $[\mathbb{R}, \mathcal{R}, P_x]$ mit $\Omega \xrightarrow{x} \mathbb{R}$

Motiv zur Behandlung von Zufallsgrößen

- Behandlung von Transformationen
- Erleichterung von Sprechweisen, übersichtliche Darstellung zur Formalisierung

Definition 2.2.1 (Zufallsgröße) Es sei $[\Omega, \mathcal{A}, P]$ ein Wahrscheinlichkeitsraum

- Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt reelle Zufallsgröße, wenn $X^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{R}$.
- Das Wahrscheinlichkeitsgesetz der Zufallsgröße X ist das Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathcal{R}]$, das gegeben ist durch

$$P_x(B) = P(X^{-1}(B)), B \in \mathcal{R}$$

- Die Verteilungsfunktion der Zufallsgröße X ist die Funktion: $F_X : \mathbb{R} \rightarrow [0, 1]$ mit

$$F_X(x) = P(X \leq x), x \in \mathbb{R}$$

Andere Schreibweisen:

$$\begin{aligned}
 F_X(x) &= P(X \leq x) \\
 &= P(X \in (-\infty, x]) \\
 &= P(\{\omega \in \Omega : X(\omega) \leq x\}) \\
 &= P \circ X^{-1}((-\infty, x]) \\
 &= P_x((-\infty, x])
 \end{aligned}$$

Wichtige Formeln:

Es sei X eine Zufallsgröße mit Verteilungsfunktion F_X , $a, b \in \mathbb{R}$, $a < b$

$$P(X \leq a) = F_X(a)$$

$$P(X > a) = 1 - F_X(a)$$

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

$$P(X = a) = F_X(a) - F_X(a - 0)$$

2.3 Unabhängigkeit von Zufallsgrößen

Vorstellung:

Zwei Zufallsgrößen X, Y sollen unabhängig heißen, wenn alle Paare von Ergebnissen, die mit Hilfe von X, Y formuliert werden können, voneinander unabhängig sind.

Definition 2.3.1 (Unabhängigkeit von Zufallsgrößen) *Es seien X, Y, X_1, Y_1 Zufallsgrößen aus dem selben Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P]$.*

a) *Zwei Zufallsgrößen X, Y heißen unabhängig, wenn*

$$\forall B_1, B_2 \in \mathcal{R} : P(X \in B_1, Y \in B_2) = P(X \in B_1) \cdot P(Y \in B_2)$$

b) *Die Zufallsgrößen X_1, \dots, X_n heißen vollständig unabhängig, wenn*

$$\forall B_1, \dots, B_n \in \mathcal{R} : P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n)$$

Folgerung 2.3.1 *Wenn X_1, \dots, X_n unabhängig sind und $m < n$, dann sind auch X_1, \dots, X_m unabhängig*

Beweis:

$$\begin{aligned}
 P(X_1 \in B_1, \dots, X_m \in B_m) &= P(X_1 \in B_1, \dots, X_m \in B_m, X_{m+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \\
 &= P(X_1 \in B_1) \cdot \dots \cdot P(X_m \in B_m) \cdot \underbrace{P(X_{m+1} \in \mathbb{R}) \cdot \dots \cdot P(X_n \in \mathbb{R})}_1 \\
 &= P(X_1 \in B_1) \cdot \dots \cdot P(X_m \in B_m)
 \end{aligned}$$

Satz 2.3.1 (Äquivalente Bedingungen für Zufallsgrößen) *Es seien X_1, \dots, X_n Zufallsgrößen (über dem selben Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P]$). X_1, \dots, X_n sind genau dann unabhängig, wenn:*

$$\forall x_1, \dots, x_n \in \mathbb{R} : P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n)$$

Häufig vorkommende Formulierungen

X_1, \dots, X_n i.i.d. (independent identically distributed) d.h unabhängig und identisch verteilt, d.h. $P_{X_1} = P_{X_2} = P_{X_3} = \dots = P_{X_n}$

Beispiel:

n-gliedriges Bernoullischema

$[\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n), P]$ mit

$$P(\{(a_1, \dots, a_n)\}) = p^{\sum_{i=1}^n a_i} \cdot (1-p)^{n-\sum_{i=1}^n a_i}, \text{ für } a \in \{0, 1\}, \quad 0 \leq p \leq 1$$

Definition 2.3.2 (Äquivalente Beschreibung des Bernoullischemas) *Zufallsgrößen X_1, \dots, X_n mit $X_i : \{0, 1\}^n \rightarrow \{0, 1\}$ mit $X_i((a_1, \dots, a_n)) = a_i$, $i = 1, \dots, n$ X_i ... Ergebnis des i-ten Versuchs, reelle Zufallsgröße*

$$\begin{aligned} P(X_i = 1) &= P(X_{i-1}(\{1\})) = P(\{(a_1, \dots, a_n) \in \{0, 1\}^n : a_i = 1\}) \\ &= \sum_{(a_1, \dots, a_n) \in \{0, 1\}^n : a_i = 1} P(\{(a_1, \dots, a_n)\}) \\ &= \sum_{l=0}^{n-1} \sum_{(a_1, \dots, a_n) \in \{0, 1\}^n : \sum_{j \neq i} a_j = l} P(\{(a_1, \dots, a_n)\}) \\ &= \sum_{l=0}^{n-1} \binom{n-1}{l} \cdot p^{l+1} (1-p)^{n-l-1} \\ &= p \cdot (p + 1 - p)^{n-1} = p \end{aligned}$$

$$P(X_i = 0) = 1 - P(X_i = 1) = 1 - p$$

Verteilungsfunktion: $F_{X_i} : \mathbb{R} \rightarrow [0, 1]$

Untersuchen Unabhängigkeit von X_1, \dots, X_n

Seien $B_1, \dots, B_n \in \mathcal{R}$ - setzen: $B'_i = B_i \cap \{0, 1\}$, $i = 1, \dots, n$

$$\begin{aligned}
 P(X_1 \in B_1, \dots, X_n \in B_n) &= P(X_1 \in B'_1, \dots, X_n \in B'_n) \\
 &= P(\{(a_1, \dots, a_n) : a_1 \in B_1, \dots, a_n \in B_n\}) \\
 &= P(B'_1 \times B'_2 \times \dots \times B'_n) \\
 &= \sum_{(a_1, \dots, a_n) \in (B'_1 \times B'_2 \times \dots \times B'_n)} P(\{(a_1, \dots, a_n)\}) \\
 &= \sum_{a_1 \in B'_1} \dots \sum_{a_n \in B'_n} P(\{(a_1, \dots, a_n)\}) \\
 &= \sum_{a_1 \in B'_1} \dots \sum_{a_n \in B'_n} [p^{a_1}(1-p)^{1-a_1}] \cdot \dots \cdot [p^{a_n}(1-p)^{1-a_n}] \\
 &= \left(\sum_{a_1 \in B'_1} p^{a_1}(1-p)^{1-a_1} \right) \cdot \dots \cdot \left(\sum_{a_n \in B'_n} p^{a_n}(1-p)^{1-a_n} \right) \\
 &= P(X_1 \in B'_1) \cdot \dots \cdot P(X_n \in B'_n) \\
 &= P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n)
 \end{aligned}$$

$\Rightarrow X_1, \dots, X_n$ sind vollständig unabhängig

Fazit:

Das n-gliedrige Bernoullischema kann auch beschrieben werden durch

$$X_1, \dots, X_n \text{ i.i.d. mit } P(X_i = 1) = 1 - P(X_i = 0) = p, \quad i = 1, \dots, n, \quad p \in [0, 1]$$

2.4 Diskrete Zufallsgrößen

2.4.1 Formalisierung

Definition 2.4.1 (diskrete Zufallsgröße) Eine Zufallsgröße heißt diskret, wenn es eine höchstens abzählbare unendliche Menge $W_X \subset \mathbb{R}$ gibt mit $P(X \in W_X) = 1$

Verteilungsfunktion einer diskreten Zufallsgröße

X , W_X höchstens abzählbar unendlich $F_X : \mathbb{R} \rightarrow [0, 1]$, $x \in \mathbb{R}$

$$\begin{aligned}
 F_X(x) &= P(X \leq x) = P(X \in \{x_k \in W_X : x_k \leq x\}) \\
 &= P\left(\bigcup_{k: x_k \leq x} \{\omega \in \Omega, X(\omega) = x_k\}\right) \\
 &= \sum_{x_k \leq x} P(X = x_k) \\
 &= \sum_{x_k \in W_X} P(X = x_k) 1_{[x_k, \infty)}(x) \rightarrow F_x \text{ ist eine Linearkombination}
 \end{aligned}$$

Die Verteilungsfunktion und damit das Verteilungsgesetz von diskreten Zufallsgrößen ist also

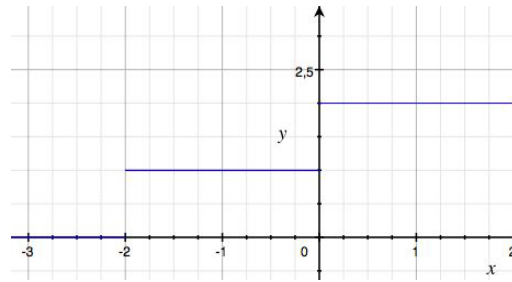


Abbildung 2.2: Verteilungsfunktion einer diskreten Zufallsgröße

gegeben durch $\{(x_k, P(X = x_k)) : x_k \in W_X\}$

Satz 2.4.1 (Unabhängigkeit) *Es seien X, Y diskrete Zufallsgrößen, W_X, W_Y höchstens abzählbar unendlich und $P(X \in W_X) = P(Y \in W_Y) = 1$. Die Zufallsgrößen sind genau dann unabhängig, wenn gilt:*

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \text{ für alle } x \in W_X, y \in W_Y$$

2.4.2 Beispiele

$[\Omega, \mathcal{A}, P]$ sei ein Wahrscheinlichkeitsraum

1. $X : \Omega \rightarrow \mathbb{R}$

$$P(X = 1) = 1 - P(X = 0) = p, \quad 0 \leq p \leq 1$$

z.B. $\Omega = \{1, \dots, 6\}$

$$X(k) = \begin{cases} 1 & \text{falls } k = 6 \\ 0 & \text{sonst} \end{cases}$$

2. X ... Anzahl der Erfolge bei n-gliedrigem Bernoullischema

Seien X_1, \dots, X_n i.i.d. mit $P(X_i = 1) = 1 - P(X_i = 0) = p$

$X = \sum_{i=1}^n X_i$, wähle $W_X = \{0, \dots, n\}$ für $k \in \{0, \dots, n\}$

$$\begin{aligned} P(X = k) &= P(X^{-1}\{k\}) = P(\{(a_1, \dots, a_n) \in \{0, 1\}^n : \sum_{i=1}^n a_i = k\}) \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

Schreibweise

$$B_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n$$

$X \sim B_{n,p}$ bzw. $P_X = B_{n,p}$... Binomialverteilung mit Parameter n und p

Satz 2.4.2 (Poisson'scher Grenzwertsatz) Es sein $(p_n)_{n \in \mathbb{N}}$ eine Zahlenfolge mit $0 \leq p_n \leq 1$ und $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda < \infty$. Dann gilt für

$$\lim_{n \rightarrow \infty} B_{n,p_n}(\{k\}) = \prod_{\lambda}(\{k\}) \text{ für } k = 0, 1, 2, \dots$$

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

Beweis:

für $n \in \mathbb{N}_0$

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} &= \lim_{n \rightarrow \infty} \frac{1}{k!} (n \cdot p_n)^k \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{n^k} \left(1 - \frac{n \cdot p_n}{n}\right)^n \left(1 - \frac{n \cdot p_n}{n}\right)^{-k} \\ &= \frac{1 \cdot \lambda^k}{k!} \cdot \underbrace{1 \cdot 1 \cdot \dots \cdot 1}_{k\text{-mal}} e \cdot 1^{-k} \end{aligned}$$

Wichtige Anwendung für den Poisson'schen Grenzwertsatz: Falls n "groß" und p "klein" und $\lambda = n \cdot p$ gesetzt wird, dann kann $B_{n,p}$ durch \prod_{λ} approximiert werden.

$$B_{n,p}(\{k\}) \approx \frac{(n \cdot p)^k}{k!} \cdot e^{-n \cdot p}$$

Zusammenhang zu Satz 2.4.2:

Wenn X_1, \dots, X_n i.i.d. mit $P(X_i = 1) = 1 - P(X_i = 0) = p$, n "groß", p "klein", dann ist $X = \sum_{i=1}^n X_i$ annähernd Poissonverteilt mit Parameter $\lambda = n \cdot p$

3. Beispiele:

- Tod durch Hufschlag in der Preußischen Armee
- Radioaktiver Zerfall
Vorstellung: große Anzahl n von Atomen, feste Zeitintervalle der Länge t
Annahme: Wahrscheinlichkeit, dass ein bestimmtes Teilchen im Zeitintervall zerfällt sei $p = \text{const}$
→ vollständige Unabhängigkeit der Ereignisse: i -tes Teilchen zerfällt, $i = 1, \dots, n$

Formalisierung:

X_1, \dots, X_n

$$X_i = \begin{cases} 0 & \dots \text{ i-tes Teilchen zerfällt} \\ 1 & \dots \text{ i-tes Teilchen zerfällt nicht} \end{cases}$$

X_1, \dots, X_n i.i.d. mit $P(X_i = 1) = 1 - P(X_i = 0) = p$

$X = \sum_{i=1}^n X_i$... Anzahl der zerfallenen Teilchen

Approximation der Binomialverteilung durch Poissonverteilung mit Parameter $\lambda = n \cdot p$.

$$P(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

4. Sei X_1, \dots, X_n eine Folge von i.i.d. Zufallsgrößen mit $P(X_i = 1) = 1 - P(X_i = 0) = p$, $0 \leq p \leq 1$ (unendliches Bernoullischema mit Erfolgswahrscheinlichkeit p)

Neu Zufallsgrößen:

$$X = \min\{n \in \{0, 1, 2, \dots\} : X_{n+1} = 1\}$$

$$\text{Für } k \in \{0, 1, 2, \dots\} = W_X$$

$$\begin{aligned} P(X = k) &= P(X_1 = 0, X_2 = 0, \dots, X_n = 0, X_{n+1} = 1) \\ &= P(X_1 = 0) \cdot P(X_2 = 0) \cdot \dots \cdot P(X_n = 0) \cdot P(X_{n+1} = 1) \\ &= (1 - p)^k \cdot p \end{aligned}$$

P_x ist die geometrische Verteilung mit Parameter $P \in \{0, 1\}$.

Da $\sum_{k=0}^{\infty} P(X = k) = 1$ folgt $P(X = \infty) = 0$

Satz 2.4.3 (geometrische Verteilung) Es sei X geometrisch verteilt mit Parameter $p \in \{0, 1\}$, dann gilt für alle $k, l \in \{0, 1, 2, \dots\}$

$$P(X = k + l | X \geq k) = P(X = l)$$

Bemerkung:

Die geometrische Verteilung wird verwendet als Verteilung des Zustandes (des Frequenzniveaus eines harmonischen Oszillators) im Gleichgewichtszustand.

2.5 Stetige Zufallsgrößen

2.5.1 Formalisierung

Definition 2.5.1 (stetige Zufallsgröße) Eine Zufallsgröße heißt stetige Zufallsgröße falls ihr Verteilungsgesetz P_X eine Verteilungsdichte f_X besitzt, wobei eine integrierbare Funktion $f_X : \mathbb{R} \rightarrow [0, \infty)$ existiert, sodass

$$P(X \leq x) = F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \text{für alle } x \in \mathbb{R}$$

Eigenschaften:

Sei X eine stetige Zufallsgröße und f_X eine Verteilungsdichte von X . Dann gilt:

$$\int_{-\infty}^{\infty} f_X(t) dt = \lim_{x \rightarrow \infty} \int_{-\infty}^x f_X(t) dt = \lim_{x \rightarrow \infty} F_X(x) = 1$$

$$\text{Für } a < b: P(a < X < b) = P_X((a, b]) = F_X(b) - F_X(a) = \int_a^b f_X(t) dt$$

$$P(X = a) = P_X(\{a\}) = F_X(a) - \lim_{n \rightarrow \infty} F_X(a - \frac{1}{n}) = \lim_{n \rightarrow \infty} \int_{a - \frac{1}{n}}^a f_X(t) dt = 0$$

$$\rightarrow P(X = a) = 0 \text{ für alle } x \in \mathbb{R}$$

Damit:

- Wenn $B \subseteq \mathbb{R}$ endlich oder abzählbar unendlich (z.B. Menge der rationalen Zahlen), dann gilt $P(x \in B) = 0$
- F_X hat keine Sprungstellen, ist also stetig
- für alle $a < b$ gilt: $P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b)$

Mit Hauptsatz der Differential- und Integralrechnung

- Die Zufallsgröße X besitzt genau dann eine stetige Dichtefunktion f_X , wenn ihre Verteilungsfunktion stetig differenzierbar ist und falls X eine stetige Dichtefunktion f_X besitzt, gilt: $\int f(x)dx = F_X$
- Achtung: $f_x(t)$ ist keine Wahrscheinlichkeit!
Aber: Das Differential $f_X(t)dt$ kann als Wahrscheinlichkeit für $x \in (t, t + dt)$ interpretiert werden.

2.5.2 Wichtige Spezialfälle

1. Die Zufallsgröße X heißt gleichverteilt auf dem Intervall (a, b) , $a, b \in \mathbb{R}, a < b$, wenn das Verteilungsgesetz von X : $P_X = U(a, b)$. Schreibweise: $X \sim U(a, b)$
z.B. X ... zufällig aus (a, b) ausgewählter Punkt in der Ebene $X \sim [0, 2\pi)$, $P_X = U(a, b)$

Satz 2.5.1 (Verteilungsfunktion) Es sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Funktion mit den Eigenschaften (1), (2), (3) aus Satz 1.2.1 (d.h. F ist eine Verteilungsfunktion). Es sei X eine Zufallsgröße mit $X \sim U(0, 1)$

- (a) Wenn F streng monoton und stetig ist, dann besitzt die Zufallsgröße $Y = F^{-1}(x)$ die Verteilungsfunktion $F_Y = F$ (wobei F^{-1} die inverse Funktion von F ist)
- (b) Die Zufallsgröße $Y = F^{-1}(x) = \sup\{x \in \mathbb{R} : F(x) < x\}$ besitzt die Verteilungsfunktion $F_Y = F$.

2. Die Zufallsgröße X heißt normalverteilt (gaußverteilt) mit Erwartungswert $\mu \in \mathbb{R}$ und Varianz $\sigma^2 > 0$, wenn $P_X = \mathcal{N}_{\mu, \sigma^2}$, $X \sim \mathcal{N}_{\mu, \sigma^2}$
Verteilungsdichte: $f_X(X) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Nachweis, dass dies eine Dichte ist:

$f_X \geq 0$ klar!

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}, \quad \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \text{ mit Substitution: } t = \frac{x-\mu}{\sigma}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \dots \text{Verteilungsfunktion der Standardnormalverteilung } (\mu = 0, \sigma^2 = 1)$$

Umrechnungsformel:

Sei $X \sim \mathcal{N}_{\mu, \sigma^2}$

$$\begin{aligned} F_X(x) = P(X \leq x) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad \text{Substitution: } s = \frac{t-\mu}{\sigma}, \frac{dt}{ds} = \sigma \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

und für $a < b$:

$$P(a < X < b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Folgerung:

Wenn $X \sim \mathcal{N}_{\mu, \sigma^2}$ dann ist $\frac{X-\mu}{\sigma} \sim \mathcal{N}_{0,1}$

Beweis:

Für $x \in \mathbb{R}$ gilt

$$\begin{aligned} F_{\frac{X-\mu}{\sigma}}(x) &= P\left(\frac{X-\mu}{\sigma} \leq x\right) = P(X \leq \sigma x + \mu) \\ &= F_X(\sigma x + \mu) = \Phi\left(\frac{\sigma x + \mu - \mu}{\sigma}\right) = \Phi(x) \end{aligned}$$

3. Die Zufallsgröße X heißt exponentialverteilt mit Parameter $\lambda > 0$, wenn sie die folgende Verteilungsdichte besitzt:

$$f_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ \lambda \cdot e^{-\lambda x} & \text{für } x \geq 0 \end{cases}$$

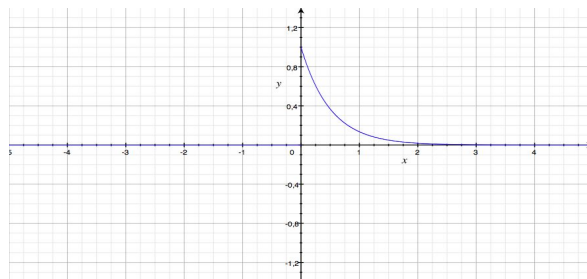


Abbildung 2.3: Verteilungsdichte einer Exponentialverteilung

Verteilungsfunktion:

$$\int_{-\infty}^x f_X(t)dt = \int_{-\infty}^x \lambda \cdot e^{-\lambda t} 1_{[0,\infty)}(t)dt = \begin{cases} 0 & \text{falls } x < 0 \\ \int_0^x \lambda \cdot e^{-\lambda t} dt & \text{falls } x \geq 0 \end{cases} = (1 - e^{-\lambda x}) \cdot 1_{[0,\infty)}(x)$$

Schreibweise: $X \sim \mathcal{E}_\lambda$

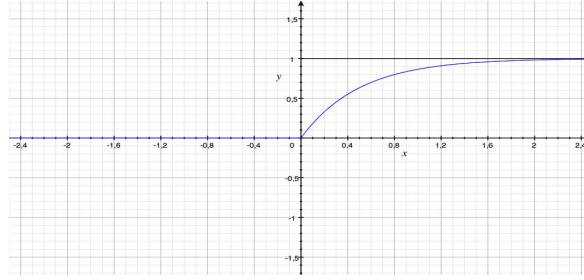


Abbildung 2.4: Verteilungsfunktion einer Exponentialverteilung

Satz 2.5.2 (nicht-alter Eigenschaft, Gedächtnislosigkeit) Sei $X \sim \mathcal{E}_\lambda$, $\lambda > 0$. Dann gilt für alle reellen $s, t > 0$

$$P(X > s + t | X > s) = P(X > t)$$

Beweis:

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda(s)})} = e^{-\lambda t} \\ &= 1 - (1 - e^{-\lambda t}) = P(X > t) \end{aligned}$$

2.6 Zufällige Vektoren

2.6.1 Formalisierung

Definition 2.6.1 (zufälliger Vektor) Es sei $[\Omega, \mathcal{U}, P]$ ein Wahrscheinlichkeitsraum und X_1, \dots, X_n reelle Zufallsgrößen über diesem Raum. Dann heißt der Vektor $\underline{X} = (X_1, \dots, X_n)$ zufälliger Vektor über $[\Omega, \mathcal{U}, P]$ mit den Koordinaten X_1, \dots, X_n

Schreibweise:

$[\Omega, \mathcal{U}, P]$ und $[\mathbb{R}, \mathcal{R}, P_{X_i}]$ mit $\Omega \xrightarrow{X_i} \mathbb{R}, i = 1, \dots, n$

$[\Omega, \mathcal{U}, P]$ und $[\mathbb{R}^n, \mathcal{R}_n, P_X]$ mit $\Omega \xrightarrow{X} \mathbb{R}^n$

wobei $X(\omega) = (X_1(\omega), \dots, X_n(\omega)), \omega \in \Omega$

$\mathcal{R}_n = \sigma(\{(-\infty, x_1] \times \dots \times (-\infty, x_n] : x_i \in \mathbb{R}, i = 1, \dots, n\})$

Für $A \in \mathcal{R}_n$:

$$P_{\underline{X}}(A) = P \circ X^{-1}(A) = P(X^{-1}(A)) = P(X \in A) = P(\{\omega \in \Omega : \underline{X}(\omega) \in A\})$$

speziell für $A = B_1 \times \dots \times B_n, B_i \in \mathcal{R}_i$

$$P_{\underline{X}}(B_1 \times \dots \times B_n) = P(X_1 \in B_1, \dots, X_n \in B_n)$$

$$P(\{\omega \in \Omega : X_1(\omega) \in B_1, \dots, X_n \in B_n\}) = P\left(\bigcap_{i=1}^n \{\omega \in \Omega : X_i(\omega) \in B_i\}\right)$$

heißt Verteilungsgesetz $P_{\underline{X}}$ des zufälligen Vektors \underline{X} / Verteilungsgesetz der Zufallsgrößen X_1, \dots, X_n
 Anwendung: bei Beobachtungen/ Messungen mehrerer Merkmale an einem Objekt (z.B. Druck, Temperatur)

Definition 2.6.2 (Randverteilung) Es sei $\underline{X} = X_1, \dots, X_n$ ein zufälliger Vektor. Die Randverteilung (marginal Distribution) von $P_{\underline{X}}$ sind die Verteilungsgesetze P_{X_i} der einzelnen Koordinaten.

Verallgemeinerter Randwert aus Wert von Vektoren

(X_{i1}, \dots, X_{im}) mit $1 \leq i \leq \dots \leq im \leq n, m < n$.

Lemma 2.6.1 Alle Randverteilungen sind durch $P_{\underline{X}}$ eindeutig bestimmt.

Beweis:

Sei $i \in \{1, \dots, n\}, B \in \mathcal{R}$. Dann gilt

$$P_{X_i}(B) = P(X_i \in B) = P(X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i \in B, X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) = P_{\underline{X}}(\mathbb{R}^{i-1} \times B \times \mathbb{R}^{n-i-1})$$

Satz 2.6.1 (Eindeutigkeit der Randverteilung) Wenn X_1, \dots, X_n unabhängige Zufallsgrößen sind, dann ist das Verteilungsgesetz $P_{\underline{X}}$ des zufälligen Vektors $\underline{X} = (X_1, \dots, X_n)$ durch die Randverteilung eindeutig bestimmt und es gilt für alle $B_1, \dots, B_n \in \mathcal{R}$:

$$P_{\underline{X}}(B_1 \times B_2 \times \dots \times B_n) = P_{X_1}(B_1) \cdot \dots \cdot P_{X_n}(B_n)$$

Beispiel

	\underline{X}				
\underline{Y}	$X = 1$	$X = 2$	$X = 3$	$X = 4$	
$Y = 0$	$\frac{1}{4}$	$\frac{1}{12}$		$\frac{2}{12}$	$\frac{1}{2}$
$Y = 1$		$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{2}$
	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$

Allgemeiner:

Randverteilung eines zweidimensionalen diskreten Vektors $\underline{X} = \{X, Y\}$

Seien W_X, W_Y endlich oder abzählbar unendlich mit $P(X \in W_X) = 1, P(Y \in W_Y) = 1$

Bezeichnung: $P_{ij} = P(X = x_i, Y = y_j)$ für $x_i \in W_X, y_j \in W_Y, P_i = P(X = x_i), P_j = P(Y = y_j)$

Es gilt:

$$P(X = x_i) = P(X = x_i, Y \in W_Y) = \sum_{y_j \in W_Y} P(X = x_i, Y = y_j)$$

$$P(Y = y_j) = P(X \in W_X, Y = y_j) = \sum_{x_i \in W_X} P(X = x_i, Y = y_j)$$

In Kurzform: $P_i = \sum_j P_{ij}$, $P_j = \sum_i P_{ij}$

Definition 2.6.3 (gemeinsame Verteilungsfunktion) Es sei $\underline{X} = (X_1, \dots, X_n)$ ein zufälliger Vektor. Unter der gemeinsamen Verteilungsfunktion von \underline{X} versteht man die Funktion $F_{\underline{X}} : \mathbb{R} \rightarrow [0, 1]$, die gegeben ist durch

$$F_{\underline{X}}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) \text{ mit } (x_1, \dots, x_n) \in \mathbb{R}$$

Die Verteilungsfunktion F_{X_i} der Koordinaten X_i , $i = 1, \dots, n$ heißen Randverteilungsfunktion von X_i .

Satz 2.6.2 (Gleichheit von Wahrscheinlichkeiten zufälliger Vektoren) Wenn \underline{X} , \underline{Y} n -dimensionale zufällige Vektoren sind mit $F_{\underline{X}} = F_{\underline{Y}}$, dann gilt $P_{\underline{X}} = P_{\underline{Y}}$.

Bestimmung der Randverteilung aus gemeinsamer Verteilungsfunktion

$$F_{X_1} = P(X_1 \leq x_1) = P(X_1 \leq x_1, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R})$$

$$= \lim_{x_2 \rightarrow \infty}, \dots, \lim_{x_n \rightarrow \infty} P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

Anwendung von Satz 2.4.1 liefert:

Es sei $\underline{X} = (X_1, \dots, X_n)$ ein zufälliger Vektor. Die Koordinaten X_1, \dots, X_n sind genau dann unabhängig, wenn

$$F_{\underline{X}}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n) \quad \text{für alle } (x_1, \dots, x_n) \in \mathbb{R}^n$$

Spezielle Klasse: stetige zufällige Vektoren

Definition 2.6.4 (Dichtefunktion, Verteilungsdichte)

a) Eine Funktion $f : \mathbb{R}^n \rightarrow [0, \infty)$ heißt Dichtefunktion auf \mathbb{R} , wenn f integrierbar ist und

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \dots dt_n = 1$$

b) Der zufällige Vektor $\underline{X} = (X_1, \dots, X_n)$ besitzt die gemeinsame Verteilungsdichte $f_{\underline{X}}$, wenn $f_{\underline{X}}$ eine Dichtefunktion ist und

$$F_{\underline{X}}(x_1, \dots, x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\underline{X}}(t_1, \dots, t_n) dt_1 \dots dt_n \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

Für $A \in \mathcal{R}_n$:

$$P_{\underline{X}}(A) = P(\underline{X} \in A) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} 1_{[A]} \cdot f_{\underline{X}}(t_1, \dots, t_n) dt_1 \dots dt_n$$

Folgerung 2.6.1 Wenn der zufällige Vektor $\underline{X} = (X_1, \dots, X_n)$ eine gemeinsame Verteilungsdichte $f_{\underline{X}}$ besitzt, dann besitzen auch alle Koordinaten X_i eine Verteilungsdichte f_{X_i} und es gilt

$$f_{X_i} = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{n-1 \text{ Integrale}} f_{\underline{X}}(t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_n) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_n$$

Beweis:

$i = 1, x \in \mathbb{R}$

$$\begin{aligned} F_{X_1}(x) &= P(X_1 \leq x) = \lim_{x_2 \rightarrow \infty} \dots \lim_{x_n \rightarrow \infty} F(x, x_2, \dots, x_n) \\ &= \lim_{x_2 \rightarrow \infty} \dots \lim_{x_n \rightarrow \infty} \int_{-\infty}^x \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_X(t_1, \dots, t_n) dt_n \dots dt_2 dt_1 \\ &= \int_{-\infty}^x \left(\underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(t_1, \dots, t_n) dt_2 \dots dt_n}_{\text{ist eine Verteilungsdichte von } X_1} \right) dt_1 \end{aligned}$$

Es gilt: Es sei $\underline{X} = (X_1, \dots, X_n)$ ein zufälliger Vektor mit der Randverteilungsdichte f_{X_i} der Koordinaten X_i , $i = 1, \dots, n$. X_1, \dots, X_n sind genau dann unabhängig, wenn $f : \mathbb{R} \rightarrow [0, \infty)$ mit

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

eine Verteilungsdichte von \underline{X} ist.

2.6.2 Wichtige Spezialfälle

1. Gleichverteilter Punkt auf achsenparallelem Rechteck

Sei $Q = (a, b) \times (c, d) | a, b, c, d \in \mathbb{R} \ a < b, c < d$ und es sei $\underline{X} = (X_1, X_2)$ ein zufälliger Vektor mit gemeinsamer Verteilungsdichte.

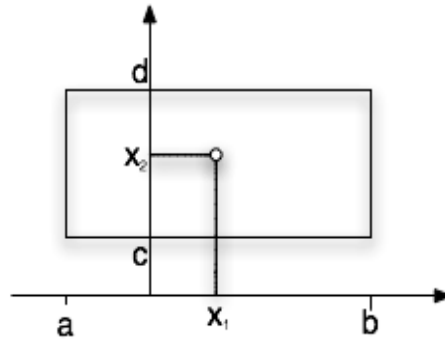


Abbildung 2.5: Punkt auf Rechteck

$$\begin{aligned} f_{\underline{X}}(x_1, x_2) &= \frac{1}{(b-a) \cdot (d-c)} \cdot 1_{(a,b) \times (c,d)}(x_1, x_2) \\ &= \frac{1}{(b-a)} \cdot 1_{(a,b)}(x_1) \frac{1}{(d-c)} \cdot 1_{(c,d)}(x_2) \end{aligned}$$

Randdichte:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{\underline{X}}(x_1, x_2) dx_2 = \int_c^d \frac{1_{(a,b)}(x_1)}{(b-a)(d-c)} = \frac{1}{b-a} \cdot 1_{(a,b)}(x_1)$$

Analog:

$$f_{X_2}(x_2) = \frac{1}{d-c} \cdot 1_{(c,d)}(x_2)$$

Also: $f_{\underline{X}} = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$ für alle $(x_1, x_2) \in \mathbb{R}^2$

Damit folgt $\underline{X} = (X_1, X_2)$ ist genau dann gleichverteilt auf dem achsenparallelen Rechteck $Q = (a, b) \times (c, d)$, wenn die kartesische Koordinaten unabhängig und jeweils gleichverteilt auf (a, b) bzw. (c, d) sind.

2. Gleichverteilter Punkt auf Kreis

$$B_r = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < r^2\}, r > 0$$

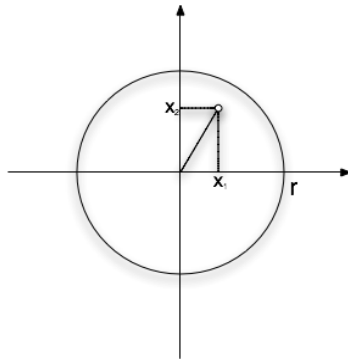


Abbildung 2.6: Punkt auf Kreis

- kartesische Koordinaten

Sei $\underline{X} = (X_1, X_2)$ ein zufälliger Vektor mit gemeinsamer Verteilungsdichte

$$f_{\underline{X}}(x_1, x_2) = \frac{1}{\pi r^2} \cdot 1_B(x_1, x_2)$$

Randdichte:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{\underline{X}}(x_1, x_2) dx_2 = \frac{1}{\pi r^2} \cdot 1_{(-r, r)}(x_1) \int_{-\sqrt{r^2 - x_1^2}}^{\sqrt{r^2 - x_1^2}} dx_2 = \frac{2}{\pi r^2} \sqrt{r^2 - x_1^2} \cdot 1_{(-r, r)}(x_1)$$

$f_{X_2}(x_2) = f_{X_1}(x_1)$ (Rotationssymmetrie)

$\Rightarrow f_{X_1}(x_1) \cdot f_{X_2}(x_2)$ ist keine Verteilungsdichte von \underline{X} .

$0 = P(x_1 \in I_1, x_2 \in I_2) \neq P(x_1 \in I_1) \cdot P(x_2 \in I_2) > 0 \Rightarrow$ nicht unabhängig

- Polarkoordinaten

$$B_r = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 < r^2\} = \{(\rho \cos \alpha, \rho \sin \alpha) : 0 \leq \rho \leq r, 0 \leq \alpha \leq 2\pi\}$$

$\underline{X} = (X_1, X_2)$ sei gleichverteilt auf B_r

Man hat $[\Omega, \mathcal{A}, P]$, $[\mathbb{R}^2, \mathcal{R}_2, P_{X_1, X_2}]$ und $[\mathbb{R}^2, \mathcal{R}_2, P_{X_1, X_2}]$

mit $\Omega \xrightarrow{\underline{X} = (X_1, X_2)} \mathbb{R}^2$ und $\mathbb{R}^2 \xrightarrow{\underline{X} = (R, \varphi)} \mathbb{R}^2$

Gemeinsame Verteilungsfunktion des zufälligen Vektors $(R\varphi)$ auf $[0, \infty) \times [0, 2\pi)$,
 $t \in [0, \infty), \beta \in [0, 2\pi)$:

$$\begin{aligned}
 F_{(R,\varphi)}(t, \beta) &= P(R \leq t, \varphi \leq \beta) = P((X_1, X_2) \in (x_1, x_2) \in \mathbb{R} : \underbrace{R(x_1, x_2) \leq t, \varphi(x_1, x_2) \leq \beta}_{s(t, \beta)}) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X_1, X_2)}(t_1, t_2) \cdot 1_{s(t, \beta)}(t_1, t_2) dt_1 dt_2 \\
 \text{Substitution : } \begin{cases} t_1 = \rho \cos \alpha \\ t_2 = \rho \sin \alpha \end{cases} \\
 &= \int_0^{\infty} \int_0^{2\pi} f_{(X_1, X_2)}(\rho \cos \alpha, \rho \sin \alpha) \cdot 1_{[0, t]}(\rho) \cdot 1_{[0, \beta]}(\alpha) \rho d\alpha d\rho \\
 &= \int_0^t \int_0^{\beta} (f_{(X_1, X_2)}(\rho \cos \alpha, \rho \sin \alpha) \rho) d\alpha d\rho
 \end{aligned}$$

Also Verteilungsdichte von $(R\varphi)$:

$$f_{R,\varphi}(\rho, \alpha) = \rho \cdot f_{(X_1, X_2)}(\rho \cos \alpha, \rho \sin \alpha) \cdot 1_{[0, \infty)}(\rho) \cdot 1_{[0, 2\pi)}(\alpha)$$

Hier speziell: $f_{(X_1, X_2)}(x_1, x_2) = \frac{1}{\pi r^2} \cdot 1_{B_r}(x_1, x_2)$

$$f_{R,\varphi}(\rho, \alpha) = \rho \frac{1}{\pi r^2} \cdot 1_{[0, r)}(\rho) \cdot 1_{[0, 2\pi)}(\alpha) = \frac{2\rho}{r^2} 1_{[0, r)}(\rho) \frac{1}{2\pi} 1_{[0, 2\pi)}(\alpha)$$

Damit: Polarkoordinaten (R, φ) eines auf B_r gleichverteilten Punktes sind unabhängig und φ ist gleichverteilt auf $[0, 2\pi)$ und R besitzt die Verteilungsdichte $f_R(\rho) = \frac{2\rho}{r^2} 1_{[0, r)}(\rho)$.

3. Zweidimensionale Normalverteilung

Vorbetrachtung: Seien X_1, X_2 unabhängig, $X_i \sim \mathcal{N}_{0,1}, i = 1, 2$. Dann ist eine gemeinsame Verteilungsdichte von $\underline{X} = (X_1, X_2)$

$$f_{\underline{X}} = f_{X_1}(x_1) \cdot f_{X_2}(x_2) = \frac{1}{2\pi} \cdot e^{-\frac{x_1^2 + x_2^2}{2}}$$

Allgemeiner: X_1, X_2 unabhängig, $X_1 \sim \mathcal{N}_{\mu_1, \sigma_1^2}$ und $X_2 \sim \mathcal{N}_{\mu_2, \sigma_2^2}$

$$f_{\underline{X}} = f_{X_1}(x_1) \cdot f_{X_2}(x_2) = \frac{1}{2\pi(\sigma_1^2 \sigma_2^2)} \cdot e^{-\frac{1}{2} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right)}$$

Definition 2.6.5 (zweidimensionale Normalverteilung) Der zufällige Vektor $\underline{X} = (X_1, X_2)$ ist zweidimensional normalverteilt mit Erwartungswert (μ_1, μ_2) , dem Vektor (σ_1^2, σ_2^2) der Varianz und dem Korrelationskoeffizient $\rho \in (-1, 1)$, wenn er folgende Verteilungsdichte besitzt:

$$\begin{aligned}
 f_{\underline{X}} &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \\
 &= \frac{1}{2\pi\sigma_1^2\sigma_2^2\sqrt{1-\rho^2}} \cdot \exp \left(-\frac{1}{2\sqrt{1-\rho^2}} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right) \right) \\
 (x_1, x_2) &\in \mathbb{R}^2
 \end{aligned}$$

Berechnung der Randdichte:

Vorbereitung: für $a > 0, b \in \mathbb{R}$:

$$\int_{-\infty}^{\infty} e^{-a(x^2+bx)} dx = e^{\frac{ab^2}{4}} \cdot \sqrt{\frac{\pi}{a}} \quad \text{mit Substitution: } y = \sqrt{2a}(x + \frac{b}{2})$$

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{\underline{X}}(x_1, x_2) dx_2 \quad \text{Substitution: } y_2 = \frac{x_2 - \mu_2}{\sigma_2} \\ &= \frac{1}{2\sigma_1 \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \frac{(x_1 - \mu_1)^2}{\sigma_1^2}\right) \int_{-\infty}^{\infty} \exp\left(\frac{1}{2(1-\rho^2)} \left(-2\rho \frac{x_1 - \mu_1}{\sigma_2} y_2 + y_2^2\right)\right) dy_2 \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}} \Rightarrow X_1 \sim \mathcal{N}_{\mu_1, \sigma_1^2} \end{aligned}$$

analog: $X_2 \sim \mathcal{N}_{\mu_2, \sigma_2^2}$

Folgerung 2.6.2 Wenn $\underline{X} = (X_1, X_2)$ eine 2-D Normalverteilung mit $(\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2), \rho$ besitzt, dann sind X_1, X_2 genau dann unabhängig, wenn $\rho = 0$.

Satz 2.6.3 (Polarkoordinaten eines Standardnormalverteilten Vektors)

Der Vektor $\underline{X} = (X_1, X_2)$ besitze eine 2-D Normalverteilung. Für den Vektor $(R; \varphi)$ seiner Polarkoordinaten gilt dann:

- R, φ sind unabhängig
- φ ist gleichverteilt auf $[0, 2\pi)$
- R besitzt die Verteilungsdichte $f_R(\rho) = \rho e^{-\frac{\rho^2}{2}} 1_{[0, \infty)}(\rho)$.

Beweis:

Aus Satz 2.6.3 und 2.6.2 folgt:

$$\begin{aligned} f_{(R, \varphi)}(\rho, \alpha) &= \rho \frac{1}{2\pi} e^{-\frac{\rho^2}{2}} \cdot 1_{[0, \infty)}(\rho) 1_{[0, 2\pi)}(\alpha) \\ &= \underbrace{\frac{1}{2\pi} 1_{[0, 2\pi)}(\alpha)}_{f_\varphi} \underbrace{\rho \cdot e^{-\frac{\rho^2}{2}} \cdot 1_{[0, \infty)}(\rho)}_{f_R} \end{aligned}$$

Anwendung dieses Satzes: Box-Müller-Verfahren zur Simulation von normalverteilten Zufallsgrößen.

4. n-dimensionale Normalverteilung

Definition 2.6.6 (n-dimensionale Normalverteilung) Der zufällige Vektor $\underline{X} = (X_1, \dots, X_n)$ besitzt eine reguläre n-dimensionale Normalverteilung mit Erwartungswert $\underline{\mu} \in \mathbb{R}^n$ mit regulärer (n, n) -Matrix Σ , wenn er folgende Verteilungsdichte $f_{\underline{X}}$ besitzt:

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} \cdot e^{-\frac{1}{2}(\underline{x} - \underline{\mu})^T \Sigma^{-1}(\underline{x} - \underline{\mu})} \quad \underline{x} \in \mathbb{R}^n$$

Man kann zeigen:

- Alle Randverteilungen sind normalverteilt: $X_i \sim \mathbb{N}_{\mu_i, \sigma_i^2}$, $\underline{\mu} = (\mu_1, \dots, \mu_n)^T$, σ_i^2 ... i-tes Diagonalelement von Σ
- X_1, \dots, X_n sind stochastisch unabhängig $\Leftrightarrow \Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix}$

Für $n=3$, Aussage die auf MAXWELL zurück geht (Bezug zur Geschwindigkeit von Molekülen eines idealen Gases im \mathbb{R}^3 , vgl. FISZ, S.193/194)

Satz 2.6.4 () Es sei $\underline{X} = (X_1, X_2, X_3)$ ein zufälliger Vektor mit Verteilungsdichte $f_{\underline{X}}$ und differenzierbare Randdichte. Wenn X_1, X_2, X_3 stochastisch unabhängig sind und eine differenzierbare Funktion $h : [0, \infty) \rightarrow (0, \infty)$ existiert, sodass $f_{\underline{X}}(x_1, x_2, x_3) = h(x_1^2 + x_2^2 + x_3^2)$ für alle $(x_1, x_2, x_3) \in \mathbb{R}^3$, dann besitzt \underline{X} eine dreidimensionale Normalverteilung.

Beweis:

Wegen Unabhängigkeit kann gesetzt werden

$$f_{\underline{X}}(x_1, x_2, x_3) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot f_{X_3}(x_3) = h(x_1^2 + x_2^2 + x_3^2)$$

$$\rightarrow \ln h(x_1^2 + x_2^2 + x_3^2) = \sum_{i=1}^3 \ln f_{X_i}(x_i) \text{ Ableiten liefert}$$

$$2 \frac{h'(x_1^2 + x_2^2 + x_3^2)}{h(x_1^2 + x_2^2 + x_3^2)} \cdot x_i = \frac{f'_{X_i}(x_i)}{f_{X_i}(x_i)} \text{ für } i = 1, 2, 3$$

$$2 \frac{h'(\|x\|)}{h(\|x\|)} = \frac{f'_{X_1}(x_1)}{x_1 f_{X_1}(x_1)} = \frac{f'_{X_2}(x_2)}{x_2 f_{X_2}(x_2)} = \frac{f'_{X_3}(x_3)}{x_3 f_{X_3}(x_3)} = a = \text{const}$$

$$\rightarrow f_{X_i}(x_i) = c \cdot e^{\frac{1}{2} a x_i}, i = 1, 2, 3, a < 0, \text{ setze } a = -\frac{1}{\sigma^2}, \sigma > 0$$

$$\rightarrow c = \frac{1}{\sqrt{2\pi\sigma}} \text{ und damit } f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{x_i^2}{2\sigma^2}}$$

$$h(x_1^2 + x_2^2 + x_3^2) = f_{\underline{X}}(x_1, x_2, x_3) = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma^3} \cdot e^{-\frac{1}{2\sigma^2} (x_1^2 + x_2^2 + x_3^2)}$$

Satz 2.6.5 (Bedingung für Normalverteiltetheit)

Es seien $\underline{X} = (X_1, X_2), \underline{Y} = (Y_1, Y_2)$ zufällige Vektoren und $A = a_{ij}$ eine reguläre (2,2)-Matrix mit $a_{ij} \neq 0$ für alle i, j . Wenn X_1, X_2 unabhängig und $X^T = A \cdot Y^T$ und Y_1, Y_2 unabhängig sind, dann sind X_1, X_2, Y_1, Y_2 normalverteilt.

Folgerung 2.6.3 Falls X_1, X_2 unabhängig sind und wenigstens ein Winkel $\alpha \neq k \cdot \frac{\pi}{2}$, $k = 1, 2, 3, \dots$ existiert, sodass der zufällige Vektor $\underline{Y} = (Y_1, Y_2)$, der durch Drehung von $\underline{X} = (X_1, X_2)$ um α entsteht, unabhängige Koordinaten besitzt, dann sind X_1, X_2, Y_1, Y_2 normalverteilt.

Beweis:

Setze in 2.6.5: $A = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}$

Kapitel 3

Weitere Verteilungsgesetze von transformierten zufälligen Vektoren

3.1 Transformation von eindimensionalen Zufallsgrößen

Sei X eine reelle Zufallsgröße und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine (meßbare) Funktion, dann ist $g(X)$ eine Zufallsgröße.

$[\Omega, \mathcal{A}, P]$ und $[\mathbb{R}, \mathcal{R}, P_X]$ mit $\Omega \xrightarrow{\quad} \mathbb{R}$
 $[\mathbb{R}, \mathcal{R}, P_X]$ und $[\mathbb{R}, \mathcal{R}, P_{g(X)}]$ mit $\Omega \xrightarrow{g(x)} \mathbb{R}$

Bereits betrachtet:

- $F(x)^{-1}$ für $X \sim U(0, 1)$
- $\frac{X-\mu}{\sigma}$ für $X \sim \mathcal{N}_{\mu, \sigma^2}$
- logarithmische Normalverteilung

Prinzip zur Berechnung von Verteilungsfunktion von $g(X)$:

Für $x \in \mathbb{R}$:

$$F_{g(X)}(x) = P(X \leq x) = P(g(X) \in (-\infty, x]) = P(X \in g^{-1}(-\infty, x])$$

falls $g^{-1}((-\infty, x])$ als Vereinigung von Intervallen darstellbar ist, dann kann $F_{g(X)}$ durch F_X dargestellt werden.

Falls eine Darstellung gefunden werden kann mit:

$$F_{g(X)}(x) = \int_{-\infty}^x h(t) dt \text{ für alle } x \in \mathbb{R}$$

dann besitzt $g(X)$ die Verteilungsdichte h .

3.2 Summe zweier Zufallsgrößen

$$\underline{X} = (X_1, X_2), g(X) = X_1 + X_2$$

3.2.1 diskreter Fall

Es seien W_1, W_2 höchstens abzählbar unendliche Mengen $P(X_1 \in W_1) = P(X_2 \in W_2) = 1$. Dann ist $X_1 + X_2$ wieder diskret mit $P(X_1 + X_2 \in W_S) = 1$ mit $W_S = \{X_1 + X_2 : x_1 \in W_1, x_2 \in W_2\}$.

Für $s \in W_S$:

$$\begin{aligned} P(X_1 + X_2 = s) &= \sum_{x_1 \in W_1} \sum_{x_2 \in W_2} 1_{\{s\}}(x_1 + x_2) P(X_1 = x_1, X_2 = x_2) \\ &= \sum_{x_1 \in W_1} P(X_1 = x_1, X_2 = s - x_1) \\ &= \sum_{x_1 \in W_1} P(X_1 = x_1) \cdot P(X_2 = s - x_1 | X_1 = x_1) \\ &P(X_1 = x_1) > 0 \end{aligned}$$

Daraus folgt speziell:

Satz 3.2.1 (Summe zweier diskreter Zufallsgrößen) *Es seien X_1, X_2 unabhängige diskrete Zufallsgrößen, dann gilt:*

$$P(X_1 + X_2 = s) = \sum_{x_1 \in W_1} P(X_1 = x_1) \cdot P(X_2 = s - x_1) \text{ für alle } s \in \mathbb{R}$$

Beispiel:

$X_1 \sim \prod_{\lambda_1}, X_2 \sim \prod_{\lambda_2}$ und $\lambda_1, \lambda_2 > 0$

Voraussetzung: X_1, X_2 unabhängig $W_1 = W_2 = \mathbb{N}_0 = \{0, 1, 2, \dots\}$

Für $s \in \mathbb{N}_0$:

$$\begin{aligned} P(X_1 + X_2 = s) &= \sum_{k=0}^{\infty} P(X_1 = k) \cdot P(X_2 = s - k) \\ &= \sum_{k=0}^s \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_1} \cdot \frac{\lambda_2^{s-k}}{(s-k)!} \cdot e^{-\lambda_2} \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{s!} \sum_{k=0}^s \binom{s}{k} \lambda_1^k \lambda_2^{s-k} \\ &= \frac{(\lambda_1 + \lambda_2)^s}{s!} \cdot e^{-(\lambda_1 + \lambda_2)} \end{aligned}$$

$$\Rightarrow X_1 + X_2 \sim \prod_{\lambda_1 + \lambda_2}$$

3.2.2 stetiger Fall

Sei $\underline{X} = (X_1, X_2)$ ein stetiger zufälliger Vektor mit Verteilungsdichte

$$\begin{aligned}
 F_{X_1+X_2}(x) &= P(X_1 + X_2 \leq x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[-\infty, x]}(t_1 + t_2) f_{\underline{X}}(t_1, t_2) dt_2 dt_1 \\
 &\quad \text{Substitution: } s = t_1 + t_2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[-\infty, x]}(s) f_{\underline{X}}(t_1, s - t_2) ds dt_1 \\
 &= \int_{-\infty}^x \left(\int_{-\infty}^{\infty} f_{\underline{X}}(t_1, s - t_2) dt_1 \right) ds
 \end{aligned}$$

$X_1 + X_2$ besitzt die Verteilungsdichte:

$$f_{X_1+X_2}(s) = \int_{-\infty}^{\infty} f_{\underline{X}}(t_1, s - t_2) dt_1 \text{ für } s \in \mathbb{R}$$

Satz 3.2.2 (Summe zweier stetiger Zufallsgrößen)

Es seien X_1, X_2 unabhängige, stetige Vektoren mit Verteilungsdichte f_{X_1}, f_{X_2} . Dann besitzt die Summe $X_1 + X_2$ die Verteilungsdichte:

$$f_{X_1+X_2}(s) = \int_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(s - t) dt \text{ für } s \in \mathbb{R}$$

Beispiel:

$X_1 \sim \mathcal{N}_{\mu_1, \sigma_1^2}, X_2 \sim \mathcal{N}_{\mu_2, \sigma_2^2}$ und X_1, X_2 sind unabhängig

Sei $\mu = \mu_1 + \mu_2$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2$

Für $s \in \mathbb{R}$:

$$\begin{aligned}
 f_{X_1+X_2}(s) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{t-\mu_1}{2\sigma_1^2}} \cdot e^{-\frac{t-\mu_2}{2\sigma_2^2}} dt \quad \text{Substitution: } u = t - \mu_1 \\
 &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{(s-\mu)^2}{2\sigma_2^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{\sigma^2}{2\sigma_1^2\sigma_2^2} u \cdot \left(u - 2\frac{\sigma_1^2(s-\mu)}{\sigma_2^2}\right)\right) du \\
 &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot e^{-\frac{(s-\mu)^2}{2\sigma_2^2}} \cdot e^{\frac{4 \cdot \sigma^2 \sigma_1^2 (s-\mu)^2}{2\sigma_1^2 \sigma_2^2 \sigma^4}} \cdot \sqrt{\frac{2\pi\sigma_1^2\sigma_2^2}{\sigma^2}} \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(s-\mu)^2}{2\sigma^2}}
 \end{aligned}$$

$$\Rightarrow X_1 + X_2 \sim \mathcal{N}_{\mu_1+\mu_2, \sigma_1^2+\sigma_2^2}$$

Es gilt sogar, wenn X_1, X_2 unabhängig und $X_1 + X_2$ normalverteilt, dann sind auch X_1, X_2 normalverteilt.

Wichtige weitere Familie von Verteilungen:

Definition 3.2.1 (Gamma-Verteilung) Die Zufallsgröße X ist gammaverteilt mit Parameter (a, b) , $a > 0, b > 0$, wenn sie folgende Verteilungsdichte besitzt:

$$f_X(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \cdot e^{-bx} \cdot 1_{(0,\infty)}(x)$$

Schreibweise: $X \sim \Gamma_{a,b}$; $P_X = \Gamma_{a,b}$

Speziell: $\mathcal{E}_\lambda = \Gamma_{1,2}$

Satz 3.2.3 (Summe zweier Gammaverteilungen) Es seien X_1, X_2 unabhängige Zufallsgrößen mit $X_i \sim \Gamma_{a_i,b}$, $a_1, a_2, b > 0$. Dann gilt:

$$X_1 + X_2 \sim \Gamma_{a_1+a_2,b}$$

Beweis:

Für $s \leq 0$: $f_{X_1+X_2}(s) = \int_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(s-t) dt = 0$

Für $s > 0$:

$$\begin{aligned} f_{X_1+X_2}(s) &= \int_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(s-t) dt \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1) \cdot \Gamma(a_2)} \cdot e^{-b \cdot s} \int_0^s t^{a_1-1} (s-t)^{a_2-1} dt \\ \text{Substitution: } u &= \frac{t}{s}, \quad dt = s du \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1) \cdot \Gamma(a_2)} e^{-b \cdot s} \cdot \int_0^1 (su)^{a_1-1} (s-su)^{a_2-1} s du \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1) \cdot \Gamma(a_2)} e^{-b \cdot s} \cdot s^{a_1+a_2-1} \cdot \underbrace{\int_0^1 u^{a_1-1} (1-u)^{a_2-1} du}_{\text{EULERSCHE-}\beta\text{-Funktion}(a_1, a_2) = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+a_2)}} \\ &= \frac{b^{a_1+a_2}}{\Gamma(a_1+a_2)} \cdot s^{a_1+a_2-1} \cdot e^{-bs} \end{aligned}$$

Folgerung 3.2.1 Seien X_1, \dots, X_n i.i.d. mit $X_i \sim \mathcal{E}_\lambda > 0$, dann gilt:

$$\sum_{i=1}^n X_i \sim \Gamma_{n,\lambda}$$

$\Gamma_{n,\lambda}$... heißt ERLANG-Verteilung n -ter Stufe mit Parameter λ .

(X_1, X_2) zufälliger Vektor $X_1 - X_2 = X_{1+(-X_2)}$

$$\begin{aligned} F_{-X_2}(x) &= P(-X_2 \leq x) = P(X_2 \geq -x) = 1 - P(X_2 < -1) \\ &= 1 - F_{X_2}(-x) + P(X_2 = -x) \end{aligned}$$

und $f_{X_1, -X_2}(x_1, x_2) = f_{X_1, X_2}(x_1, -x_2)$

\Rightarrow Wenn $\underline{X} = (X_1, X_2)$ zufälliger Vektor mit Verteilungsdichte $f_{\underline{X}}$ ist, dann:

$$f_{X_1 - X_2}(s) = \int_{-\infty}^{\infty} f_{\underline{X}}(t, t - s) dt$$

falls X_1, X_2 unabhängig dann:

$$f_{X_1 - X_2}(s) = \int_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(t - s) dt$$

3.3 Produkt und Quotient zweier Zufallsgrößen

Hier nur stetiger Fall für $x \in \mathbb{R}$

$$\begin{aligned} F_{X_1 \cdot X_2}(x) &= P(X_1, X_2 \leq x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{(-\infty, x]}(t_1, t_2) f_{\underline{X}}(t_1, t_2) dt_1 dt_2 \\ &\text{im inneren Integral } s = t_1 \cdot t_2, ds = t_2 dt_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{(-\infty, x]}(s) f_{\underline{X}}\left(\frac{s}{t_2}, t_2\right) \cdot \frac{1}{|t_2|} |t_2| ds dt_2 \end{aligned}$$

$\Rightarrow X_1 \cdot X_2$ besitzt die Verteilungsdichte

$$f_{X_1 \cdot X_2}(s) = \int_{-\infty}^{\infty} \frac{1}{|t|} f_{\underline{X}}\left(\frac{s}{t}, t\right) dt$$

Analog $\frac{X_1}{X_2}$ besitzt die Verteilungsdichte

$$f_{\frac{X_1}{X_2}}(s) = \int_{-\infty}^{\infty} |t| f_{\underline{X}}(s \cdot t, t) dt$$

Satz 3.3.1 (Produkt und Quotient zweier Zufallsgrößen) Es seien X_1, X_2 unabhängige Zufallsgrößen mit Verteilungsdichte f_{X_1}, f_{X_2} , dann sind die Funktionen:

$$f_{\frac{X_1}{X_2}}(s) = \int_{-\infty}^{\infty} |t| f_{\underline{X}}(s \cdot t, t) dt$$

$$f_{X_1 \cdot X_2}(s) = \int_{-\infty}^{\infty} \frac{1}{|t|} f_{\underline{X}}\left(\frac{s}{t}, t\right) dt$$

Verteilungsdichten von $X_1 \cdot X_2$ bzw $\frac{X_1}{X_2}$ für $s \in \mathbb{R}$.

3.4 Injektive differenzierbare Transformationen von zufälligen Vektoren

Satz 3.4.1 () Es sei $\underline{X} = (X_1, \dots, X_n)$ ein n -dimensionaler zufälliger Vektor mit Verteilungsdichte $f_{\underline{X}}$ und $V \subset \mathbb{R}^n$ eine offene Menge mit $f_{\underline{X}}(x) = 0$ für $x \notin V$. Weiter sei $T : V \rightarrow \mathbb{R}^n$ stetig differenzierbar und injektiv und es sei die Funktionaldeterminante $\det T'(x) \neq 0$ für alle $x \in V$. Dann besitzt der zufällige Vektor $\underline{Y} = T(\underline{X})$ die Verteilungsdichte

$$f_{\underline{Y}}(\underline{u}) = \frac{f_{\underline{X}}(T^{-1}(\underline{u}))}{|\det T'(T^{-1}(\underline{u}))|}$$

für alle $\underline{u} \in T(V)$ und sonst $f_{\underline{Y}}(\underline{u}) = 0$.

Beweis:

Für Borelmenge $B \subset T(V)$ gilt:

$$\begin{aligned} P(\underline{Y} \in B) &= P(T(\underline{X}) \in B) \\ &= P(\underline{X} \in T^{-1}(B)) \\ &= \int_{\mathbb{R}^n} 1_{T^{-1}(B)}(\underline{x}) \cdot f_{\underline{X}}(\underline{x}) d\underline{x} \\ &= \int_{\mathbb{R}^n} 1_B(T(\underline{x})) \cdot f_{\underline{X}}(\underline{x}) d\underline{x} \\ \text{Substitution: } \underline{u} &= T(\underline{x}) \rightarrow \underline{x} = T^{-1}(\underline{u}) \\ \det \frac{\partial \underline{x}}{\partial \underline{u}} &= \frac{1}{\det(T'(T^{-1}(\underline{u})))} \\ &= \int_{\mathbb{R}^n} 1_B(\underline{u}) \frac{f_{\underline{X}}(T^{-1}(\underline{u}))}{|\det(T'(T^{-1}(\underline{u})))|} d\underline{u} \end{aligned}$$

Beispiele:

Transformation kartesisches Koordinaten \rightarrow Polarkoordinaten
Transformation bei logarithmischer Normalverteilung

Beispiel: Affine Abbildungen

Es sei A eine reguläre (n,n) -Matrix, $\underline{b} \in \mathbb{R}^n$. $T(\underline{x}) = A\underline{x} + \underline{b}$
Dann hat der Vektor \underline{Y}^T mit $\underline{Y} = A\underline{x} + \underline{b}$ die Dichte:

$$f_{\underline{Y}}(\underline{u}) = f_{\underline{X}}(A^{-1}(\underline{u} - \underline{b})) \cdot \frac{1}{|\det A|}$$

Speziell für $|\det A| = 1$ und $\underline{b} = 0$:

$$f_{\underline{Y}}(\underline{u}) = f_{\underline{X}}(A^{-1}\underline{u})$$

für $\underline{u} \in \mathbb{R}^n$ (Drehung oder Drehspiegelung)

Folgerung 3.4.1 *Es sei $\underline{X} = (X_1, \dots, X_n)$ normalverteilt mit Erwartungswert $\underline{\mu} \in \mathbb{R}^n$ und regulärer Kovarianz-Matrix Σ .*

a) *Wenn A eine reguläre (n, n) -Matrix ist und $\underline{b} \in \mathbb{R}^n$, dann ist $Y^T = AX^T + b^T$ normalverteilt mit Erwartungswert $\underline{\mu}_Y^T = A\underline{\mu}^T + b$ und Kovarianz-Matrix $\Sigma_Y = A\Sigma A^T$*

b) *Wenn X_1, \dots, X_n i.i.d. standardnormalverteilt, dann ist $\underline{Y}^T = A \cdot X^T + b^T$ normalverteilt mit $\mu_Y = \underline{b}$ und $\Sigma_Y = A \cdot A^T$.*

c) *Es existiert eine orthogonale Matrix A (d.h. $A^T = A^{-1}$), sodass $Y^T = AX^T$ normalverteilt ist mit $\mu_Y^T = A\underline{\mu}^T$ und $\Sigma_Y = \begin{bmatrix} \tilde{\sigma}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{\sigma}_n^2 \end{bmatrix} = \text{diag}(\tilde{\sigma}_1^2, \dots, \tilde{\sigma}_n^2)$, d.h. es existiert eine Drehung bzw. Drehung und Spiegelung von \underline{X} , sodass die Koordinaten von \underline{Y} unabhängig sind.*

d) *Es existiert eine reguläre Matrix \tilde{A} , sodass*

$$\underline{Y}^T = \tilde{A}(X^T - \underline{\mu}^T)$$

normalverteilt mit $\mu_Y = \underline{0}$ und $\Sigma_Y = I_n$.

Nachweis d):

Für A aus c) setzten: $\tilde{A} = \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\tilde{\sigma}_n^2} \end{bmatrix} \cdot A$

Kapitel 4

Erwartungswert, Varianz und Kovarianz

4.1 Vorbemerkung

Bisher: vollständige beschreibung des Verteilungsgesetzes einer Zufallsgröße, Verteilungsfunktion:

F_X

bzw $\begin{cases} \text{Einzelwahrscheinlichkeit } P(X = x_k), x_k \in W_X \text{ falls X diskret ist} \\ \text{Verteilungsdichte: } f_X, \text{ falls X stetig ist} \end{cases}$

Jetzt:

- wichtige quantitative Merkmale des Verteilungsgesetzes
- wenige numerische Parameter sollen Vorstellung von P_X vermitteln
⇒ Informationsdichte (Informationsverlust)
- inhaltliche Deutung der Parameter von Verteilungsfamilien

Beispiel:

X ... Monatseinkommen einer zufällig ausgewählten Person eines Landes

P_X, F_X beschreiben das Verteilungsgesetz

Spezielle Parameter:

- Durchschnittseinkommen → Erwartungswert
- mittlere quadratische Abweichung des Einkommens vom Durchschnittseinkommen → Varianz/Streuung
- $P(X \leq a)$ für einen bestimmten Wert a → Wert der Verteilungsfunktion an der Stelle a
- Zahl c, für die $P(X > c) = \alpha \rightarrow \alpha$ – Quantil z.B.
 - $\alpha = 0.1 \rightarrow$ Dezantil
 - $\alpha = 0.01 \rightarrow$ Perzentil
 - $\alpha = 0.5 \rightarrow$ Median

4.2 Erwartungswert einer Zufallsgröße

Vorstellung: Mittelwert - nicht: "der erwartete Wert"

Beispiel: Radioaktiver Zerfall

n ... Gesamtzahl der Versuche

n_i ... Anzahl der Versuche, bei denen i Teilchen zerfallen sind, $i = 1, 2, 3, \dots$

→ Gesamtzahl der zerfallenen Teilchen $\sum_{i=1}^{\infty} i \cdot n_i$

mittlere Anzahl der zerfallenen Teilchen pro Versuch: $\frac{1}{n} \sum_{i=1}^{\infty} i \cdot n_i = \sum_{i=1}^{\infty} i \cdot \frac{n_i}{n}$

Sei X ... Anzahl der zerfallenen Teilchen bei zufällig ausgewähltem Versuch (gemäß der Gleichverteilung auf $\{1, \dots, n\}$)

$$P(X = i) = \frac{n_i}{n}$$

→ Mittlere Teilchenzahl pro Versuch: $\sum_{i=1}^{\infty} i \cdot P(X = i)$

oder folgende Vorstellung: X' ... zufällige Anzahl der zerfallenen Teilchen bei einem Versuch

$$P(X' = i) = p_i \approx \frac{n_i}{n}$$

(relative Häufigkeit \approx Wahrscheinlichkeit)

Mittlere Anzahl der Zerfälle in einem Versuch:

$$\sum_{i=1}^{\infty} i \cdot \frac{n_i}{n} \approx \sum_{i=1}^{\infty} i \cdot P(X' = i)$$

Definition 4.2.1 (Erwartungswert)

a) Es sei X eine diskrete Zufallsgröße. Falls

$$\sum_{x_k \in W_X} |x_k| \cdot P(X = x_k) < \infty$$

dann heißt

$$\mathbb{E}X = \sum_{x_k \in W_X} x_k \cdot P(X = x_k)$$

Erwartungswert von X .

b) Es sei X eine stetige Zufallsgröße mit Verteilungsdichte f_X - Falls

$$\int_{-\infty}^{\infty} |x| \cdot f(x) dx < \infty$$

dann heißt

$$\mathbb{E}X = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

Erwartungswert von X .

- Schreibweise in der Physik: $\bar{X} = \mathbb{E}X = \langle X \rangle$
- falls f_X als Massenverteilung auf \mathbb{R} interpretiert wird, dann ist $\mathbb{E}X$ gerade die Koordinate des Schwerpunktes
- falls eine endliche Menge W_X existiert mit $P(X \in W_X) = 1$, dann existiert $\mathbb{E}X$ immer
- falls X nicht negativ, d.h. $P(X \geq 0) = 1$, dann wird auch $\mathbb{E}X = \infty$ zugelassen, d.h. $\mathbb{E}X$ existiert dann immer

Beispiele:

- a) $X \sim \Pi_\lambda, \lambda > 0, X$ nicht negativ \rightarrow Existenz gesichert

$$\begin{aligned}
 \mathbb{E}X &= \sum_{k=0}^{\infty} k \cdot P(X = k) \\
 &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\
 &= \lambda \cdot e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
 &= \lambda \cdot e^{-\lambda} \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{(k')!} = \lambda
 \end{aligned}$$

- b) $X \sim \mathcal{N}_{\mu, \sigma^2}, \mu \in \mathbb{R}, \sigma^2 > 0$
Existenz:

$$\begin{aligned}
 \int_{-\infty}^{\infty} |x| f_X(x) dx &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \left(- \int_{-\infty}^0 x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \int_0^{\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right) \\
 \text{Substitution: } t &= \frac{x-\mu}{\sigma}, dx = \sigma dt \\
 &= \frac{1}{\sqrt{2\pi}} \left(- \int_{-\infty}^{-\frac{\mu}{\sigma}} (\sigma t + \mu) e^{-\frac{t^2}{2}} dt + \int_{-\frac{\mu}{\sigma}}^{\infty} (\sigma t + \mu) e^{-\frac{t^2}{2}} dt \right) \\
 &= \frac{1}{\sqrt{2\pi}} \left(\underbrace{\sigma e^{-\frac{\mu^2}{2\sigma^2}} - \mu \int_{-\infty}^{-\frac{\mu}{\sigma}} e^{-\frac{t^2}{2}} dt}_{\mu\sqrt{2\pi}} + \underbrace{\sigma e^{-\frac{\mu^2}{2\sigma^2}} + \mu \int_{-\frac{\mu}{\sigma}}^{\infty} e^{-\frac{t^2}{2}} dt}_{\mu\sqrt{2\pi}} \right) < \infty
 \end{aligned}$$

$$\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) e^{-\frac{t^2}{2}} dt \\
&= \frac{1}{\sqrt{2\pi}} \left(\underbrace{\sigma \int_{-\infty}^{\infty} t \cdot e^{-\frac{t^2}{2}} dt}_0 + \mu \underbrace{\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt}_{\sqrt{2\pi}} \right) = \mu
\end{aligned}$$

Eigenschaften des Erwartungswertes

- Wenn X_1, X_2 Zufallsgrößen mit $P_{X_1} = P_{X_2}$ und $\mathbb{E}X_1$ existiert, dann existiert auch $\mathbb{E}X_2$ und es gilt: $\mathbb{E}X_1 = \mathbb{E}X_2$.

Satz 4.2.1 () *Es sei X eine Zufallsgröße und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine messbare Funktion.*

a) *Falls X diskret ist und $\mathbb{E}g(X)$ existiert, dann gilt:*

$$\mathbb{E}g(X) = \sum_{x_k \in W_X} g(x_k) \cdot P(X = x_k)$$

b) *Falls X stetig ist mit der Verteilungsdichte f_X und $\mathbb{E}g(X)$ existiert, dann gilt:*

$$\mathbb{E}g(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Beweis: nur für a)

Wenn $\mathbb{E}g(x)$ existiert, dann

$$\begin{aligned}
\mathbb{E}g(X) &= \sum_{y_j \in W_{g(X)}} y_j \cdot P(g(X) = y_j) \\
&= \sum_{y_j \in W_{g(X)}} y_j \cdot P(X \in g^{-1}(\{y_j\})) \\
&= \sum_{y_j} y_j \sum_{x_k \in W_X : g(x_k) = y_j} P(X = x_k) \\
&= \sum_{y_j} \sum_{x_k \in W_X : g(x_k) = y_j} g(x_k) \cdot P(X = x_k) \\
&= \sum_{x \in W_X} g(x_k) \cdot P(X = x_k)
\end{aligned}$$

Wichtiger Spezialfall: $g(X) = X^m$

$\mathbb{E}X^m$... m-tes Moment von X (falls es existiert)

$g(X) = (X - \mathbb{E}X)^m$

$\mathbb{E}(X - \mathbb{E}X)^m$... m-tes zentrales Moment

Satz 4.2.2 (Erwartungswert von Summe Produkt zweier Zufallsgrößen) *Es seien X_1, X_2 Zufallsgrößen, deren Erwartungswert existiert.*

- a) Für alle $a, b \in \mathbb{R}$ existiert $\mathbb{E}(aX_1 + b)$ und es gilt $\mathbb{E}(aX_1 + b) = a\mathbb{E}X_1 + b$.
- b) Es existiert $\mathbb{E}(X_1 + X_2)$ und es gilt $\mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2$.
- c) Falls X_1, X_2 unabhängig sind, dann existiert $\mathbb{E}(X_1 \cdot X_2)$ und es gilt $\mathbb{E}(X_1 \cdot X_2) = \mathbb{E}X_1 \cdot \mathbb{E}X_2$.

Beweis nur teilweise:

zu a) Sei X stetig mit Verteilungsdichte f_X auf $\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx < \infty$

Für $g(X) = aX + b$ gemäß Satz 4.2.2 a) :

$$\text{Existenz: } \int_{-\infty}^{\infty} |ax + b| \cdot f_X(x) dx \leq |a| \int_{-\infty}^{\infty} |x| f_X(x) dx + |b| \underbrace{\int_{-\infty}^{\infty} f_X(x) dx}_{=1} < \infty$$

$$\mathbb{E}(aX + b) = \int_{-\infty}^{\infty} (aX + b) f_X(x) dx = a\mathbb{E}X + b$$

zu b) Sei $X = (X_1, X_2)$ stetig mit gemeinsamer Verteilungsdichte f_X

$$\text{und } \int_{-\infty}^{\infty} |x| \underbrace{f_{X_i}(x)}_{\text{Randdichte}} dx < \infty, i = 1, 2$$

Betrachte: $g(X_1, X_2) = X_1 + X_2$

Existenz:

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x_1 + x_2| f_X(x_1, x_2) dx_1 dx_2 &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x_1| f_X(x_1, x_2) dx_1 dx_2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x_2| f_X(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} |x_1| f_{X_1}(x_1) dx_1 + \int_{-\infty}^{\infty} |x_2| f_{X_2}(x_2) dx_2 \end{aligned}$$

$$\begin{aligned} \mathbb{E}(X_1 + X_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) f_X(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 + \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2 \\ &= \mathbb{E}X_1 + \mathbb{E}X_2 \end{aligned}$$

zu c) ähnlich wie b), dabei Unabhängigkeit verwenden

Folgerung 4.2.1 Wenn X_1, X_2 Zufallsgrößen sind deren Erwartungswerte existieren und $X_1 \leq X_2$, dann ist $\mathbb{E}X_1 \leq \mathbb{E}X_2$.

Satz 4.2.3 () Es sei X eine nicht negative Zufallsgröße (d.h. $P(X \geq 0) = 1$) mit Verteilungsfunktion F_X . Dann gilt:

$$\mathbb{E}X = \int_0^{\infty} 1 - F_X(x) dx$$

, wobei hier der Wert ∞ möglich ist.

Beweis (nur für stetigen Fall):

Es sei X eine nicht negative, stetige Zufallsgröße mit Verteilungsdichte $f_X \cdot 1_{[0, \infty)}$. Dann gilt:

$$\begin{aligned} \int_0^{\infty} 1 - F_X(x) dx &= \int_0^{\infty} \int_x^{\infty} f_X(t) dt dx \\ &= \int_0^{\infty} \int_0^{\infty} f_X(t) \cdot 1_{(x, \infty)}(t) dt dx \\ &= \int_0^{\infty} f_X(t) \int_0^{\infty} 1_{(0, t)}(x) dx dt \\ &= \int_0^{\infty} t \cdot f_X(t) dt = \mathbb{E}X \end{aligned}$$

Folgerung 4.2.2

$$\mathbb{E}X^2 = 0 \Rightarrow P(X = 0) = 1$$

Beweis:

$$\begin{aligned} 0 &= \mathbb{E}X^2 = \int_0^{\infty} 1 - F_X(x) dx \\ &\rightarrow F_{X^2} = 1_{[0, \infty)} \rightarrow P(X^2) = 1 \rightarrow P(X = 0) = 1 \end{aligned}$$

Beispiel

$X \sim \mathcal{E}_{\lambda}, \lambda > 0$.

$$\mathbb{E}X = \int_0^{\infty} 1 - (1 - e^{-\lambda x}) dx = \frac{1}{\lambda}$$

Definition 4.2.2 (Erwartungswert eines zufälligen Vektors) Es sei $\underline{X} = (X_1, \dots, X_n)$ ein zufälliger Vektor. Unter dem Erwartungswertvektor von \underline{X} versteht man den Vektor:

$$\mathbb{E}\underline{X} = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)$$

falls alle Erwartungswerte $\mathbb{E}X_i, i = 1, \dots, n$ existieren.

- falls $\underline{X} \sim \mathcal{N}_{\mu, \Sigma}$, dann $\mathbb{E}\underline{X} = \underline{\mu}$
(z.B. $\mathbb{E}X_1 = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1 f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_n$)

Definition 4.2.3 () Es sei $Z = X + iY$ eine komplexwertige zufällige Variable, wobei X, Y reelle Zufallsgrößen sind, deren Erwartungswert existiert. Dann wird der Erwartungswert von Z definiert als:

$$\mathbb{E}Z = \mathbb{E}X + i\mathbb{E}Y$$

Wichtige Transformationen von Verteilungsgesetzen

Definition 4.2.4 () Es sei X eine reelle Zufallsgröße.

- a) Es gilt $P(X \in \mathbb{N}_0) = 1$, d.h. X sei eine nicht negative ganze zufällige Zufallsgröße. Dann wird die (Wahrscheinlichkeit-)erzeugende Funktion von X (bzw von P_X) definiert durch

$$F_X(t) = \mathbb{E}t^X = \sum_{k=0}^{\infty} t^k \cdot P(X = k) \text{ für } |t| \leq 1$$

- b) Die charakteristische Funktion von X (bzw P_X) wird definiert als

$$\varphi_X(t) = \mathbb{E}e^{itX} \text{ für } t \in \mathbb{R}$$

Falls X stetig mit Verteilungsdichte f_X dann gilt:

$$f_X(t) = \int_{-\infty}^{\infty} f_X(x) \cdot e^{itx} dx$$

Wichtige Aussagen für diese Transformationen

- Eindeutigkeitssatz, Konvergenzsätze/Stetigkeitsaussagen
- Formel zur Berechnung von Momenten (falls diese existieren)

$$\frac{d}{dt} \varphi_X^{(t)}|_{t=0} = \mathbb{E} \cdot iX \cdot e^{itX}|_{t=0} = i\mathbb{E}X$$

- Behandlung von Summen unabhängiger Zufallsgrößen,

$$\begin{aligned}
 \varphi_{X_1+X_2}(t) &= \mathbb{E}e^{it(X_1+X_2)} \\
 &= \mathbb{E}(e^{itX_1} \cdot e^{itX_2}) \\
 &= (\mathbb{E}e^{itX_1})(\mathbb{E}e^{itX_2}) \\
 &= \varphi_{X_1}(t) \cdot \varphi_{X_2}(t)
 \end{aligned}$$

4.3 Varianz einer Zufallsgröße

Abweichung: $X - \mathbb{E}X$ ist Zufallsgröße, $\mathbb{E}(X - \mathbb{E}X) = 0$

Definition 4.3.1 (Varianz) Es sei X eine Zufallsgröße. Falls $\mathbb{E}X^2 < \infty$, dann heißt

$$\text{var}X = \mathbb{E}(X - \mathbb{E}X)^2$$

die Varianz von X . Falls $\mathbb{E}X^2 < \infty$, dann $\mathbb{E}|X| < \infty$, d.h. $\mathbb{E}X$ existiert und $\mathbb{E}X < \infty$. Es gilt:

$$\text{var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

Beweis:

$$\begin{aligned}
 \text{var}X &= \mathbb{E}(X - \mathbb{E}X)^2 \\
 &= \mathbb{E}(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2) \\
 &= \mathbb{E}X^2 - (\mathbb{E}X)^2
 \end{aligned}$$

Schreibweise: $\text{var}X = \sigma^2(X) = D^2(X) = \langle (X - \langle X \rangle)^2 \rangle$

Sprechweise:

$\text{var}X$... Varianz, Streuung, Dispersion

$\sqrt{\text{var}X}$... Standardabweichung

$\frac{\sqrt{\text{var}X}}{\mathbb{E}X}$... Varianzkoeffizient, falls $\mathbb{E}X \neq 0$

Beispiele:

- a) Zufallsgröße X mit $P(X = 1) = 1 - P(X = 0) = p$, $p \in [0, 1]$

Hier: $P(X^2 = 1) = 1 - P(X^2 = 0) = p$

d.h. $P_X = P_{X^2}$

$\mathbb{E}X^2 = \mathbb{E}X = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = p$

$\text{var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = p - p^2 = p(1 - p)$

b) $X \sim \mathcal{N}_{\mu, \sigma^2}$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ $\mathbb{E}X = \mu$

$$\begin{aligned} \text{var}X &= \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X - \mu)^2 \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ \text{Substitution: } t &= \frac{x - \mu}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 t^2 \cdot e^{-\frac{t^2}{2}} dt \quad \text{partielle Integration: } \underbrace{t}_u \cdot \underbrace{te^{-\frac{t^2}{2}}}_{v'} \Rightarrow u' = 1 \quad v = -e^{-\frac{t^2}{2}} \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left(-t \cdot e^{-\frac{t^2}{2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-\frac{t^2}{2}} dt \right) \\ &= \sigma^2 \end{aligned}$$

c) $X \sim \Pi_{\lambda}$, $\lambda > 0$

$$\begin{aligned} \mathbb{E}X^2 &= \sum_{k=0}^{\infty} k^2 \cdot P(X = k) \\ &= \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \\ &= e^{-\lambda} \left(\sum_{k=1}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty} 1 \frac{\lambda^k}{(k-1)!} \right) \\ &= e^{-\lambda} \left(\sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \right) \\ &= e^{-\lambda} (\lambda^2 e^{\lambda} + \lambda e^{\lambda}) = \lambda^2 + \lambda \end{aligned}$$

$$\begin{aligned} \text{var}X &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

Satz 4.3.1 ()

a) Es sei X eine Zufallsgröße mit $\mathbb{E}X^2 < \infty$, Dann gilt für alle $a, b \in \mathbb{R}$, dass $\mathbb{E}(aX + b)^2 < \infty$ und

$$\text{var}(aX + b) = a^2 \text{var}X$$

b) Es seien X_1, X_2 unabhängige Zufallsgrößen mit $\mathbb{E}X_i^2 < \infty, i = 1, 2$. Dann gilt für all $\mathbb{E}(X_1 + X_2)^2 < \infty$ und $\text{var}(X_1 + X_2) = \text{var}X_1 + \text{var}X_2$

$$\text{var}(X_1 - X_2) = \text{var}X_1 + \text{var}X_2$$

Beweis:zu a) $\mathbb{E}(aX + b)^2 \mathbb{E}(a^2 X^2 + 2abX + b^2) = a^2 \mathbb{E}X^2 + 2ab\mathbb{E}X + b^2 < \infty$, falls $\mathbb{E}X^2 < \infty$

$$\begin{aligned}
\text{var}(aX + b) &= \mathbb{E}(aX + b - \mathbb{E}(aX + b))^2 \\
&= \mathbb{E}(aX + b - a\mathbb{E}X - b)^2 \\
&= \mathbb{E}a^2(X - \mathbb{E}X)^2 \\
&= a^2 \text{var}X
\end{aligned}$$

zu b) $\mathbb{E}(X_1 + X_2)^2 = \mathbb{E}X_1^2 + \mathbb{E}X_2^2 + 2\mathbb{E}X_1X_2 < \infty$
Falls X_1, X_2 unabhängig, $\mathbb{E}(X_1X_2) = \mathbb{E}X_1 \cdot \mathbb{E}X_2 < \infty$

$$\begin{aligned}
\text{var}(X_1 + X_2) &= \mathbb{E}(X_1 + X_2 - \mathbb{E}(X_1 + X_2))^2 \\
&= \mathbb{E}(X_1 - \mathbb{E}X_1)^2 + \mathbb{E}(X_2 - \mathbb{E}X_2)^2 + 2\mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)) \\
&= \text{var}X_1 + \text{var}X_2 + 2\mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)) \\
&= \text{var}X_1 + \text{var}X_2 \text{ falls } X_1, X_2 \text{ unabhängig}
\end{aligned}$$

4.4 Die Kovarianz zweier Zufallsgrößen / die Kovarianz-Matrix eines zufälligen Vektors

Zunächst: 2-dimensionaler zufälliger Vektor $\underline{X} = (X_1, X_2)$ **Vorstellung:** Parameter, der den Grad der stochastischen Abhängigkeit von X_1 und X_2 ausdrückt.**Eine Möglichkeit:** $\mathbb{E}(X_1 \cdot X_2) - (\mathbb{E}X_1) \cdot (\mathbb{E}X_2)$ **Vorbemerkung:** Aus der Tatsache $(|X_1| - |X_2|)^2 \geq 0$, folgt $|X_1 \cdot X_2| \leq X_1^2 + X_2^2$ und damit $\mathbb{E}|X_1 \cdot X_2| \leq \mathbb{E}X_1^2 + \mathbb{E}X_2^2$ **Definition 4.4.1 (Kovarianz)** Es sei $\underline{X} = (X_1, X_2)$ ein zufälliger Vektor. Falls $\mathbb{E}X_i^2 < \infty, i = 1, 2$, dann heißt

$$\text{cov}(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2))$$

die Kovarianz von X_1 und X_2 .

Es gilt:

$$\text{cov}(X_1, X_2) = \mathbb{E}(X_1 \cdot X_2) - \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)$$

$$\text{cov}(X_1, X_2) = \text{cov}(X_2, X_1)$$

$$\text{var}(X_1 + X_2) = \text{var}X_1 + \text{var}X_2 + 2\text{cov}(X_1, X_2)$$

- Beachte: $\text{cov}(X_1, X_2)$ kann negativ sein
- falls X_1, X_2 stochastisch unabhängig, dann $\text{cov}(X_1, X_2) = 0$.
Umkehrung gilt im Allgemeinen nicht.
Wenn $\text{cov}(X_1, X_2) = 0$, dann heißen X_1, X_2 unkorreliert.

Beispiele:

a)

$$\begin{aligned}
\text{cov}(X, aX + b) &= \mathbb{E}((X - \mathbb{E}X)(aX + b - \mathbb{E}(aX + b))) \\
&= a \cdot \mathbb{E}(X - \mathbb{E}X)^2 = a \cdot \text{var}X
\end{aligned}$$

b) (X_1, X_2) besitze zweidimensionale Normalverteilung mit (μ_1, μ_2) , (σ_1^2, σ_2^2) , ρ

$$\begin{aligned}
 \text{cov}(X_1, X_2) &= \mathbb{E}((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) \cdot f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\
 \text{Substitution: } y_1 &= \frac{x_1 - \mu_1}{\sigma_1}, y_2 = \frac{x_2 - \mu_2}{\sigma_2} \\
 &= \frac{\sigma_1 \cdot \sigma_2}{2\pi \sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 \cdot \exp\left(-\frac{1}{2(1 - \rho^2)}(y_1^2 - 2\rho y_1 y_2 + y_2^2)\right) dy_1 dy_2 \\
 &= \sigma_1 \sigma_2 \int_{-\infty}^{\infty} y_2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1-\rho^2}{2(1-\rho^2)} y_2^2} \underbrace{\int_{-\infty}^{\infty} y_1 \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2}} e^{-\frac{1(y_1 - \rho y_2)^2}{2(1-\rho^2)}} dy_1}_{\rho y_2} dy_2 \\
 &= \rho \sigma_1 \sigma_2 \cdot \underbrace{\int_{-\infty}^{\infty} y_2^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y_2^2}{2}} dy_2}_{\text{Wert der Standardnormalverteilung}=1} \\
 &= \rho \sigma_1 \sigma_2
 \end{aligned}$$

→ für Zweidimensionale Normalverteilung:

$\rho = 0 \Leftrightarrow$ Unabhängigkeit von $X_1, X_2 \Leftrightarrow \text{cov}(X_1, X_2) = 0$ (d.h. X_1, X_2 unkorreliert)

Definition 4.4.2 (Korrelationskoeffizient) Es sei (X_1, X_2) ein zufälliger Vektor mit $\mathbb{E}X_i^2 < \infty$ und $\text{var}X_i > 0, i = 1, 2$

$$\rho_{X_1, X_2} = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}X_1 \cdot \text{var}X_2}}$$

heißt Korrelationskoeffizient von X_1 und X_2 .

- Aus der Cauchy-Schwartz'schen-Ungleichung folgt $(\text{cov}(X_1, X_2))^2 \leq \text{var}X_1 \cdot \text{var}X_2$ und damit $-1 \leq \rho_{X_1, X_2} \leq 1$
- Wenn X_1, X_2 unabhängig und $\text{var}X_i > 0, i = 1, 2$ dann: $\rho_{X_1, X_2} = 0$

Beispiel:

b) zweidimensionale Normalverteilung: $\rho_{X_1, X_2} = \rho$

a) $a \neq 0, \text{var}X > 0$:

$$\rho_{X, aX+b} = \frac{a \cdot \text{var}X}{\sqrt{\text{var}X \cdot a^2 \cdot \text{var}X}} = \frac{a}{|a|} = \text{sgn } a \begin{cases} 1, a > 0 \\ -1, a < 0 \end{cases}$$

- man kann sogar zeigen $|\rho_{X_1, X_2}| = 1 \Leftrightarrow \exists a, b \in \mathbb{R}, a \neq 0 : X_2 = aX_1 + b$ und $\text{sgn } a = \text{sgn } \rho_{X_1, X_2}$

Es sei $A = (Y_{ij})_{m \times n}$ eine zufällige Matrix, d.h. Y_{ij} sind reelle Zufallsgrößen. Man definiert

$$\mathbb{E}A = (\mathbb{E}Y_{ij})_{m \times n}$$

falls alle $\mathbb{E}Y_{ij}$ existieren.

Definition 4.4.3 (Kovarianzmatrix) Es sei $\underline{X} = (X_1, \dots, X_n)$ ein zufälliger Vektor mit $\mathbb{E}X^2 < \infty$, $i = 1, \dots, n$. Dann heißt die $(n \times n)$ -Matrix:

$$\Sigma_{\underline{X}} = \mathbb{E}((\underline{X} - \mathbb{E}\underline{X})^T(\underline{X} - \mathbb{E}\underline{X})) = (\text{cov}(X_i, X_j))_{n \times n}$$

die Kovarianzmatrix des Vektors \underline{X} .

Eigenschaften von $\Sigma_{\underline{X}}$:

- auf der Hauptdiagonalen: $\text{cov}(X_i, X_j) = \text{var}X_i$
- $\Sigma_{\underline{X}}$ ist symmetrisch, $\Sigma_{\underline{X}} = \Sigma_{\underline{X}}^T$
- $\Sigma_{\underline{X}}$ ist positiv semidefinit, d.h. $\underline{x}\Sigma_{\underline{X}}\underline{x}^T \geq 0$ für alle $\underline{x} \in \mathbb{R}^n$
Beweis: $\underline{x}\Sigma_{\underline{X}}\underline{x}^T = \underline{x}(\mathbb{E}((\underline{X} - \mathbb{E}\underline{X})^T(\underline{X} - \mathbb{E}\underline{X})))\underline{x}^T = \mathbb{E}(\underline{x}(\underline{X} - \mathbb{E}\underline{X})^T)^2 \geq 0$
- falls $\Sigma_{\underline{X}}$ regulär, dann ist $\Sigma_{\underline{X}}$ positiv definit, d.h. $\underline{x}\Sigma_{\underline{X}}\underline{x}^T > 0$ für alle $\underline{x} \in \mathbb{R}^n \setminus \{\underline{x} = 0\}$
- $\Sigma_{\underline{X}}$ ist Diagonalmatrix $\Leftrightarrow X_1, \dots, X_n$ paarweise unkorreliert
- für zufällige Vektoren: $\underline{X} \sim \mathcal{N}_{\mu, \Sigma} \Rightarrow \Sigma_{\underline{X}} = \Sigma$
 speziell für zweidimensionale Normalverteilung mit $(\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2), \rho$

$$\Sigma_{\underline{X}} = \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \sigma_2 \\ \rho \cdot \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

Kapitel 5

Ungleichungen und Grenzwertsätze

5.1 Einführung

Gesetze der Großen Zahlen:

relative Häufigkeit $\xrightarrow[\text{lineares Ereignis}]{\text{Anzahl der Versuche} \rightarrow \infty}$ Wahrscheinlichkeit des Ereignisses

Arithmetisches Mittel der beobachteten Werte von Realisierungen einer Zufallsgröße $\xrightarrow[\text{Erwartungswert der Zufallsgröße}]{\text{Anzahl der Realisierungen} \rightarrow \infty}$

Zentrale Grenzwertsätze:

Asymptotische Verteilung der Summe von i.i.d. Zufallsgrößen ist Normalverteilt.
solche Aussagen:

- illustrieren, dass Modelle und Begriffe der Wahrscheinlichkeitstheorie gelungen sind
- können interpretiert werden als „Gesetzmässigkeit des Zufalls“
- bilden wichtige Grundlage der mathematischen Statistik: „Aus Beobachtungen können Informationen über unbekannte Verteilung bzw. deren Parameter gewonnen werden“.

Hilfsmittel:

Ungleichungen zur Abschätzung von Wahrscheinlichkeiten für „Abweichungen“.

5.1.1 Vorbetrachtung

betrachten speziell unendliches Bernoulli-Schema:

X_1, \dots, X_n i.i.d. Zufallsgrößen mit $P(X_i = 1) = P(X_i = 0) = \frac{1}{2} = p$

für gerades $n = 2k, k \in \mathbb{N}$

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i = p\right) = P\left(\sum_{i=1}^{2k} X_i = k\right) = \binom{2k}{k} \frac{1}{2^{2k}} \xrightarrow{k \rightarrow \infty} 0$$

Nachweis mit Hilfe der STIRLINGschen Formel:

$$\lim_{n \rightarrow \infty} \frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} = 1$$

5.2 Markovsche und Tschebyschersche Ungleichung

Satz 5.2.1 (Markovsche Ungleichung)

Es seien X eine Zufallsgröße und $g : [0, \infty) \rightarrow [0, \infty)$ monoton nicht fallend und $g(x) > 0$ für $x > 0$. Dann gilt für alle $c > 0$:

$$P(|X| \geq c) \leq \frac{\mathbb{E}g(|X|)}{g(c)}$$

Beweis:

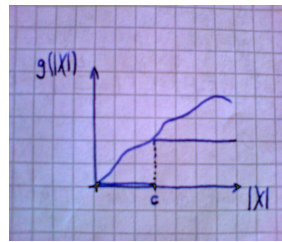


Abbildung 5.1: $g(|X|)$ über $|X|$

$$\begin{aligned} \text{Für } c > 0 \text{ gilt } g(c) \cdot 1_{[c, \infty)}(|X|) &\leq g(|X|) \\ \rightarrow g(c) \cdot P(|X| \geq c) &= g(c) \mathbb{E} \cdot 1_{[c, \infty)}(|X|) \\ &= \mathbb{E}g(c) \cdot 1_{[c, \infty)}(|X|) \leq \mathbb{E}g(|X|) \quad \square \end{aligned}$$

Satz 5.2.2 (Tschebyschevsche Ungleichung) Es sei X eine Zufallsgröße mit $\mathbb{E}X^2 < \infty$. Für alle $c > 0$ gilt:

$$P(|X - \mathbb{E}X| \geq c) \leq \frac{\text{var}X}{c^2}$$

Beweis:

Ersetzen in Satz 5.2.1 X durch $X - \mathbb{E}X$ und setzen $g(x) = x^2$ für $x \geq 0$ \square

Beispiel:

Sei $X \sim \mathcal{N}_{\mu, \sigma^2}$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$

Setzen $c = k\sigma$, $k = 1, 2, 3$

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} = \begin{cases} 1 & \text{für } k = 1 \\ \frac{1}{4} & \text{für } k = 2 \\ \frac{1}{9} & \text{für } k = 3 \end{cases}$$

5.3 Gesetze der Großen Zahlen

Satz 5.3.1 (schwaches Gesetz der Großen Zahlen) Es sei X_1, X_2, \dots eine Folge von i.i.d. Zufallsgrößen mit $\mathbb{E}X_i^2 < \infty$. Dann gilt für alle $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}X_1 \right| > \epsilon \right) = 0$$

Beweis:

Verwenden Tschebyschevsche Ungleichung und ersetzen dort X durch $\frac{1}{n} \sum_{i=1}^n X_i$. Für alle $\epsilon > 0$

$$0 \leq P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}X_1 \right| > \epsilon \right) \leq \frac{\text{var} X_1}{n \cdot \epsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

Folgerung 5.3.1 Es sei X_1, X_2, \dots eine Folge von i.i.d. Zufallsgrößen mit $P(X_i = 1) = 1 - P(X_i = 0) = p$; $p \in [0, 1]$. Dann gilt für alle $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{\sum_{i=1}^n X_i}{n} - p \right| > \epsilon \right) = 0$$

(Bereits 1713 von Jakob Bernoulli gezeigt)

Bemerkung

Für Gültigkeit der Aussage zum Satz des schwachen Gesetzes der Großen Zahlen genügt schon die paarweise Unkorreliertheit und die Existenz einer oberen Schranke von $\text{var} X_i$, $i = 1, 2, \dots$ allerdings $\mathbb{E}X_1 = \mathbb{E}X_2 = \dots$

Beispiel Würfeln:

$$X_i = \begin{cases} 0, & \text{falls im } i\text{-ten Wurf keine „6“} \\ 1, & \text{falls im } i\text{-ten Wurf eine „6“} \end{cases} \quad p = \frac{1}{6}$$

$$\text{var} X_i = p(1-p) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36} \quad \text{zum Beispiel } \epsilon = 0,01$$

$$P \left(\left| \frac{\sum_{i=1}^n X_i}{n} - \frac{1}{6} \right| > 0,01 \right) \leq \frac{50000}{36n} = \frac{1383}{n}$$

Definition 5.3.1 (Konvergenz) Es seien X, X_1, X_2, \dots Zufallsgrößen (auf einem Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P]$). Die Folge (X_n) konvergiert im Wahrscheinlichkeitsraum gegen X , wenn

$$\forall \epsilon > 0 : \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

Schreibweise: $X_n \xrightarrow{P} X$

Damit kann der Satz der schwachen Gesetze großer Zahlen auch geschrieben werden:

$$X_1, X_2, \dots \text{ i.i.d. , } \mathbb{E}X_1^2 < \infty \Rightarrow \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mathbb{E}X_1$$

ausführliche Schreibweise

$$\forall \epsilon > 0 : \lim_{n \rightarrow \infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0$$

Definition 5.3.2 (Konvergenz (fast sicher)) Es seien X, X_1, X_2, \dots Zufallsgrößen (auf einem Wahrscheinlichkeitsraum $[\Omega, \mathcal{U}, P]$) Die Folge (X_n) konvergiert fast sicher (f.s.) gegen X , wenn

$$P\left(\lim_{n \rightarrow \infty} |X_n - X| = 0\right) = 1$$

Schreibweise: $X_n \xrightarrow{f.s.} X$

ausführlicher Schreibweise:

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| = 0\right\}\right) = 1$$

Es gilt: Wenn $X_n \xrightarrow{f.s.} X$, dann $X_n \xrightarrow{P} X$

Verschärfung von Satz des schwachen Gesetzes großer Zahlen.

Satz 5.3.2 (starkes Gesetz der großen Zahlen) Es seien X_1, X_2, \dots eine Folge von i.i.d. Zufallsgrößen, für die $\mathbb{E}X_1$ existiert, dann gilt:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} \mathbb{E}X_1, \text{ d.h. } P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}X_1\right) = 1$$

Folgerung 5.3.2 Es sei X_1, X_2, \dots eine Folge von i.i.d. Zufallsgrößen mit $P(X_i = 1) = 1 - P(X_i = 0) = p$, $p \in [0, 1]$. Dann gilt:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} p$$

5.4 Der zentrale Grenzwertsatz

Satz 5.4.1 (Zentraler Grenzwertsatz (ZGWS)) Es sei X_1, X_2, \dots eine Folge von i.i.d. Zufallsgrößen mit $\mathbb{E}X_1^2 < \infty$ und $\text{var}X_1 = \sigma^2 > 0$, $\mathbb{E}X_1 = \mu$. Dann gilt für alle $x \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} P \left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x \right) = \Phi(x)$$

- Summe von i.i.d. Zufallsgrößen mit positiv endlicher Varianz sind asymptotisch normalverteilt.
 - Modellbildung zum Beispiel für Meßwerte
 - mathematische Statistik
- Konvergenzart: Punktweise Konvergenz von Verteilungsfunktionen (man kann sogar zeigen: gleichmäßige Konvergenz)

$\hat{=}$ schwache Konvergenz der Verteilung

$\hat{=}$ Konvergenz in Verteilung der Folge der Zufallsgrößen ($X_n \xrightarrow{d} X$)

Es gilt: $X_n \xrightarrow{\text{f.s.}} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$
nach zentralem Grenzwertsatz gilt:

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \xrightarrow{d} Z, Z \sim \mathcal{N}_{0,1}$$

Aus dem Gesetz der Großen Zahlen folgt:

$$\frac{1}{n} \sum_{i=1}^n X_i - \mu \xrightarrow{d} 0$$

- Eine Aussage zur Konvergenzgeschwindigkeit:
Satz von Berry-Esseen (siehe Georgii S.135). Falls die Voraussetzung für den zentralen Grenzwertsatz erfüllt und ausserdem $\mathbb{E}|X_1|^3 < \infty$, dann:

$$\sup_{x \in \mathbb{R}} \left| P \left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x \right) - \Phi(x) \right| \leq 0,8 \frac{\mathbb{E}(|X_1 - \mathbb{E}X_1|)^3}{\sigma^3} \cdot \frac{1}{\sqrt{n}}$$

Anmerkung

Möglichkeit zur Simulation von $\mathcal{N}_{0,1}$ -verteilten Zufallsgrößen.

$n = 12$ $X_1, X_2, \dots, X_{12} \sim U(0, 1)$, i.i.d

$$\sum_{i=1}^n X_i - 6 \sim \mathcal{N}_{0,1} \text{ approximiert}$$

Achtung: Es kann fraglich sein, dass ein Zufallsgenerator unabhängige Zufallsgrößen gut simuliert!