

1 Mathematische Modelle von Zufallsexperimenten

Grundbegriffe

Grundraum:	nicht leere Menge $\Omega \neq \emptyset$ Enthält alle möglichen Ergebnisse eines Zufallsexperimentes.
Ereignis:	Teilmenge $A \subset \Omega$ Nicht immer wird jede Teilmenge als Ereignis bezeichnet, sondern nur Mengen aus einem Mengensystem $\mathcal{A} \subset 2^\Omega$ Sprechweise: Ereignis A tritt ein \Leftrightarrow Ergebnis ω liegt in A
Elementarereignis:	$\omega \in \Omega$ Vorsicht: ω ist kein Ereignis! Aber $\{\omega\}$ ist ein Ereignis, falls \mathcal{A} geeignet gewählt ist.

Beispiel 1.1.

(i) Werfen eines Würfels

$$\Omega = \{1, \dots, 6\}$$

$$\text{Ereignis "Augenzahl ist gerade": } A = \{2, 4, 6\}$$

(ii) In einem Netzwerk werden Längen (in Byte) der ersten $n = 10^5$ Datenpakete beobachtet

$$\Omega = \mathbb{N}^n = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \mathbb{N} \forall 1 \leq i \leq n\}$$

Interpretation: $\omega_i =$ Länge des i -ten Paketes

Ereignis "Das größte Paket umfasst maximal 10^7 Byte"

$$A := \{(\omega_1, \dots, \omega_n) \mid \omega_i \leq 10^7 \forall 1 \leq i \leq n\}$$

"Rechnen mit Ereignissen"

Seien $A, B \subset \Omega$ Ereignisse. Dann ist

- $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ oder } \omega \in B\}$
Ereignis, dass A eintritt oder B eintritt
- $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ und } \omega \in B\}$
Ereignis, dass A und B eintritt
- $A \setminus B = \{\omega \in \Omega \mid \omega \in A, \omega \notin B\}$ Ereignis, dass A eintritt, aber nicht B eintritt

- $A \subset B$

Wenn A eintritt, dann tritt auch B ein.

Im Wahrscheinlichkeitsmodell soll jedem Ereignis eine Wahrscheinlichkeit (W.) zugeordnet werden.

Diese müssen "zueinander passen".

Z.B. muss bei $A \subset B$ die Wahrscheinlichkeit von $A \leq$ Wahrscheinlichkeit von B sein (Monotonie)

Mathematisch:	Abbildung	$P : \mathcal{A} \rightarrow \mathbb{R}$	
	mit Normierung	$P(\emptyset) = 0$	Ereignis, das <u>nie</u> eintritt, hat W. 0
		$P(\Omega) = 1$	Ereignis, das <u>immer</u> eintritt, hat W. 1
		$\rightsquigarrow P : \mathcal{A} \rightarrow [0, 1]$	

Welche Abb. P sind sinnvoll?

→ Folien "1. Ansatz: Inhaltlicher Zusammenhang"
bis "Subjektivistische" ...

Axiomatischer Zugang - Kolmogorov (1933)

Fordere Eigenschaften von P

Zur Motivation: Für die relativen Häufigkeiten ("empirischen Wahrscheinlichkeiten") P_n gilt

$$P_n(\Omega) = 1, \quad P_n(\emptyset) = 0$$

$$A, B \text{ disjunkt (d.h. } A \cap B = \emptyset) \Rightarrow P_n(A \cup B) = P_n(A) + P_n(B)$$

$$\text{Induktiv folgt: Ereignisse } A_1, \dots, A_k \text{ disjunkt} \Rightarrow P_n\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k P_n(A_i)$$

endl. Wahrscheinlichkeitsraum & Wahrscheinlichkeitsmaß

Definition 1.2 (endl. Wahrscheinlichkeitsraum & Wahrscheinlichkeitsmaß). *Ein endlicher Wahrscheinlichkeitsraum ist ein Paar (Ω, P) bestehend aus einem endlichen Grundraum $\Omega \neq \emptyset$ und einer Abbildung $P : 2^\Omega \rightarrow [0, 1]$ mit den Eigenschaften:*

- $P(\Omega) = 1$
- $A, B \subset \Omega$ disjunkt $\Rightarrow P(A \cup B) = P(A) + P(B)$ (Additivität)

P heißt dann Wahrscheinlichkeitsmaß.

Beispiel 1.3.

(i) vgl. Beispiel 1.1.(i): einmaliges Werfen eines fairen Würfels

$$\Omega = \{1, \dots, 6\}$$

$$A \subset \Omega : P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6}$$

$$\text{insbesondere } P(\{\omega\}) = \frac{1}{6}$$

(ii) Werfen zweier ununterscheidbarer fairer Würfel

$$\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, \dots, 6\}, \omega_1 \leq \omega_2\}$$

Hier ist es nicht sinnvoll, jeder Menge $\{(\omega_1, \omega_2)\}$ die gleiche Wahrscheinlichkeit zuzuweisen, denn z.B. gibt es 2 Möglichkeiten (1,2) zu erhalten, aber nur eine für (1,1).

$$\rightsquigarrow P(\{\omega_1, \omega_2\}) = \begin{cases} \frac{1}{36}, & \omega_1 = \omega_2 \\ \frac{2}{36}, & \omega_1 \neq \omega_2 \end{cases}$$

Für $A \subset \Omega$ folgt

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

Satz 1.4. Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum, dann gilt

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

(ii) $A_1, \dots, A_k \subset \Omega$ disjunkt $\Rightarrow P\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k P(A_i)$

(iii) $P(A^c) = 1 - P(A) \quad \forall A \subset \Omega$

(iv) $A \subset B \subset \Omega \Rightarrow P(B \setminus A) = P(B) - P(A)$

(v) $A, B \subset \Omega \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(vi) $\forall A_1, \dots, A_k \subset \Omega \Rightarrow P\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k P(A_i)$ (Subadditivität)

Beweis: (i) - (v) üben

(vi) Def. $B_1 := A_1, B_i := A_i \setminus \left(\bigcup_{j=1}^{i-1} A_j\right) \subset A_i \quad \forall 2 \leq i \leq k$

$\Rightarrow B_1, \dots, B_k$ disjunkt (denn für $i < l$ gilt $B_i \subset A_i, A_i \cap B_l = \emptyset$)

$$\bigcup_{i=1}^k B_i = \bigcup_{i=1}^k A_i$$

$$\Rightarrow P\left(\bigcup_{i=1}^k A_i\right) = P\left(\bigcup_{i=1}^k B_i\right) \stackrel{(ii)}{=} \sum_{i=1}^k P(B_i) \leq \sum_{i=1}^k P(A_i)$$

(Ist $\omega \in A_i$ für ein i und ist i_ω das kleinste solche i , dann ist $\omega \in B_{i_\omega}$)

Definition 1.5 (Gleichverteilung oder Laplace-Verteilung). Ist Ω endlich, so heißt das durch

$P(A) = \frac{|A|}{|\Omega|}$, $A \subset \Omega$,
definierte Wahrscheinlichkeitsmaß die Gleichverteilung oder Laplace-Verteilung auf Ω .

1.6 Urnenmodelle

Wenn Gleichverteilung betrachtet wird, sind nur Mächtigkeiten von Mengen zu bestimmen.

Oft hilfreich sind Urnenmodelle

Situation

Es sind die Kugeln $1, \dots, n$ in der Urne und es wird k -mal aus der Urne gezogen.

Es wird unterschieden, ob

- mit oder ohne Zurücklegen gezogen wird
- mit oder ohne Berücksichtigung der Reihenfolge gezogen wird

1.6 (a) Ziehen mit Zurücklegen und mit Berücksichtigung der Reihenfolge

Grundraum $\Omega_a = \{1, \dots, n\}^k = \{(\omega_1, \dots, \omega_k) \mid \omega_i \in \{1, \dots, n\}\}$

Interpretation: $\omega = (\omega_1, \dots, \omega_k) \in \Omega_a$
 $\hat{=}$ im i -ten Zug werde Kugel ω_i gezogen

$$|\Omega_a| = n^k$$

1.6 (b) Ziehen ohne Zurücklegen, aber mit Berücksichtigung der Reihenfolge

$\Omega_b = \{(\omega_1, \dots, \omega_k) \in \Omega_a \mid \omega_i \neq \omega_j \forall 1 \leq i < j \leq k\}$

Interpretation wie bei (a), $|\Omega_b| = n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}$

Begründung: n Möglichkeiten für ω_1

dann $n-1$ Möglichkeiten für ω_2 bei gegebenem ω_1

dann $n-2$ Möglichkeiten für ω_3 bei gegebenen ω_1, ω_2

... ..

dann $n-k+1$ Möglichkeiten für ω_k bei gegebenen $\omega_1, \dots, \omega_{k-1}$

-
- 1.6 (c) **Ziehen ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge**
Ordne gezogene Kugelnummern nach Größe

$$\Omega_c = \{(\omega_1, \dots, \omega_k) \in \Omega_a \mid \omega_1 < \omega_2 < \dots < \omega_k\}$$

Es gibt zu jedem $(\omega_1, \dots, \omega_k) \in \Omega$ genau $k!$ verschiedene Möglichkeiten die k verschiedenen Kugelnummern anzuordnen.

D.h. jedem $\omega \in \Omega_c$ entsprechen $k!$ Elemente aus Ω_b

$$\Rightarrow |\Omega_c| = \frac{|\Omega_b|}{k!} = \frac{n!}{k!(n-k)!} =: \binom{n}{k}$$

- 1.6 (d) **Ziehen mit Zurücklegen und ohne Berücksichtigung der Reihenfolge**

$$\Omega_d = \{(\omega_1, \dots, \omega_k) \in \Omega_a \mid \omega_1 \leq \omega_2 \leq \dots \leq \omega_k\}$$

Die Abbildung

$$S : \Omega_d \rightarrow \{(\tilde{\omega}_1, \dots, \tilde{\omega}_k) \in \{1, \dots, n+k-1\}^k \mid \tilde{\omega}_1 < \tilde{\omega}_2 < \dots < \tilde{\omega}_k\} =: \tilde{\Omega}_c$$

$$S(\omega_1, \dots, \omega_k) := (\omega_1, \omega_2 + 1, \omega_3 + 2, \dots, \omega_k + k + 1)$$

ist eine Bijektion.

$\tilde{\Omega}_c$ ist vom gleichen Typ wie Ω_c mit n ersetzt durch $n+k-1$.

$$\text{Also } |\Omega_d| = |\tilde{\Omega}_c| \stackrel{(c)}{=} \binom{n+k-1}{n}$$

Beispiel 1.7. (i) *Lotto 6 aus 49*

6 Kugeln werden ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge aus 49 Kugeln gezogen

$$\stackrel{1.6(c)}{\Rightarrow} \text{Es gibt } \binom{49}{6} = 13.983.816 \text{ Möglichkeiten}$$

Alle Möglichkeiten sind gleich wahrscheinlich

$$\Rightarrow \text{Gewinnwahrscheinlichkeit für ein Los } \frac{1}{\binom{49}{6}} \approx 7,15 \cdot 10^{-8}$$

- (ii) Sucht man bei Yahoo nach Schlagwörtern "Hausdepp Kartenspiel", so gibt es unter den ersten 10 ausgegebenen Links genau einen, unter dem die Regeln dieses Kartenspiels zu finden sind. Sei $r \in \{1, \dots, 10\}$ die Nummer dieses Links.

Sie probieren die 10 Links in zufälliger Reihenfolge durch, bis Sie auf den Link mit den Regeln stoßen.

Wie groß ist die Wahrscheinlichkeit, dass Sie genau k Versuche brauchen?

\rightsquigarrow Ziehen ohne Zurücklegen und mit Berücksichtigung der Reihenfolge

$$\Omega = \{(\omega_1, \dots, \omega_k) \in \{1, \dots, 10\}^k \mid \omega_i \neq \omega_j \forall i \neq j\}$$

Ereignis "Erfolg nach genau k Versuchen"

$A = \{(\omega_1, \dots, \omega_{k-1}, r) \mid \omega_i \in \{1, \dots, 10\} \setminus \{r\} \forall 1 \leq i \leq k-1\}$
entspricht 9-maligem Ziehen ohne Zurücklegen aus $\{1, \dots, 10\} \setminus \{r\}$

$$\Rightarrow |A| = \frac{9!}{(9-(k-1))!}$$

$$|\Omega| = \frac{10!}{(10-k)!}$$

Bei Gleichverteilung

$$P(A) = \frac{|A|}{|\Omega|} = \frac{9!}{10!} = \frac{1}{10}$$

Bei 2 richtigen Links erhält man auf ähnliche Weise

$$|A| = \frac{8! \cdot 2}{(8-(k-1))!} \quad (\text{da es nun 2 Möglichkeiten für den } k\text{-ten Versuch gibt})$$

$$|\Omega| = \frac{10!}{(10-k)!}$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{8! \cdot 2 \cdot (10-k)!}{10! (8-(k-1))!} = \frac{2(10-k)}{10 \cdot 9} = \frac{10-k}{45}$$

Liegt wie z.B. im Beispiel 1.1(ii) kein endlicher, sondern ein abzählbarer Grundraum vor, dann ist die Additivität nicht mehr ausreichend.

Definition 1.8 (diskretes Wahrscheinlichkeitsmaß & -raum). Sei $\Omega \neq \emptyset$ eine beliebige nicht-leere Menge.

Eine Abbildung $P : 2^\Omega \rightarrow [0, 1]$ heißt diskretes Wahrscheinlichkeitsmaß, falls

(i) $P(\Omega) = 1$

(ii) $\forall A_n \subset \Omega, n \in \mathbb{N}$ disjunkt: $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$ ("σ-Additivität")

(iii) Es existiert eine (höchstens) abzählbare Menge $\Omega_0 \subset \Omega$ mit $P(\Omega_0) = 1$.

Dann heißt (Ω, P) diskreter Wahrscheinlichkeitsraum.

Satz 1.4 gilt völlig analog auch für diskrete Wahrscheinlichkeitsräume.
Es gilt sogar in Verallgemeinerung von 1.4(vi)

Satz 1.9. Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum und $A_n \subset \Omega \forall n \in \mathbb{N}$
Dann gilt $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} P(A_n)$ ("σ-Subadditivität")

Satz & Definition 1.10.

(i) Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum.

Dann heißt die Funktion $f : \Omega \rightarrow [0, 1], f(\omega) = P(\{\omega\})$ Zähldichte oder Wahrscheinlichkeitsfunktion von P .

Sie besitzt folgende Eigenschaften:

(a) $\Omega_T := \{\omega \in \Omega \mid f(\omega) > 0\}$ ist abzählbar (und heißt Träger von P bzw. von f)

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

Für alle $A \subset \Omega$ gilt dann:

$$P(A) = \sum_{\omega \in A} f(\omega) = \sum_{\omega \in A \cap \Omega_T} f(\omega) \quad (1.1)$$

(ii) Ist umgekehrt $\Omega \neq \emptyset$ und $f : \Omega \rightarrow [0, 1]$ eine Funktion, die (a) und (b) erfüllt, so existiert genau ein diskretes Wahrscheinlichkeitsmaß auf Ω , das f zur Zähldichte hat.

Beweis:

(i) $\exists \Omega_0 \subset \Omega$ abzählbar, $P(\Omega_0) = 1$

$$\Rightarrow P(\Omega_0^c) = 0$$

$$\Rightarrow \forall \omega \in \Omega_0^c : f(\omega) = P(\{\omega\}) \leq P(\Omega_0^c) = 0$$

$$\Rightarrow \Omega_T \subset \Omega_0 \Rightarrow \Omega_T \text{ abzählbar, d.h. (a).}$$

(b) ist Spezialfall von (1.1) mit $A = \Omega$

(1.1) folgt aus der σ -Additivität von P , denn

$$\begin{aligned} P(A) &= P(A \cap \Omega_T) + P(A \cap \Omega_T^c) \\ &= P(A \cap \Omega_T) \\ &= P\left(\bigcup_{\omega \in A \cap \Omega_T} \{\omega\}\right) \\ &= \sum_{\omega \in A \cap \Omega_T} P(\{\omega\}) = \sum_{\omega \in A \cap \Omega_T} f(\omega) = \sum_{\omega \in A} f(\omega) \end{aligned}$$

da $f(\omega) = 0 \forall \omega \notin \Omega_T$

(ii) Def. $P(A)$ gemäß (1.1) für alle $A \subset \Omega$

Dann ist P ein diskretes Wahrscheinlichkeitsmaß, denn

- $P(\Omega) \stackrel{(1.1)}{=} \sum_{\omega \in \Omega} f(\omega) \stackrel{(b)}{=} 1$

- $A_n \subset \Omega, n \in \mathbb{N}$, disjunkt

$$\Rightarrow P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{\omega \in \bigcup_{n \in \mathbb{N}} A_n} f(\omega) = \sum_{n \in \mathbb{N}} \sum_{\omega \in A_n} f(\omega) \stackrel{(1.1)}{=} \sum_{n \in \mathbb{N}} P(A_n)$$

- $P(\Omega_T) \stackrel{(1.1)}{=} \sum_{\omega \in \Omega_T} f(\omega) = \sum_{\omega \in \Omega} f(\omega) \stackrel{(b)}{=} 1$
und Ω_T ist abzählbar nach (a)

Also ist P ein diskretes Wahrscheinlichkeitsmaß mit $P(\{\omega\}) \stackrel{(1.1)}{=} f(\omega)$ d.h. f ist Zähldichte von P . Dass P das einzige diskrete Wahrscheinlichkeitsmaß mit diesen Eigenschaften ist, folgt aus (1.1).

Beispiel 1.11.

(i) Sei $p \in [0, 1]$

$$1 = (p + (1 - p))^n = \sum_{k=0}^n \binom{n}{k} p^k \cdot (1 - p)^{n-k} =: f(k)$$

Dies definiert eine Funktion $f : \{0, \dots, n\} =: \Omega \rightarrow [0, 1]$, die 1.10(a) und 1.10(b) erfüllt, also die Zähldichte eines diskreten Wahrscheinlichkeitsmaßes auf Ω ist.

Dieses Wahrscheinlichkeitsmaß heißt Binomialverteilung $\mathcal{B}_{(n,p)}$

$$\text{d.h. } \mathcal{B}_{(n,p)}(\{k\}) = \binom{n}{p} p^k \cdot (1 - p)^{n-k} \quad \forall k \in \{0, \dots, n\}$$

Später: $\mathcal{B}_{(n,p)}$ beschreibt zufällige Anzahl der Erfolge bei n -maliger unabhängiger Durchführung eines Zufallsexperiments mit Erfolgswahrscheinlichkeit p .

(ii) Frage : Wie groß ist die Wahrscheinlichkeit, dass man k mal würfeln muss, bevor das erste Mal 6 gewürfelt wird?

Bei fairem Würfel:

$$\text{Wahrscheinlichkeit, } k\text{-mal keine 6 zu würfeln} = \left(\frac{5}{6}\right)^k$$

$$\text{Wahrscheinlichkeit, beim } (k+1)\text{-ten Mal 6 zu würfeln} = \left(\frac{1}{6}\right)$$

$$\Rightarrow \text{gesuchte Wahrscheinlichkeit} = \left(\frac{5}{6}\right)^k \cdot \frac{1}{6} \quad \forall k \in \mathbb{N}_0$$

Allgemein: Für $p \in (0, 1]$ definiert

$$f(k) = (1 - p)^k \cdot p, \quad k \in \mathbb{N}_0,$$

eine Zähldichte auf \mathbb{N}_0 , denn \mathbb{N}_0 ist abzählbar und

$$\sum_{k \in \mathbb{N}_0} f(k) = p \cdot \sum_{n \in \mathbb{N}_0} (1 - p)^n \stackrel{\text{geometrische Reihe}}{=} p \cdot \frac{1}{1 - (1 - p)} = 1$$

Das zugehörige Wahrscheinlichkeitsmaß heißt geometrische Verteilung \mathcal{G}_p und beschreibt die Anzahl der Misserfolge bis zum ersten Erfolg bei unabhängiger Durchführung eines Zufallsexperiments mit Erfolgswahrscheinlichkeit p .

2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Gesucht: Wahrscheinlichkeit, dass Ereignis A eintritt, wenn bekannt ist, dass Ereignis B eintritt/eintreten wird.

Ein Experiment wird n mal durchgeführt
 n_A = Anzahl der Experimente, bei denen A eintritt
 n_B = Anzahl der Experimente, bei denen B eintritt
 $n_{A \cap B}$ = Anzahl der Experimente, bei denen A und B eintreten

empirisches Gesetz der großen Zahlen:

Für "große" n gilt $\frac{n_A}{n} \approx P(A)$, $\frac{n_B}{n} \approx P(B)$, $\frac{n_{A \cap B}}{n} \approx P(A \cap B)$

Zur Bestimmung obiger Wahrscheinlichkeit betrachtet man nur Experimente, bei denen B eingetreten ist.

$\rightsquigarrow \frac{n_{A \cap B}}{n_B} = \frac{n_{A \cap B}/n}{n_B/n} \approx \frac{P(A \cap B)}{P(B)}$, falls der Nenner > 0 .

Definition 2.1 (bedingte Wahrscheinlichkeit von A gegeben B). Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum, $B \subset \Omega$ mit $P(B) > 0$, $A \subset \Omega$. Dann heißt

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

die bedingte Wahrscheinlichkeit von A gegeben B

Beispiel 2.2. (vergleiche Präsenzübungsblatt 2, A3)

Beim Skatenspiel hat Spieler 1 keinen Buben erhalten.

Wie groß ist dann die Wahrscheinlichkeit, dass Spieler 2 oder Spieler 3 alle 4 Buben erhalten hat?

Modell: $\Omega = \{(\underbrace{\omega_1, \dots, \omega_{10}}_{\text{Karten Sp.1}}, \underbrace{\omega_{11}, \dots, \omega_{20}}_{\text{Karten Sp.2}}, \underbrace{\omega_{21}, \dots, \omega_{30}}_{\text{Karten Sp.3}}, \underbrace{\omega_{31}, \omega_{32}}_{\text{Stock}}) \in K^{32} \mid \omega_i \neq \omega_j \forall i \neq j\}$

mit $K = \{\text{KreuzAs}, \text{PikAs}, \dots, \text{Karo7}\}$ Menge aller Karten

P Laplace-Verteilung

$$C = \{\omega \in \Omega \mid \{\omega_1, \dots, \omega_{10}\} \cap \{\text{KreuzBube}, \text{PikBube}, \text{HerzBube}, \text{KaroBube}\} = \emptyset\}$$

Ergebnis, dass Spieler 1 keinen Buben erhält.

$$B_2 = \{\omega \in \Omega \mid \{\text{KreuzBube}, \text{PikBube}, \text{HerzBube}, \text{KaroBube}\} \subset \{\omega_{11}, \dots, \omega_{20}\}\}$$

Ereignis, dass Spieler 2 alle Buben erhält.

analog: B_3 Ereignis, dass Spieler 3 alle Buben erhält.

$$\Rightarrow B_2 \cap B_3 = \emptyset, \quad B_2 \cup B_3 \subset C$$

$$|C| = \underbrace{\frac{28!}{18!}}_1 \cdot \underbrace{22!}_2$$

¹: Anzahl der Möglichkeiten, 10 Karten für Spieler 1 aus 28 "Nicht-Buben" zu ziehen (siehe 1.6(b))

²: Anzahl der Möglichkeiten, die restlichen 22 Karten zu verteilen.

$|B_2| \notin |B_3|$ in Präsenzaufgabe.

Gesuchte Wahrscheinlichkeit:

$$\begin{aligned} P(B_2 \cup B_3 | C) &\stackrel{2.1}{=} \frac{P((B_2 \cup B_3) \cap C)}{P(C)} \\ &= \frac{P(B_2) + P(B_3)}{P(C)} \\ &= \frac{\frac{|B_2|}{|\Omega|} + \frac{|B_3|}{|\Omega|}}{\frac{|C|}{|\Omega|}} = \frac{|B_2| + |B_3|}{|C|} = \frac{2 \cdot |B_2|}{|C|} = \overset{\text{Präsenzaufg.}}{\dots} = \frac{12}{209} \approx 0,057 \end{aligned}$$

Satz 2.3 (totale Wahrscheinlichkeit & Bayes). Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum, $B_i \subset \Omega (i \in I)$ disjunkt, $\bigcup_{i \in I} B_i = \Omega$, I (höchstens) abzählbar und $P(B_i) > 0, A \subset \Omega$

(i) Satz von der totalen Wahrscheinlichkeit

$$P(A) = \sum_{i \in I} P(A | B_i) \cdot P(B_i)$$

(ii) Satz von Bayes

Falls $P(A) > 0, k \in I$, dann

$$P(B_k | A) = \frac{P(A | B_k) \cdot P(B_k)}{\sum_{i \in I} P(A | B_i) \cdot P(B_i)}$$

Beweis:

$$(i) \sum_{i \in I} P(A | B_i) \cdot P(B_i) = \sum_{i \in I} \frac{P(A \cap B_i)}{P(B_i)} \cdot P(B_i) = P\left(\bigcup_{i \in I} (A \cap B_i)\right) = P(A),$$

da $A \cap B_i$ disjunkt.

$$(ii) P(B_k | A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A | B_k) \cdot P(B_k)}{P(A)} \Rightarrow \text{Beh. mit (i)}$$

Beispiel 2.4. *Trisomie 21: genetischer Defekt, verursacht Down-Syndrom*

Der Defekt kann schon bei Föten durch eine Fruchtwasseruntersuchung erkannt werden.

In 99% der Fälle, in denen Trisomie 21 vorliegt, ist der Test positiv;

in 98% der Fälle, in denen Trisomie 21 nicht vorliegt, fällt der Test negativ aus.

Betrachte die Ereignisse:

D: Trisomie 21 liegt vor

T: Test positiv

Dann gilt

$$P(T | D) = 0,99 \quad \text{Sensitivität des Tests}$$

$$P(T^c | D^c) = 0,98 \quad \text{Spezifität des Tests}$$

$$\Rightarrow P(T | D^c) = 0,02$$

Was bedeutet es, wenn der Test positiv ist?

$$P(D | T) \stackrel{\text{Bayes}}{=} \frac{P(T | D) \cdot P(D)}{P(T | D) \cdot P(D) + P(T | D^c) \cdot P(D^c)} = \frac{1}{1 + \frac{P(T|D^c) \cdot P(D^c)}{P(T|D) \cdot P(D)}}$$

P(D) entspricht der relativen Häufigkeit, mit der Trisomie 21 in der betrachteten Bevölkerungsgruppe auftritt.

$$\text{25 jährige Schwangere: } P(D) \approx \frac{1}{1250} \Rightarrow P(D | T) \approx 0,035$$

$$\text{43 jährige Schwangere: } P(D) \approx \frac{1}{50} \Rightarrow P(D | T) \approx 0,503$$

Während das Baby einer 45jährigen Schwangeren also etwa mit Wahrscheinlichkeit 1/2 tatsächlich Trisomie 21 aufweist, wenn der Test positiv ausfällt, ist dies bei 25jährigen Schwangeren dennoch nur mit einer Wahrscheinlichkeit von unter 4% der Fall.

Generell ist die Aussagekraft von derartigen medizinischen Tests selbst bei hoher Sensitivität und Spezifität sehr beschränkt, wenn die Erkrankungswahrscheinlichkeit sehr gering ist. Daher ist in einem solchen Fall ein allgemeines Screening sinnlos.

Wird A nicht von B beeinflusst, so sollte $P(A | B) = P(A)$ gelten.

Definition 2.5 ((P-)stochastisch unabhängig). *Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum. Ereignisse $A, B \subset \Omega$ heißen (P-)stochastisch unabhängig, falls*

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

Definition 2.6 ((P-)stochastisch unabhängig allgemein). *Ereignisse $A_1, \dots, A_n \subset \Omega$ in einem diskreten Wahrscheinlichkeitsraum (Ω, P) heißen (P-)stochastisch unabhängig, wenn für jede Indexmenge $I \subset \{1, \dots, n\}, I \neq \emptyset$, gilt*

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

Bemerkung 2.7.

(i) Definition 2.6 stellt sicher, dass jede beliebige Auswahl $A_i, i \in I \subset \{1, \dots, n\}$ aus unabhängigen Ereignissen A_1, \dots, A_n auch wieder unabhängig sind.

(ii) Mehr als 2 Ereignisse A_1, \dots, A_n sind im allgemeinen nicht stochastisch unabhängig, wenn nur $P(\bigcup_{i=1}^n A_i) = \prod_{i=1}^n P(A_i)$

Ist z.B. $A_1 = \emptyset$, so gilt $P(\bigcap_{i=1}^n A_i) = P(\emptyset) = 0 = \prod_{i=1}^n P(A_i)$

Ist $A_2 = A_3 = A$ mit $P(A) \in (0, 1)$, so gilt

$$P(A_2 \cap A_3) = P(A) \neq P(A_2) \cdot P(A_3) = (P(A))^2$$

$\Rightarrow A, A_2, A_3$ sind nicht stochastisch unabhängig

(iii) Ebenso sind mehr als 2 Ereignisse i.d.R. nicht stochastisch unabhängig, wenn jeweils 2 der Ereignisse stochastisch unabhängig sind.

Bsp.: 2 maliges Werfen eines fairen Würfels

$\Omega = \{1, \dots, 6\}$, P Laplace-Verteilung

$A_1 = \{1, 3, 5\} \times \{1, \dots, 6\}$ (Erste Augenzahl ist ungerade)

$A_2 = \{1, \dots, 6\} \times \{1, 3, 5\}$ (Zweite Augenzahl ist ungerade)

$A_3 = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 \text{ ungerade}\}$ (Summe ist ungerade)
 $= (\{1, 3, 5\} \times \{2, 4, 6\}) \cup (\{2, 4, 6\} \times \{1, 3, 5\})$

A_1, A_2 sind stochastisch unabhängig

A_1, A_3 sind stochastisch unabhängig

A_2, A_3 sind stochastisch unabhängig

$$\begin{aligned} \text{z.B. } P(A_2 \cap A_3) &= P(\{2, 4, 6\} \times \{1, 3, 5\}) = \frac{|\{2, 4, 6\} \times \{1, 3, 5\}|}{|\Omega|} = \frac{9}{36} = \frac{1}{4} \\ &= \frac{1}{2} \cdot \frac{1}{2} = P(A_2) \cdot P(A_3), \end{aligned}$$

d.h. A_2, A_3 sind stochastisch unabhängig.

Aber A_1, A_2, A_3 sind nicht stochastisch unabhängig, denn

$$A_1 \cap A_2 \cap A_3 = \emptyset \Rightarrow P(A_1 \cap A_2 \cap A_3) = 0 \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{8}$$

3 Zufallsvariablen und ihre Verteilungen

Oft ist nicht das genaue Ergebnis eines Zufallsexperiments von Interesse, sondern nur eine summarische Größe, die einen gewissen Aspekt des Ergebnisses widerspiegelt

↔ Abbildung $X : \Omega \rightarrow S, S \neq \emptyset$ beliebige Menge.

———— Zeichnung —————

Urbild von B unter $X : X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} =: \{X \in B\}$

Erinnerung:

$$X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$$

$$X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$$

$$X^{-1}(A \setminus B) = X^{-1}(A) \setminus X^{-1}(B)$$

Satz & Definition 3.1 (Zufallsvariable). *Ist (Ω, P) ein diskreter Wahrscheinlichkeitsraum und $S \neq \emptyset$ eine beliebige Menge, so wird eine Abbildung $X : \Omega \rightarrow S$ auch S -wertige Zufallsvariable genannt.*

Durch $P^X(B) := P(X^{-1}(B)) \forall B \subset S$ wird ein Wahrscheinlichkeitsmaß P^X auf S definiert, die sogenannte *Verteilung von X* .

(S, P^X) ist ein diskreter Wahrscheinlichkeitsraum

Beweis: Offensichtlich $P^X(B) = P(X^{-1}(B)) \in [0, 1]$

$$P^X(S) = P(X^{-1}(S)) = P(\Omega) = 1$$

$B_i \subset S$ disjunkt ($i \in \mathbb{N}$)

$$\Rightarrow P^X\left(\bigcup_{i \in \mathbb{N}} B_i\right) = P\left(X^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right)\right) = P\left(\underbrace{\bigcup_{i \in \mathbb{N}} X^{-1}(B_i)}_{\text{disjunkt}}\right) = \sum_{i \in \mathbb{N}} P(X^{-1}(B_i)) = \sum_{i \in \mathbb{N}} P^X(B_i),$$

d.h. P^X ist σ -additiv

(Ω, P) diskreter Wahrscheinlichkeitsraum $\Rightarrow \exists \Omega_0 \subset \Omega$ abzählbar mit $P(\Omega_0) = 1$

$S_0 := X(\Omega_0) = \{X(\omega) \mid \omega \in \Omega_0\} \subset S$ abzählbar.

$$P^X(S_0) = P(X^{-1}(X(\Omega_0))) \geq P(\Omega_0) = 1$$

d.h. $P^X(S_0) = 1$

Beispiel 3.2. *2 maliges Werfen eines fairen Würfels*

Es interessiert nur die Augensumme

Modell: $\Omega = \{1, \dots, 6\}^2, P$ Laplace-Verteilung

$$X : \Omega \rightarrow \{2, \dots, 12\} =: S$$

$$X(\omega_1, \omega_2) := \omega_1 + \omega_2$$

Verteilung von P^X von X hat Zähldichte f_X gegeben durch

$$\begin{aligned} k \in S : f_X(k) &= P^X(\{k\}) = P\{X = k\} \\ &= P\{(\omega_1, \omega_2) \in \Omega \mid X(\omega_1, \omega_2) = \omega_1 + \omega_2 = k\} \\ &= P\{(\omega_1, k - \omega_1) \mid 1 \leq \omega_1 \leq 6, 1 \leq k - \omega_1 \leq 6\} \\ &= P\{(\omega_1, k - \omega_1) \mid \max(1, k - 6) \leq \omega_1 \leq \min(6, k - 1)\} \\ &= \begin{cases} \frac{k-1}{36} & \text{falls } 2 \leq k \leq 7 \\ \frac{13-k}{36} & \text{falls } 7 < k \leq 12 \end{cases} \\ &= \frac{6 - |k - 7|}{36} \quad \forall k \in S \end{aligned}$$

Zufallsvariablen sollen als unabhängig angesehen werden, wenn alle durch sie ausdrückbaren Ereignisse stochastisch unabhängig sind.

Definition 3.3 ((P-)stochastische Unabhängigkeit bei Zufallsvariablen). Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum.

Dann heißen Zufallsvariablen $X_i : \Omega \rightarrow S_i, 1 \leq i \leq n$ (P-)stochastisch unabhängig, wenn für beliebige $B_i \subset S_i, 1 \leq i \leq n$, die Ereignisse $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ stochastisch unabhängig sind.

Man beachte: $(X_1, \dots, X_n) : \Omega \rightarrow S_1 \times \dots \times S_n$ ist eine Zufallsvariable.

Satz 3.4. In der Situation von Definition 3.3 sind äquivalent:

(i) X_1, \dots, X_n sind stochastisch unabhängig

(ii) $\forall B_i \subset S_i (1 \leq i \leq n) : P\{X_i \in B_i \forall 1 \leq i \leq n\} = \prod_{i=1}^n P\{X_i \in B_i\}$

(iii) Die Zähldichte $f_{(X_1, \dots, X_n)}$ von $P^{(X_1, \dots, X_n)}$ hat die Form

$$f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i=1}^n h_i(t_i) \quad \forall t_i \in S_i (1 \leq i \leq n)$$

für geeignete Funktionen $h_i : S_i \rightarrow [0, \infty)$

In dem Fall hat P^{X_i} die Zähldichte

$$f_{X_i}(t_i) = c_i \cdot h_i(t_i) \quad \forall t_i \in S_i \text{ mit } c_i = \frac{1}{\sum_{t \in S_i} h_i(t)} \quad (3.1)$$

Beweis:

(i) \Rightarrow (ii) ist klar, da $\{X_i \in B_i\}, 1 \leq i \leq n$, stochastisch unabhängig.

$$\Rightarrow P\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = P\{X_i \in B_i \forall 1 \leq i \leq n\} = \prod_{i=1}^n P\{X_i \in B_i\}$$

(ii) \Rightarrow (iii) Speziell für $B_i = \{t_i\}$

$$\begin{aligned} P\{X_i = t_i \forall 1 \leq i \leq n\} &= P\{X_i \in B_i \forall 1 \leq i \leq n\} \\ &\stackrel{(ii)}{=} \prod_{i=1}^n P\{X_i \in B_i\} \\ &= \prod_{i=1}^n P\{X_i = t_i\} \end{aligned}$$

$$\text{d.h. } f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = P\{(X_1, \dots, X_n) = (t_1, \dots, t_n)\} = \prod_{i=1}^n P\{X_i = t_i\} = \prod_{i=1}^n f_{X_i}(t_i)$$

d.h. (iii)

(iii) \Rightarrow (3.1)

$$\begin{aligned} f_{X_i}(t_i) &= P\{X_i = t_i\} \\ &= P\{X_i = t_i, X_j \in S_j \forall j \neq i\} \\ &= \sum_{\substack{t_j \in S_j \\ j \neq i}} f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) \\ &\stackrel{(iii)}{=} \sum_{\substack{t_j \in S_j \\ j \neq i}} \prod_{k=1}^n h_k(t_k) \\ &= \sum_{t_1 \in S_1} \sum_{t_2 \in S_2} \dots \sum_{t_{i-1} \in S_{i-1}} \sum_{t_{i+1} \in S_{i+1}} \dots \sum_{t_n \in S_n} \prod_{k=1}^n h_k(t_k) \\ &= h_i(t_i) \cdot \prod_{\substack{j=1 \\ i \neq j}}^n \left(\sum_{t_j \in S_j} h_j(t_j) \right) =: c_i \quad (\text{Distributivgesetz}) \end{aligned}$$

Die Darstellung für c_i folgt aus der Bedingung $\sum_{t \in S_i} f_{X_i}(t) = 1$ an Zähldichten

(iii) \Rightarrow (ii) O.E. $h_i = f_{X_i}$

$\forall B_i \in \mathcal{S}_i$

$$\begin{aligned} P\{X_i \in B_i \forall 1 \leq i \leq n\} &= P\{(X_1, \dots, X_n) \in B_1 \times \dots \times B_n\} \\ &= \sum_{(t_1, \dots, t_n) \in B_1 \times \dots \times B_n} f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) \\ &= \prod_{i=1}^n \sum_{t_i \in B_i} f_{X_i}(t_i) \quad (\text{Distributivgesetz}) \\ &= \prod_{i=1}^n P\{X_i \in B_i\} \end{aligned}$$

(ii) \Rightarrow (i) z.Z. $\forall J \subset \{1, \dots, n\}, J \neq \emptyset: \{X_j \in B_j \forall j \in J\} = \prod_{j \in J} P\{X_j \in B_j\}$

Dies folgt sofort aus (ii) mit $B_i := \mathcal{S}_i \forall i \notin J$, denn dann $P\{X_i \in B_i\} = P(\Omega) = 1$ und $\{X_j \in B_j \forall j \in J\} = \{X_i \in B_i \forall 1 \leq i \leq n\}$

Beispiel 3.5. Zufallsexperiment mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ werde n -mal unabhängig durchgeführt.

Modell: $\Omega = \{0, 1\}^n$,
wobei 0 einem Misserfolg und 1 einem Erfolg entspricht.

Def. Wahrscheinlichkeitsmaß auf Ω mit Zähldichte
 $f(\omega) = P\{(\omega_1, \dots, \omega_n)\} = p^{\sum_{i=1}^n \omega_i} \cdot (1-p)^{\sum_{i=1}^n (1-\omega_i)}$
(denn bei jeder Durchführung Erfolgswahrscheinlichkeit p
 \rightsquigarrow Faktor p , wenn $\omega_i = 1$, d.h. i -tes Experiment erfolgreich
Faktor $(1-p)$, wenn $\omega_i = 0$, d.h. i -tes Experiment Misserfolg)

Zufallsvariable $X_i: \Omega \rightarrow \{0, 1\}, X_i(\omega) = \omega_i$, d.h.

$$X_i = \begin{cases} 0, & \text{falls } i\text{-tes Experiment Misserfolg} \\ 1, & \text{falls } i\text{-tes Experiment Erfolg} \end{cases}$$

Beh.: X_1, \dots, X_n sind stochastisch unabhängig.

Denn: Zähldichte von $P^{(X_1, \dots, X_n)}$

$$\begin{aligned} f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) &= P\{(X_1, \dots, X_n) = (t_1, \dots, t_n)\} \\ &= P\{\omega \in \Omega \mid X_i(\omega) = \omega_i = t_i \forall 1 \leq i \leq n\} \\ &= P\{(t_1, \dots, t_n)\} = p^{\sum_{i=1}^n t_i} \cdot (1-p)^{\sum_{i=1}^n (1-t_i)} \\ &= \prod_{i=1}^n (p^{t_i} \cdot (1-p)^{(1-t_i)}) \end{aligned}$$

hat Produktgestalt.

Betrachte die Zufallsvariable $Y := \sum_{i=1}^n X_i$, also die Anzahl der Erfolge in n Experimenten.

P^Y hat Zähldichte

$$\begin{aligned} f_Y(k) &= P\{Y = k\} = P\{(\omega_1, \dots, \omega_n) \in \Omega \mid Y(\omega_1, \dots, \omega_n) = \sum_{i=1}^n X_i(\omega) = \sum_{i=1}^n \omega_i = k\} \\ &= \sum_{\substack{(\omega_1, \dots, \omega_n) \in \Omega \\ \sum_{i=1}^n \omega_i = k}} p^k \cdot (1-p)^{n-k} \\ &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \end{aligned}$$

denn es gibt genau $\binom{n}{k}$ Möglichkeiten, die k Stellen im Vektor $(\omega_1, \dots, \omega_n)$ auszuwählen, an denen eine 1 steht.

Daher: $P^Y = \mathcal{B}_{(n,p)}$: Binomialverteilung (\rightarrow Bsp 1.11.(i))

Speziell $n = 1$: $\mathcal{B}_{(1,p)}$: Bernoulli-Verteilung

Also sind X_1, \dots, X_n stochastisch unabhängig mit $P^{X_i} = \mathcal{B}_{(1,p)}$

Beispiel 3.6. Capture-Recapture-Verfahren

Ziel: Schätze Anzahl N der Fische in einem See

Dazu:

1. Fange M Fische, markiere sie und lasse sie wieder frei.
2. Fange wieder n Fische, darunter seien m markierte.

Annahme: Fangwahrscheinlichkeit unter 2. sei für markierte und unmarkierte Fische gleich

Modell: n -mal Ziehen ohne Zurücklegen aus N Fischen und ohne Berücksichtigung der Reihenfolge.

$$\Omega = \{(\omega_1, \dots, \omega_n) \in \{1, \dots, N\}^n \mid \omega_1 < \omega_2 < \dots < \omega_n\}$$

markierte Fische entsprechen den Nummern $1, \dots, M$

P Laplace-Verteilung

$$X(\omega) = \sum_{i=1}^n 1_{\{1, \dots, M\}}(\omega_i) : \text{Anzahl der markierten Fische}$$

$$\text{Erinnerung: } 1_{\{1, \dots, M\}}(\omega_1) = \begin{cases} 1 & \text{falls } \omega_i \in \{1, \dots, M\} \\ 0 & \text{falls } \omega_i \notin \{1, \dots, M\} \end{cases}$$

mögliche Werte von $X(\omega)$: $0 \leq X(\omega) \leq \min(M, n)$, $n - X(\omega) \leq N - M$

$$\Rightarrow \text{Wertebereich von } X : S := \{\max(0, n + M - N), \dots, \min(n, M)\}$$

Berechne: $P\{X = m\} = \frac{|\{\omega \in \Omega \mid X(\omega) = m\}|}{|\Omega|}$

Es gibt genau $\binom{M}{m}$ Möglichkeiten, m markierte Fische aus M markierten Fischen zu ziehen.

Es gibt genau $\binom{N-M}{n-m}$ Möglichkeiten, $n - m$ nicht markierte Fische aus allen $N - M$ nicht markierten Fischen zu ziehen.

$$\Rightarrow P\{X = m\} = \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}} =: \mathcal{H}_{(N,M,n)}(\{m\}) \quad \forall m \in S$$

Die rechte Seite gibt die Zähl-dichte eines Wahrscheinlichkeitsmaßes auf S an, der so genannten hypergeometrische Verteilung $\mathcal{H}_{(N,M,n)}$.

Es gilt also $P^X = \mathcal{H}_{(N,M,n)}$

Die hypergeometrische Verteilung beschreibt also die Anzahl der "Erfolge" beim Ziehen ohne Zurücklegen, während die Binomialverteilung die Anzahl beim Ziehen mit Zurücklegen beschreibt.

Falls $n \ll N$, dann ist Ziehen mit oder ohne Zurücklegen fast identisch und daher

$$\mathcal{H}_{(N,M,n)}\{m\} \approx \mathcal{B}_{(n, \frac{M}{N})}\{m\} \quad \forall 0 \leq m \leq n$$

Wir werden sehen: erwartete Anzahl markierter Fische $= n \cdot \frac{M}{N}$, wobei N unbekannt ist.

Einen Schätzer für N kann man so motivieren, dass die tatsächlich beobachtete Zahl $X(\omega)$ markierter Fische in etwa gleich der erwarteten Anzahl gesetzt wird:

$$X(\omega) \approx n \cdot \frac{M}{N} \rightsquigarrow \hat{N} := \left[n \cdot \frac{M}{X(\omega)} \right]$$

Im Fall, dass die Anzahl der Experimente n groß und die Erfolgswahrscheinlichkeit p klein ist, kann $\mathcal{B}_{(n,p)}$ approximiert werden durch eine einfachere Verteilung.

Satz & Definition 3.7 (Poisson'scher Grenzwertsatz, "Gesetz der kleinen Zahlen").
Ist $p_n \in (0, 1)$, $n \in \mathbb{N}$, so dass $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda > 0$, so gilt für $k \in \mathbb{N}_0$

$$\lim_{n \rightarrow \infty} \mathcal{B}_{(n,p_n)}(\{k\}) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} =: f_\lambda(k)$$

$f_\lambda : \mathbb{N}_0 \rightarrow [0, 1]$ ist eine Zähl-dichte.

Das zugehörige Wahrscheinlichkeitsmaß \mathcal{P}_λ heißt Poisson-Verteilung mit Parameter λ .

Beweis: $(1 - \frac{x_n}{n})^n \rightarrow e^{-x}$ falls $x_n \rightarrow x$

$$\mathcal{B}_{(n,p_n)}\{k\} = \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{n \cdot n \cdot \dots \cdot n} \cdot (n \cdot p_n)^k \cdot (1 - \frac{n \cdot p_n}{n})^n \cdot (1 - p_n)^{-k} \cdot \frac{1}{k!}$$

$$\xrightarrow{n \rightarrow \infty} \frac{\lambda^k \cdot e^{-\lambda}}{k!},$$

da $(n - i)/n \rightarrow 1 \quad \forall 0 \leq i \leq k - 1$, $(1 - np_n/n)^n \rightarrow e^{-\lambda}$ und wegen $np_n \rightarrow \lambda \Rightarrow p_n \rightarrow 0$ schließlich $(1 - p_n)^{-k} \rightarrow 1$.

Summen unabhängiger Zufallsvariablen

Definition 3.8 (Faltung). Sind X, Y stochastisch unabhängige \mathbb{R} -wertige Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) , so heißt die Verteilung P^{X+Y} der Summe $X+Y$ die Faltung von P^X und P^Y , in Zeichen $P^X * P^Y$

Ebenso wird die Zähldichte f_{X+Y} von P^{X+Y} die Faltung der Zähldichten f_X von P^X und f_Y von P^Y genannt, i. Z. $f_X * f_Y$

Satz 3.9. Sind in der Situation von Def. 3.8 X und Y \mathbb{Z} -wertig, so gilt

$$f_X * f_Y(n) = \sum_{k \in \mathbb{Z}} f_X(k) \cdot f_Y(n - k) \quad \forall n \in \mathbb{Z}$$

Beweis:

$$\begin{aligned} f_X * f_Y(n) &= P\{X + Y = n\} \\ &= \sum_{k \in \mathbb{Z}} P\{X = k, Y = n - k\} \\ &= \sum_{k \in \mathbb{Z}} P\{X = k\} \cdot P\{Y = n - k\} \\ &= \sum_{k \in \mathbb{Z}} f_X(k) \cdot f_Y(n - k) \end{aligned}$$

Beispiel 3.10. Seien $\lambda_1, \lambda_2 > 0$. Dann hat $\mathcal{P}_{\lambda_1} * \mathcal{P}_{\lambda_2}$ die Zähldichte

$$\begin{aligned} f_{\lambda_1} * f_{\lambda_2}(n) &= \sum_{k=0}^{\infty} f_{\lambda_1}(k) \cdot f_{\lambda_2}(n - k) \\ &= \sum_{k=0}^n e^{-\lambda_1} \cdot \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2} \cdot \frac{\lambda_2^{(n-k)}}{(n-k)!} \\ &= e^{-(\lambda_1 + \lambda_2)} \cdot \frac{1}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \cdot \lambda_1^k \cdot \lambda_2^{n-k} \\ &= e^{-(\lambda_1 + \lambda_2)} \cdot \frac{1}{n!} \cdot (\lambda_1 + \lambda_2)^n = f_{\lambda_1 + \lambda_2}(n) \quad \forall n \in \mathbb{N}_0 \end{aligned}$$

Also $\mathcal{P}_{\lambda_1} * \mathcal{P}_{\lambda_2} = \mathcal{P}_{\lambda_1 + \lambda_2}$

Beispiel 3.11 (Quicksort). Die Zahlen x_1, \dots, x_n , die alle als verschieden angenommen werden, sollen sortiert werden.

Der Algorithmus Quicksort erledigt diese Aufgabe wie folgt:

1. Wähle zufällig gleichverteilt ein x_j aus

-
2. Ordne Zahlen $x_i < x_j$ links von $x_j \rightsquigarrow X_l$ Vektor von Zahlen $< x_j$
 3. Ordne Zahlen $x_i > x_j$ rechts von $x_j \rightsquigarrow X_r$ Vektor von Zahlen $> x_j$
 4. Verfahre mit X_l und X_r getrennt ebenso usw.

Beispiel. 3 7 2 6 13 1

1. Wahl 7: $\underbrace{3\ 2\ 6\ 1}_{X_l} | 7 | \underbrace{13}_{X_r} \rightsquigarrow 5$ Vergleiche

2. Wahl 3: $2\ 1 | 3 | 6 | 7 | 13 \rightsquigarrow 3$ Vergleiche
 $1\ 2 | 3 | 6 | 7 | 13 \rightsquigarrow 1$ Vergleich

9 Vergleiche

worst case: gewählte Zahl jeweils kleinste oder größte
 Anzahl Vergleiche: $(n-1) + (n-2) + \dots + 1 = \frac{n \cdot (n-1)}{2}$

best case: Speziell $n = 2^k - 1$
 ausgewählte Zahl jeweils mittlere

1. Schritt $2^k - 2$ Vergleiche $\rightsquigarrow 2$ Blöcke mit Länge $2^{k-1} - 1$
 2. Schritt $2 \cdot (2^{k-1} - 2)$ Vergleiche $\rightsquigarrow 4$ Blöcke mit Länge $2^{k-2} - 1$ usw.
- \rightsquigarrow Gesamtzahl der Vergleiche:
 $(2^k - 2) + 2 \cdot (2^{k-1} - 2) + \dots + 2^{k-2} \cdot (2^2 - 2) = (k-2) \cdot 2^k + 2 \approx n \cdot \log_2 n$

Später: mittlere Zahl der benötigten Vergleiche $\approx \underbrace{\frac{2}{\log 2}}_{\approx 2,89} \cdot \text{Anzahl der Vergleiche beim best case}$

Hier: Bestimme die Zähldichte der zufälligen Anzahl von Vergleichen, die zum Sortieren benötigt werden.

Wird die Zahl jeweils zufällig gleichverteilt gewählt, so hängt die Verteilung der Anzahl von Vergleichen nur von der Anzahl n der Daten ab, nicht von deren Reihenfolge.

$Z(X)$ = Zahl der Vergleiche, um Vektor X zu sortieren
 $Z(x_1, \dots, x_n) = n - 1 + Z(X_l) + Z(X_r)$

Angenommen k -te kleinste Zahl ausgewählt:
 Dann hat X_l die Länge $k - 1$ und X_r die Länge $n - k$

Zähldichte von $Z(x_1, \dots, x_n)$: $f_n(m) := P\{Z(x_1, \dots, x_n) = m\}$ (hängt von n ab, nicht von den genauen Werten x_1, \dots, x_n)

Dann
 $P\{Z(x_1, \dots, x_n) = l\} = P\{Z(X_l) + Z(X_r) = l - (n - 1)\}$
 Bei gegebenem Wert k sind $Z(X_l), Z(X_r)$ stochastisch unabhängig, da die Zahlen jeweils

unabhängig gewählt werden.

Da jeder Wert von k mit Wahrscheinlichkeit $\frac{1}{n}$ angenommen wird, folgt

$$\begin{aligned} f_n(l) &= P\{Z(X_l) + Z(X_r) = l - n + 1\} \\ &= \sum_{k=1}^n \frac{1}{n} \cdot f_{k-1} * f_{n-k}(l - n + 1) \\ &= \frac{1}{n} \sum_{k=1}^n \sum_{j=0}^{\infty} f_{k-1}(j) \cdot f_{n-k}(l - n + 1 - j) \end{aligned}$$

$$f_1(0) = 1$$

$$f_2(1) = 1$$

$f_n, n \geq 3$ kann wie oben rekursiv berechnet werden.

----- PDF-Scan eines Buches mit Grafen zu berechneten Werten -----

Definition 3.12 (Markov-Kette und -Eigenschaft & Übergangswahrscheinlichkeit).

Eine Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}_0}$ von S -wertigen Zufallsvariablen (S höchstens abzählbar) heißt Markov-Kette, falls für alle $n \in \mathbb{N}, s_0, \dots, s_{n+1} \in S$ mit $P\{X_n = s_n, X_{n-1} = s_{n-1}, \dots, X_0 = s_0\} > 0$

$$\begin{aligned} P(X_{n+1} = s_{n+1} \mid X_n = s_n, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) \\ = P(X_{n+1} = s_{n+1} \mid X_n = s_n) \quad (\text{Markov-Eigenschaft}) \end{aligned}$$

Die Markov-Kette heißt homogen, falls die Übergangswahrscheinlichkeiten $p_{ts} := P(X_{n+1} = t \mid X_n = s)$ für alle $n \in \mathbb{N}_0$ mit $P\{X_n = s\} > 0$ gleich sind

Bei (homogenen) Markov-Ketten hängt das zukünftige Verhalten nur vom gegenwärtigen Zustand ab, nicht von der echten Vergangenheit.

Stochastisches Verhalten einer homogenen Markov-Kette ist eindeutig bestimmt durch die Startverteilung P^{X_0} (z.B. durch Angabe der zugehörigen Zähldichte $f_0(s) = P\{X_0 = s\}$) und die Übergangswahrscheinlichkeit $p_{ts}, s, t \in S$.

Dann z.B.

$$\begin{aligned} P\{X_0 = s_0, X_1 = s_1, X_2 = s_2\} &= P(X_2 = s_2 \mid X_1 = s_1, X_0 = s_0) \cdot P\{X_1 = s_1, X_0 = s_0\} \\ &= p_{s_2 s_1} \cdot P(X_1 = s_1 \mid X_0 = s_0) \cdot P\{X_0 = s_0\} \\ &= p_{s_2 s_1} \cdot p_{s_1 s_0} \cdot f_0(s_0) \end{aligned}$$

$$\text{Allgemein: } P\{X_i \in A_i \mid \forall 0 \leq i \leq n\} = \sum_{\substack{s_i \in A_i \\ (0 \leq i \leq n)}} f_0(s_0) \cdot p_{s_0 s_1} \cdot p_{s_1 s_2} \cdot \dots \cdot p_{s_{n-1} s_n}$$

Beispiel 3.13. Spieler A und B spielen das folgende Glücksspiel: Spieler A bestimmt den Einsatz und wirft eine Münze.

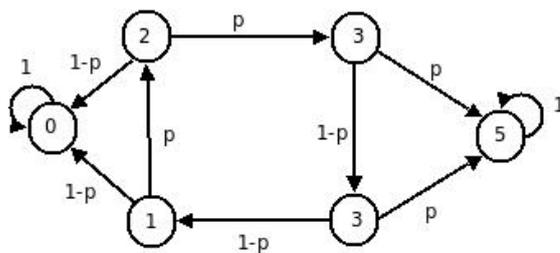
Falls Kopf fällt, so zahlt B den Einsatz an A, sonst zahlt A den Einsatz an B.

Spieler A hat 1 € und benötigt 5 €.

Er entschließt sich daher, die folgende "Kühne Strategie" anzuwenden: A setzt jeweils sein gesamtes Kapital, wenn er $\leq \frac{5}{2}$ € hat, sonst die Differenz aus 5 € und seinem Kapital (so dass er in diesem Fall, die 5 € zusammen hat, wenn Kopf fällt). Das Spiel ist beendet, sobald Spieler A sein gesamtes Kapital verspielt hat oder aber 5 € besitzt und das Spiel beendet.

Die Zufallsvariable X_n bezeichne das Kapital von A nach n Spielen. (Sobald das Spiel beendet ist, soll sich der Wert nicht mehr ändern.) Die Wahrscheinlichkeit, dass Kopf fällt, betrage jeweils p .

$(X_n)_{n \in \mathbb{N}_0}$ ist dann eine homogene Markov-Kette. Die folgende Grafik gibt alle möglichen Übergänge von X_n nach X_{n+1} sowie die zugehörigen Übergangswahrscheinlichkeiten wieder:



Z.B. ist von $X_n = 4$ nur ein Wechsel nach $X_{n+1} = 5$ mit Wahrscheinlichkeit p oder nach $X_{n+1} = 3$ mit Wahrscheinlichkeit $1 - p$ möglich, d.h. $p_{54} = p$, $p_{34} = 1 - p$, $p_{i4} = 0 \forall i \notin \{3, 5\}$

Die Zustände 1 und 5 heißen absorbierend, da sie nicht wieder verlassen werden können.

Mit welcher Wahrscheinlichkeit erhält Spieler A schließlich 5 €?

$$p_k = P(X_l = 5 \text{ für ein } l > n \mid X_n = k) \quad (\text{hängt nicht von } n \text{ ab})$$

Dann:

$$\begin{aligned}
p_1 &= P(X_l = 5 \text{ für ein } l > n | X_n = 1) \\
&= P(X_{n+1} = 2, X_l = 5 \text{ für ein } l > n | X_n = 1) + \underbrace{P(X_{n+1} = 0, X_l = 5 \text{ für ein } l > n | X_n = 1)}_0 \\
&= \frac{P\{X_n = 1, X_{n+1} = 2, X_l = 5 \text{ für ein } l > n + 1\}}{P\{X_n = 1, X_{n+1} = 2\}} \cdot \frac{P\{X_n = 1, X_{n+1} = 2\}}{P\{X_n = 1\}} \\
&= P(X_l = 5 \text{ für ein } l > n + 1 | X_n = 1, X_{n+1} = 2) \cdot P(X_{n+1} | X_n = 1) \\
&\stackrel{*}{=} P(X_l = 5 \text{ für ein } l > n + 1 | X_{n+1} = 2) p_{21} \\
&= p_2 \cdot p
\end{aligned}$$

* = Markov-Eigenschaft

Ebenso erhält man die Beziehungen:

$$p_2 = p \cdot p_4$$

$$p_3 = p + (1 - p) \cdot p_1$$

$$p_4 = p + (1 - p) \cdot p_3$$

Diese 4 Gleichungen bilden ein lineares Gleichungssystem, das in Matrixschreibweise wie folgt aussieht:

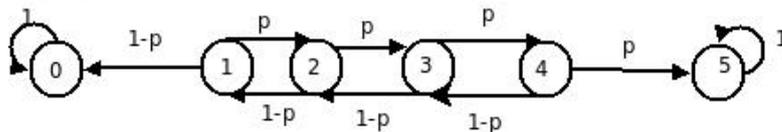
$$\underbrace{\begin{pmatrix} 1 & -p & 0 & 0 \\ 0 & 1 & 0 & -p \\ p-1 & 0 & 1 & 0 \\ 0 & 0 & p-1 & 1 \end{pmatrix}}_A \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ p \\ p \end{pmatrix}$$

Da die Matrix A invertierbar ist, existiert eine eindeutige Lösung p_1, \dots, p_4

$$\text{Es gilt insbesondere } p_1 = \frac{(2-p) \cdot p^3}{1-p^2+2p^3-p^4}$$

Zum Vergleich betrachten wir einen vorsichtigen Spieler \tilde{A} , der in der Situation von Spieler A immer nur $1 \in$ setzt.

Man erhält dann die folgende Grafik der möglichen Übergänge mit den zugehörigen Übergangswahrscheinlichkeiten:



Mit analogen Überlegungen wie oben erhält man

$$\begin{pmatrix} 1 & -p & 0 & 0 \\ p-1 & 1 & -p & 0 \\ 0 & p-1 & 1 & -p \\ 0 & 0 & p-1 & 1 \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ p \end{pmatrix}$$

Dieses System ist ebenfalls eindeutig lösbar und es gilt nun $p_1 = \frac{4}{1 - 3p + 4p^2 - 2p^3 + p^4}$.

Der Vergleich der Wahrscheinlichkeiten p_1 , dass Spieler A bzw. \tilde{A} mit der jeweiligen Strategie schließlich 5 € erhält, zeigt:

- Ist $p < 1/2$, d.h. ist die Münze ungünstig für Spieler A bzw. \tilde{A} , so ist die Erfolgswahrscheinlichkeit p_1 bei der kühnen Strategie größer als bei der vorsichtigen Strategie. (Dies gilt auch für die Wahrscheinlichkeit p_k , wenn Spieler A zu Beginn k € besitzt ($1 \leq k \leq 4$).)

Man kann zeigen: Die kühne Strategie maximiert die Erfolgswahrscheinlichkeit p_k !

- Ist $p > 1/2$, d.h. ist die Münze günstig für Spieler A bzw. \tilde{A} , so ist die Erfolgswahrscheinlichkeit p_1 bei der vorsichtigen Strategie größer als bei der kühnen Strategie.

- Ist $p = 1/2$, d.h. ist die Münze fair, so beträgt die Erfolgswahrscheinlichkeit p_k bei beiden Strategien $p_k = k/5$.

Man kann zeigen: Dies gilt für alle möglichen Strategien, d.h. bei einem fairen Spiel kann man seine Gewinnwahrscheinlichkeit nicht durch eine geschickte Spielstrategie verbessern!

Beispiel 3.14 (Ranking-Verfahren bei Google). Angenommen ein Surfer startet bei einer zufällig ausgewählten Website, wählt dann zufällig gleichverteilt einen Link, der von dieser Seite wegführt und verfährt bei der nächsten Seite wieder genauso.

Die Wahl des Links sei unabhängig von dem Weg, auf dem die Seite erreicht wurde.

Nach m Schritten (m "groß") befindet sich Surfer auf einer zufälligen Website und die Wahrscheinlichkeit, dass er sich auf einer bestimmten Seite befindet, ist Maß für die Relevanz dieser Seite.

Modell

Webseiten durchnummeriert $1, \dots, N$

$X_m = N_r$ der Website, auf der der Surfer nach m Schritten ist.

S -wertige Zufallsvariable mit $S = \{1, \dots, N\}$

L_i : Menge der Seiten, auf die von Seite i aus ein Link führt (Annahme: $L_i \neq \emptyset$)

Dann

$$\begin{aligned}
q_{ij} &:= P(X_{m+1} = j \mid X_m = i) = \begin{cases} 0 & \text{falls } j \in L_i \\ \frac{1}{|L_i|} & \text{falls } i \in L_i \end{cases} \\
&= P(X_{m+1} = j \mid X_m = i, X_{m-1} = i_{m-1}, \dots, X_0 = i_0) \quad \forall (i_0, \dots, i_{m-1}, i, j) \in S^{m+2} \\
&\quad \text{mit } P\{X_m = i, X_{m-1} = i_{m-1}, \dots, X_0 = i_0\} > 0
\end{aligned}$$

Also ist $(X_m)_{m \geq 0}$ eine homogene Markovkette

Gesucht: $P\{X_m = i\} \forall i \in S$ für "großes m "

Definierte $p^{(m)} = (p_i^{(m)})_{i \in S}$ mit $p_i^{(m)} = P\{X_m = i\}$

Übergangsmatrix $Q = (q_{ji})_{1 \leq i, j \leq N} = (q_{ji})_{i, j \in S}$

Q ist eine stochastische Matrix, d.h. alle Spaltensummen sind gleich 1

Denn

$$\sum_{j=1}^N q_{ji} = \sum_{j \in S} P(X_{m+1} = j \mid X_m = i) = P(X_{m+1} \in S \mid X_m = i) = 1$$

Mit dem Satz von der totalen Wahrscheinlichkeit ergibt sich:

$$\begin{aligned}
p_j^{(m+1)} &= P\{X_{m+1} = j\} = \sum_{i=1}^N P(X_{m+1} = j \mid X_m = i) \cdot P(X_m = i) \\
&= \sum_{i=1}^N q_{ji} \cdot p_i^{(m)} = (Q \cdot p^{(m)})_j \quad \forall j \in S \\
&\Rightarrow p^{m+1} = Q \cdot p^{(m)} = Q \cdot (Q \cdot p^{(m-1)}) = Q^2 \cdot p^{(m-1)} = \dots = Q^{m+1} \cdot p^0
\end{aligned}$$

Frage: Wie verhält sich $p^{(m)}$ für $m \rightarrow \infty$?

Erinnerung: λ Eigenwert der Matrix A mit Eigenvektor v : $\Leftrightarrow Av = \lambda v$

Satz von Perron:

Ist Matrix A nur mit strikt positiven Einträgen, so existiert ein Eigenwert $\lambda_1 > 0$, der betragsmäßig strikt größer ist als alle anderen Eigenwerte von A ; es existiert ein Eigenvektor v_1 zu λ_1 , der strikt positive Einträge hat.

Ferner:

$$\frac{A^k \cdot v}{\lambda_1^k} \longrightarrow cv_1$$

für alle Vektoren v mit positiven Einträgen und ein $c > 0$ (das von v abhängt)

Ist A eine stochastische Matrix und ist die Summe der Einträge von v gleich 1, dann ist auch bei $A \cdot v$ die Summe der Einträge = 1, also auch bei $A^k \cdot v$.

Daher $\lambda_1 = 1$ und Summen der Einträge von $cv_1 = 1$.

Annahme (*): Es gibt ein m_0 , so dass Q^{m_0} nur strikt positive Einträge hat, d.h. man kann in m_0 Schritten von jeder Website zu jeder beliebigen Website gelangen.

Dann: Satz von Perron mit $A = Q^{m_0}$

$$p^{(m)} = Q \cdot p^{(0)} = A^{\frac{m}{m_0}} \cdot p^{(0)} \xrightarrow{m \rightarrow \infty} p_s \quad (m \text{ Vielfaches von } m_0)$$

wobei p_s ein Eigenvektor mit Einträgen > 0 mit Eigenwert 1 und Spaltensumme 1 ist.

$\Rightarrow (p_s)_i$ ist das gesuchte Maß für die Relevanz der Seite i .

Zugehöriges Wahrscheinlichkeitsmaß $P_s\{i\} = (p_s)_i$ heißt stationäre Verteilung der Markov-Kette.

Ist $P\{X_0 = i\} = (p_s)_i \forall i$,

$$\text{dann } P\{X_{m_0} = i\} = (Q^{m_0} \cdot p_i^{(0)}) = (A \cdot p_s)_i = (p_s)_i = P\{X_0 = i\}$$

- Konvergenz $A^k \cdot v \xrightarrow{k \rightarrow \infty} p_s$ erfolgt exponentiell schnell.

$\rightsquigarrow p_s$ relativ effizient berechenbar

- Annahme (*) lässt sich vermeiden durch folgende Modifikation:

Für (kleines) $\alpha \in (0, 1)$ wählt der Surfer jeweils gleichverteilt einen Link mit der Wahrscheinlichkeit $1 - \alpha$ aus und wählt eine beliebige Website mit Wahrscheinlichkeit α gemäß der Verteilung \tilde{p} (d.h. Website i mit Wahrscheinlichkeit $\alpha \cdot \tilde{p}_i$).

\rightsquigarrow neue Übergangsmatrix

$$(1 - \alpha) \cdot Q + \alpha \cdot \begin{pmatrix} \tilde{q}_1 & \dots & \tilde{q}_1 \\ \vdots & & \vdots \\ \tilde{q}_N & \dots & \tilde{q}_N \end{pmatrix}$$

erfüllt die Annahme (*) falls alle $\tilde{p}_i > 0$

4 Erwartungswert und Momente von Zufallsvariablen

Definition 4.1. Erwartungswert & Mittelwert

(i) Der Erwartungswert einer \mathbb{R} -wertigen Zufallsvariable X auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) ist def. als $E_P(X) = E(X) := \sum_{x \in X(\Omega)} x \cdot P\{X = x\}$, falls

$$\sum_{x \in X} |x| \cdot P\{X = x\} < \infty \quad (\text{Andernfalls besitzt } X \text{ keinen Erwartungswert})$$

(ii) Sei (\mathbb{R}, Q) ein diskreter Wahrscheinlichkeitsraum, dann heißt $\mu(Q) := \sum_{x \in \mathbb{R}} x \cdot Q(\{x\})$

$$\underline{\text{Mittelwert}} \text{ von } Q, \text{ falls } \sum_{x \in X} |x| \cdot Q(\{x\}) < \infty$$

Bemerkung 4.2. $Id : \mathbb{R} \rightarrow \mathbb{R}, Id(x) := x$ (Identität auf \mathbb{R})

$$\Rightarrow \mu(Q) = E_Q(Id)$$

Umgekehrt: $E_P(x) = \mu(P^X)$,

denn

$$\mu(P^X) = \sum_{x \in \mathbb{R}} x \cdot P^X(\{x\}) = \sum_{x \in \mathbb{R}} x \cdot P\{X = x\} = \sum_{x \in X(\Omega)} x \cdot P\{X = x\} = E_P(X)$$

Satz 4.3. Transformationssatz Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum, $X : \Omega \rightarrow S$ eine Zufallsvariable und $g : S \rightarrow \mathbb{R}$. Dann gilt

$$E_P(\underbrace{g(X)}_{g \circ X}) = E_{P^X}(g)$$

Beweis:

$$\begin{aligned} E_{P^X}(g) &= \sum_{y \in g(S)} y \cdot P^X\{g = y\} \\ &= \sum_{y \in g(S)} y \cdot P(X^{-1}(\{g = y\})) \\ &= \sum_{y \in g(S)} y \cdot P(\underbrace{X^{-1}(\{g = y\})}_{\{g(X)=y\}}) \\ &= \sum_{y \in g(X)(\Omega)} y \cdot P\{g(X) = y\} = E_P(g(X)) \text{ ,falls alle Summen absolut konvergieren.} \end{aligned}$$

Beispiel 4.4.

$$(i) X = 1_A, \text{ d.h. } 1_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases}$$

$$\Rightarrow E(X) = 0 \cdot P\{X = 0\} + 1 \cdot P\{X = 1\} = P(A)$$

$$(ii) X \text{ sei } \mathcal{P}_\lambda\text{-verteilt, d.h. } P\{X = n\} = e^{-\lambda} \cdot \frac{\lambda^n}{n!} \quad \forall n \in \mathbb{N}_0$$

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

Satz 4.5. Seien X, Y Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) , die Erwartungswerte besitzen. Dann gilt:

$$(i) E(aX + Y) = a \cdot E(X) + E(Y) \quad (\forall a \in \mathbb{R})$$

$$(ii) \text{ gilt } X(\omega) \leq Y(\omega) \quad \forall \omega \in \Omega, \text{ dann } E(X) \leq E(Y)$$

Beweis: Im Wesentlichen direktes Nachrechnen.

Beispiel 4.6. Sei Y eine Zufallsvariable auf (Ω, P) mit $P^X = \mathcal{B}_{(n,p)}$
Gesucht: $E(X)$

1. Lösung: Berechnen nach Definition

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \cdot P\{X = k\} = \sum_{k=0}^n k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad (\rightarrow \text{Bsp. 1.11}) \\ &= n \cdot p \cdot \sum_{k=1}^n \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \cdot p^{k-1} \cdot (1-p)^{(n-1)-(k-1)} \\ &\stackrel{j=k-1}{=} n \cdot p \cdot \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} \cdot p^j \cdot (1-p)^{n-1-j} \\ &= n \cdot p \cdot (p + 1 - p)^{n-1} = n \cdot p \end{aligned}$$

2. Lösung: Bsp 3.5:

X_i seien unabhängige Zufallsvariablen mit

$$P\{X_i = 1\} = p, P\{X_i = 0\} = 1 - p$$

$\Rightarrow \sum_{i=1}^n X_i$ ist $\mathcal{B}_{(n,p)}$ -verteilt, hat also die gleiche Verteilung wie X

$$\Rightarrow E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n (0 \cdot P\{X_i = 0\} + 1 \cdot P\{X_i = 1\}) = n \cdot p$$

Beispiel 4.7. Erwartete Laufzeit von Quicksort (Fortsetzung von Bsp. 3.11)

$Z(x_1, \dots, x_n)$ = Zufällige Zahl der Vergleiche, die benötigt werden, um x_1, \dots, x_n zu sortieren.

Mit Wahrscheinlichkeit $\frac{1}{n}$ wähle eine Zahl x_j ($1 \leq j \leq n$)

Sortiere Zahlen $< x_j$ in Vektor X_l und Zahlen $> x_j$ in Vektor X_r

Dann

$$Z(x_1, \dots, x_n) = n - 1 + Z(X_l) + Z(X_r)$$

$$\mu_n = E(Z(x_1, \dots, x_n)) \quad (\text{hängt nur von } n \text{ ab!})$$

Falls x_j die k kleinste Zahl ist, dann hat X_l die Länge $k-1$ also

$$E(Z(X_l)) = \mu_{k-1} \quad \text{und ebenso}$$

$$E(Z(X_r)) = \mu_{n-k}$$

Jeder Wert von k tritt mit der Wahrscheinlichkeit $\frac{1}{n}$ auf.

Also

$$\begin{aligned} \mu_n &= E(Z(x_1, \dots, x_n)) = n - 1 + E(Z(X_l)) + E(Z(X_r)) \\ &= n - 1 + \sum_{k=1}^n \frac{1}{n} \cdot \mu_{k-1} + \sum_{k=1}^n \frac{1}{n} \cdot \mu_{n-k} \\ &= n - 1 + \frac{1}{n} \cdot \sum_{j=0}^{n-1} \mu_j + \frac{1}{n} \cdot \sum_{j=0}^{n-1} \mu_j \\ &= n - 1 + \frac{2}{n} \cdot \sum_{j=0}^{n-1} \mu_j \end{aligned}$$

$$\begin{aligned}
\Rightarrow & n \cdot \mu_n = n \cdot (n-1) + 2 \cdot \sum_{j=0}^{n-1} \mu_j \\
\Rightarrow & n \cdot \mu_n - (n-1) \cdot \mu_{n-1} = n \cdot (n-1) - (n-1) \cdot (n-2) + 2\mu_{n-1} \\
\Rightarrow & n \cdot \mu_n = (n+1) \cdot \mu_{n-1} + 2(n-1) \\
\Rightarrow & \mu_n = \frac{n+1}{n} \cdot \mu_{n-1} + 2 \cdot \frac{n-1}{n} \\
& = \frac{n+1}{n} \cdot \left(\frac{n}{n-1} \cdot \mu_{n-2} + 2 \cdot \frac{n-2}{n-1} \right) + 2 \cdot n - 1n \\
& = \text{Vollst. Induktion} = 2(n+1) \cdot \sum_{k=2}^n \frac{k-1}{k(k+1)} + \underbrace{1}_{=0} \mu_0
\end{aligned}$$

Umformungen liefern:

$$\mu_n = 2(n+1) \cdot \sum_{j=1}^{n+1} \frac{1}{j} - 4n - 2 \quad \begin{cases} \leq 2n \log n \\ \geq 2n \log n - 4n \end{cases}$$

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

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

Satz 4.8 (Siebformel von Sylvester-Poincaré). Seien $A_1, \dots, A_n \subset \Omega$ Ereignisse in einem diskreten Wahrscheinlichkeitsraum (Ω, P) . Dann gilt:

$$\begin{aligned}
& P\left(\bigcup_{i=1}^n A_i\right) \\
& = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_j \cap A_i) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n)
\end{aligned}$$

Beweis: $1_{A^c} = 1 - 1_A, 1_{A \cap B} = 1_A \cdot 1_B$

$$\begin{aligned}
 1_{\bigcup_{i=1}^n A_i} &= 1 - 1_{\left(\bigcup_{i=1}^n A_i\right)^c} = 1 - 1_{\bigcap_{i=1}^n A_i^c} = 1 - \prod_{i=1}^n 1_{A_i^c} = 1 - \prod_{i=1}^n (1 - 1_{A_i}) \\
 &= 1 - \left(1 - \sum_{i=1}^n 1_{A_i} + \sum_{1 \leq i < j \leq n} 1_{A_i} \cdot 1_{A_j} - \dots + (-1)^{n-1} \prod_{i=1}^n 1_{A_i}\right) \\
 &= \sum_{i=1}^n 1_{A_i} - \sum_{1 \leq i < j \leq n} 1_{A_i \cap A_j} + \dots + (-1)^{n-1} \prod_{i=1}^n 1_{A_i} \\
 \Rightarrow P\left(\bigcup_{i=1}^n A_i\right) &= E\left(1_{\bigcup_{i=1}^n A_i}\right) = \sum_{i=1}^n \underbrace{E(1_{A_i})}_{P(A_i)} - \sum_{1 \leq i < j \leq n} \underbrace{E(1_{A_i \cap A_j})}_{P(A_i \cap A_j)} + \dots + (-1)^{n-1} \underbrace{E\left(1_{\bigcup_{i=1}^n A_i}\right)}_{P\left(\bigcap_{i=1}^n A_i\right)}
 \end{aligned}$$

Definition 4.9 (Varianz, Standardabweichung & Moment). Sei X eine \mathbb{R} -wertige Zufallsvariable auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) .

(i) Existiert $E(X)$, so wird

$$\text{Var}(X) := E[(X - E(X))^2] = \sum_{x \in X} (x - E(X))^2 \cdot P\{X = x\}$$

die Varianz von X genannt und $\sqrt{\text{Var}(X)}$ heißt Standardabweichung von X

(ii) Für $k \in \mathbb{N}$ heißt $E(X^k)$ (im Falle der Existenz) das k -te Moment von X

Einschub:

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

$$X^k : \Omega \rightarrow \mathbb{R}$$

$$\omega \rightarrow (X(\omega))^k$$

Bemerkung 4.10. Die Varianz ist die mittlere quadratische Abweichung von X von seinem Erwartungswert, also ein Maß dafür wie stark die Zufälligen Werte um den Erwartungswert streuen.

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

Insbesondere: $0 \leq \text{Var}(X) = E(X^2) - (E(X))^2 \Rightarrow E(X^2) \geq (E(X))^2$

Beispiel 4.11. $P^X = \mathcal{P}_\lambda \Rightarrow E(X) = \lambda \quad (\rightarrow 4.4(ii))$

$$\begin{aligned} E(X \cdot (X - 1)) &= \sum_{n=0}^{\infty} n \cdot (n - 1) \cdot P\{X = n\} \\ &= \sum_{n=2}^{\infty} n \cdot (n - 1) \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} \\ &= \lambda^2 \underbrace{\sum_{n=2}^{\infty} e^{-\lambda} \frac{\lambda^{n-2}}{(n-2)!}}_{=1 \quad (\rightarrow 4.4(ii))} \\ &= \lambda^2 \end{aligned}$$

$$\begin{aligned} \Rightarrow E(X^2) &= E(X \cdot (X - 1)) + E(X) = \lambda^2 + \lambda \\ \Rightarrow \text{Var}(X) &= E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

Satz 4.12 (Markov- & und Chebyshev-Ungleichung). X sei eine Zufallsvariable mit Erwartungswert $E(X)$

$$(i) P\{|X| \geq c\} \leq \frac{E(|X|)}{c} \quad \forall c > 0 \quad \underline{\text{Markov-Ungleichung}}$$

$$(ii) P\{|X - E(X)| \geq c\} \leq \frac{\text{Var}(X)}{c^2} \quad \forall c > 0 \quad \underline{\text{Chebyshev-Ungleichung}}$$

Beweis:

$$(i) P\{|X| > c\} \stackrel{4.4(i)}{=} E(1_{\{|X| \geq c\}}) \leq E\left(\frac{|X|}{c} \cdot 1_{\{|X| \geq c\}}\right) \leq E\left(\frac{|X|}{c}\right) = \frac{E(|X|)}{c}$$

(ii) in den Übungen

Beispiel 4.13.

(i) Sei $X \mathcal{P}_\lambda$ -verteilt, $\lambda > 0 \rightarrow E(X) = \lambda = \text{Var}(X)$

$$\Rightarrow P\{X \geq c\} \leq \frac{E(X)}{c} = \frac{\lambda}{c} \quad \forall c > 0 \quad (\text{Markov-Ungleichung})$$

und für $c > \lambda$

$$P\{X \geq c\} = P\{X - \lambda \geq c - \lambda\} \leq P\{|X - \lambda| \geq c - \lambda\} \leq \frac{\text{Var}(X)}{(c - \lambda)^2} = \frac{\lambda}{(c - \lambda)^2}$$

Für $c < \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}$ ist die Markov-Ungleichung schärfer

und für $c > \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}$ ist die Chebyshev-Ungleichung schärfer

Oft sind beide Schranken viel größer als $P\{X > c\}$

(ii) Quicksort

$Z(x_1, \dots, x_n)$: Anzahl der Vergleiche, um die Zahlen x_1, \dots, x_n zu sortieren.

$$E(Z(x_1, \dots, x_n)) \begin{cases} \leq 2n \log n \\ \geq 2n \log n - 4n \end{cases}$$

Man kann zeigen: $\text{Var}(Z(x_1, \dots, x_n)) \leq 3n(n-1)$

$$\begin{aligned} &\Rightarrow P\{Z(x_1, \dots, x_n) \geq 2n \log n + a \cdot n\} \\ &\leq P\{|Z(x_1, \dots, x_n) - E(Z(x_1, \dots, x_n))| \geq \underbrace{2n \log n - E(Z(x_1, \dots, x_n)) + a \cdot n}_{\geq 0}\} \\ &\leq P\{|Z(x_1, \dots, x_n) - E(Z(x_1, \dots, x_n))| \geq a \cdot n\} \\ &\leq \frac{\text{Var}(Z(x_1, \dots, x_n))}{(a \cdot n)^2} \leq \frac{3}{a^2} \end{aligned}$$

Insbesondere: $a = \epsilon \cdot \log n$

$$\Rightarrow P\{Z(x_1, \dots, x_n) \geq (2 + \epsilon)n \log n\} \leq \frac{3}{\epsilon^2 (\log n)^2} \xrightarrow[n \rightarrow \infty]{} 0 \quad \forall \epsilon > 0$$

Also ist mit großer Wahrscheinlichkeit die Anzahl der Vergleiche von der Größenordnung $2n \log n$, wenn der Datenumfang n groß ist.

Bemerkung 4.14. Die Varianz ist im gegensatz zum Erwartungswert nicht linear im Argument.

$$\begin{aligned} \text{Var}(aX + b) &= E((aX + b - \underbrace{E(aX + b)}_{=a \cdot E(X) + b})^2) \\ &= E(a^2 \cdot (X - E(X))^2) \\ &= a^2 \cdot E((X - E(X))^2) \\ &= a^2 \cdot \text{Var}(X) \quad \forall a, b, \in \mathbb{R} \end{aligned}$$

$$\begin{aligned} \text{Var}(X + Y) &= E((X + Y - \underbrace{E(X + Y)}_{=E(X) + E(Y)})^2) \\ &= E(((X - E(X)) + (Y - E(Y)))^2) \\ &= E((X - E(X))^2 + 2(X - E(X))(Y - E(Y)) + (Y - E(Y))^2) \\ &= \text{Var}(X) + 2E((X - E(X)) \cdot (Y - E(Y))) + \text{Var}(Y) \end{aligned}$$

Definition 4.15 (Kovarianz & Korrelation).

Sind X, Y Zufallsvariablen mit Erwartungswerten $E(X)$ bzw. $E(Y)$.

$$\text{Cov}(X, Y) := E((X - E(X)) \cdot (Y - E(Y))) = E(XY) - E(X) \cdot E(Y)$$

heißt (im Falle der Existenz) die Kovarianz von X und Y .

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}Y}}$$

heißt dann Korrelation von X und Y , falls $\text{Var}(X) > 0, \text{Var}(Y) > 0$.
 X und Y heißen unkorreliert, falls $\text{Cov}(X, Y) = 0$

Satz 4.16. Sind X und Y stochastisch unabhängige Zufallsvariablen mit Erwartungswerten $E(X)$ bzw $E(Y)$, so gilt

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

, d. h. X und Y sind unkorreliert.

Beweis:

$$\begin{aligned} E(X \cdot Y) &= \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} x \cdot y \cdot P\{X = x, Y = y\} \\ &= \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} x \cdot y \cdot P\{X = x\} \cdot P\{Y = y\} \quad (\text{da stochastisch Unabhängig}) \\ &= \sum_{x \in X(\Omega)} x \cdot P\{X = x\} \cdot \sum_{y \in Y(\Omega)} y \cdot P\{Y = y\} = E(X) \cdot E(Y) \end{aligned}$$

Korollar 4.17. Sind X_1, \dots, X_n Zufallsvariablen mit Erwartungswerten $E(X_1), \dots, E(X_n)$. Dann gilt:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \cdot \sum_{1 \leq i < j \leq n} \text{Corr}(X_i, X_j) \quad (4.1)$$

Sind die Zufallsvariablen X_1, \dots, X_n unabhängig (oder schwächer: unkorreliert), so gilt insbesondere

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

Beweis:

4.1 folgt aus Bemerkung 4.14, der Rest mit Definition 4.15 und Satz 4.16

Beispiel 4.18. X sei eine $\mathcal{B}_{(n,p)}$ -verteilte Zufallsvariable.

Bsp. 3.5: X hat die selbe Verteilung wie $\sum_{i=1}^n X_i$, X_i sind stochastisch unabhängig und X_i ist $\mathcal{B}_{(n,p)}$ -verteilt, d.h. $P\{X_i = 1\} = p, P\{X_i = 0\} = (1 - p)$

$$\begin{aligned} \text{Var}(X) &= \text{Var}\left(\sum_{i=1}^n X_i\right) \stackrel{(4.1)}{=} \sum_{i=1}^n \text{Var}(X_i) \\ &= n \cdot \text{Var}(X_1) = n \cdot (E(\underbrace{X_1^2}_{=X_1}) - (E(X_1))^2) \\ &= n \cdot (p - p^2) = n \cdot p \cdot (1 - p) \end{aligned}$$

Bemerkung 4.19. Der Wert einer Zufallsvariable X wurde beobachtet. Daraus soll der Wert der Zufallsvariable Y vorhergesagt werden.

Dabei sollen zur Approximation von Y nur lineare Funktionen von X verwendet werden. Als Maß für die Approximationsgüte/Vorhersagegüte verwenden wir

$E((Y - (aX + b))^2)$: mittlerer quadratischer Vorhersagefehler

Wähle a, b , so, dass $E((Y - (aX + b))^2)$ minimal wird.

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

$E((Y - (aX + b))^2)$ wird als Funktion von b minimiert durch

$$b = E(Y) - aE(X)$$

$$\begin{aligned} \Rightarrow E((Y - (aX + b))^2) &= \text{Var}(Y) + a^2 \cdot \text{Var}(X) + 2 \cdot \underbrace{\text{Cov}(Y, -aX)}_{=-a\text{Cov}(X,Y)} \\ &= \left(a \cdot \sqrt{\text{Var}(X)} - \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}}\right)^2 + \text{Var}(Y) - \frac{(\text{Cov}(X, Y))^2}{\text{Var}(X)} \end{aligned}$$

\Rightarrow minimalstelle $a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$, falls $\text{Var}(X) > 0$

\Rightarrow beste lineare Approximation für Y ist $\frac{\text{Cov}(X, Y)}{\text{Var}(X)} \cdot (X - E(X)) + E(Y)$

\Rightarrow beste lineare Approximation für $\tilde{Y} = \frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}}$ ist $\underbrace{\frac{\text{Corr}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}}_{\text{Corr}(X, Y)} \cdot \underbrace{\frac{X - E(X)}{\sqrt{\text{Var}(X)}}}_{=: \tilde{X}}$

Beobachte: $E(\tilde{Y}) = 0, \text{Var}(\tilde{Y}) = 1$

Der zugehörige mittlere quadratische Approximationsfehler ist

$$\begin{aligned} E\left((\tilde{Y} - \text{Corr}(X, Y) \cdot \tilde{X})^2\right) &= \frac{1}{\text{Var}(Y)} \cdot E\left((Y - (aX + b))^2\right) \\ &\stackrel{\text{s.o.}}{=} \frac{1}{\text{Var}(Y)} \cdot \left(\text{Var}\left(Y - \frac{\text{Cov}(X, Y)}{\text{Var}(X)}\right)\right) \\ &= 1 - (\text{Corr}(X, Y))^2 \quad (\Rightarrow \text{Corr}(X, Y) \in [-1, 1]) \end{aligned}$$

$|\text{Corr}(X, Y)| = 1 \Rightarrow$ kein Vorhersagefehler

$\text{Corr}(X, Y) = 0 \Rightarrow$ maximaler Vorhersagefehler

Korrelation ist nur Maß für Stärke des linearen Zusammenhangs zwischen X und Y .
Es ist möglich, dass X, Y unkorreliert sind, aber Y vollständig durch X bestimmt ist.

(Blödes) Beispiel:

Sei $P\{X = 1\} = P\{X = -1\} = \frac{1}{2}$

$Y = X^2 \Rightarrow X \cdot Y = X^3 = X$

$\Rightarrow E(X \cdot Y) = E(X) = E(X) \cdot E(Y)$, d.h. unkorreliert.

(Blödes Beispiel, da Y konstant ist.)

(Besseres) Beispiel:

Sei $P\{X = 1\} = P\{X = 0\} = P\{X = -1\} = \frac{1}{3}$

$Y = X^2$

$E(X) = \sum_{x=-1}^1 x \cdot P\{X = x\} = 0$

$X \cdot Y = X^3 = X \Rightarrow E(X \cdot Y) = 0 = E(X) \cdot E(Y)$ d.h. unkorreliert, aber X, Y sind nicht unabhängig, sondern Y ist durch X vollständig bestimmt

5 Grenzwertsätze

Man betrachtet $\sum_{i=1}^n X_i$ von Zufallsvariablen X_i für $n \rightarrow \infty$

Für große n ist die Verteilung von $\sum_{i=1}^n X_i$ in der Regel nicht exakt berechenbar.

Ziel: Eine gute Approximation für große n zu finden.

Satz 5.1 (Schwaches Gesetz der Großen Zahlen). Seien $X_i, i \in \mathbb{N}$ unkorrelierte Zufallsvariablen mit Erwartungswert $E(X_i)$ und $\text{Var}(X_i) \leq M \quad \forall i \in \mathbb{N}$ und ein $M < \infty$. Dann gilt für alle $\epsilon > 0$

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

Beweis: $E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$, $\text{Var}\left(\sum_{i=1}^n X_i\right) \stackrel{\text{(unkorreliert)}}{=} \sum_{i=1}^n \text{Var}(X_i)$

$$\begin{aligned} P\left\{\frac{1}{n} \left| \sum_{i=1}^n (X_i - E(X_i)) \right| \geq c\right\} &= P\left\{\left| \sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right) \right| \geq n \cdot c\right\} \\ &\stackrel{\text{Chebyshev}}{\leq} \frac{\text{Var}\left(\sum_{i=1}^n X_i\right)}{(n \cdot c)^2} \leq \frac{n \cdot M}{n \cdot c^2} \xrightarrow{n \rightarrow \infty} 0 \Rightarrow \text{Beh.} \end{aligned}$$

Definition 5.2. Seien Y, Y_n \mathbb{R} -wertige Zufallsvariablen. Gilt für alle $\epsilon > 0$

$$P\{|Y_n - Y| \geq \epsilon\} \xrightarrow{n \rightarrow \infty} 0,$$

so sagt man, dass Y_n (P-)Stochastisch gegen Y konvergiert (in Zeichen $Y_n \xrightarrow{P} Y$ oder $Y_n \rightarrow Y$ P-Stochastisch)

Bemerkung 5.3. Unter der Bedingung von Satz 5.1 gilt also

$$\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \xrightarrow{P} 0$$

Ist $E(X_i) = \mu$ für alle $i \in \mathbb{N}$, so gilt auch

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$$

Beispiel 5.4. X_i seien unabhängig und $\mathcal{B}_{(1,p)}$ -verteilt.

Dann ist $X_n := \sum_{i=1}^n X_i \sim \mathcal{B}_{(n,p)}$

$$E(X_i) = p \stackrel{5.1}{\underset{\text{bzw. 5.3}}{\Rightarrow}} \frac{Y_n}{n} \xrightarrow{P} p$$

Möchte man z.B. testen, ob eine Münze fair ist, dann werfe man sie n Mal.

$$X_i \begin{cases} 1 & \text{falls im } i\text{-ten Wurf Kopf fällt} \\ 0 & \text{falls im } i\text{-ten Wurf Zahl fällt} \end{cases}$$

Dann $\underbrace{\frac{1}{n} \sum_{i=1}^n X_i}_{\text{rel. Häufigkeit}} \rightarrow p$

Falls die Münze fair ist, dann ist die relative Häufigkeit $\approx \frac{1}{2}$.

Frage: Bei welchen Abweichungen von $\frac{1}{2}$ kann man dies als deutlichen Hinweis auffassen, dass die Münze unfair ist (also $p \neq \frac{1}{2}$)?

Dazu: $\text{Var}(X_i) = p \cdot (1-p) \leq \frac{1}{4}$

Der Beweis von Satz 5.1 zeigt:

$$P\left\{\frac{1}{n} \left| \sum_{i=1}^n (X_i - p) \right| \geq \epsilon\right\} \leq \frac{\frac{1}{4}}{n \cdot \epsilon^2} = \frac{1}{4n \cdot \epsilon^2}$$

Wähle ϵ so, dass $\frac{1}{4n \cdot \epsilon} = 0,05 \Leftrightarrow \epsilon = \frac{1}{\sqrt{4n \cdot 0,05}}$

Dann $P\left\{\left|\frac{1}{4} \sum_{i=1}^n X_i - p\right| \geq \epsilon\right\} \leq 0,05$

Ist $n = 500$, s ist also bei einer fairen Münze (d.h. $p = \frac{1}{2}$)

$$P\left\{\frac{1}{n} \sum_{i=1}^n X_i \notin (0,4; 0,6)\right\} = P\left\{\left|\frac{1}{4} \sum_{i=1}^n X_i - \frac{1}{2}\right| \geq \epsilon = \frac{1}{4}\right\} \leq 0,05$$

, d.h. falls bei 500 Würfeln die relative Häufigkeit $\leq 0,4$ oder $\geq 0,6$ ist, dann ist die Münze vermutlich unfair.

Das Intervall $(0,4; 0,6)$ ist unnötig lang.

Der Zentrale Grenzsatz liefert genauere Abschätzungen.

Beispiel 5.5 (Monte-Carlo-Situation). Ziel: Berechne für Zufallsvariable Z Wahrscheinlichkeit von Typ $P\{Z \in A\}$ oder Erwartungswert $E(Z)$

Problem: Verteilung P^Z ist oft nicht analytisch bestimmbar

Insbesondere ist oft der Fall, wenn $Z = f(Y_1, \dots, Y_k)$, wobei Y_i "einfach" ist.

Dann: Simuliere "Pseudozufallszahlen" z_i , die sich Wahrscheinlich so verhalten, als würde man unabhängige Zufallsvariablen Z_i beobachten, die alle die selbse Verteilung wie Z haben.

Gesetz der Großen Zahlen:

$$\frac{1}{n} \sum_{i=1}^n Z_i \rightarrow E(Z)$$

Daher ist zu hoffen, dass $\frac{1}{n} \sum_{i=1}^n Z_i \approx E(Z)$, wenn n hinreichend groß ist.

Konkretes Beispiel:

Seien Y_1, Y_2 gleichverteilt auf $\{1, \frac{1}{L}, \frac{2}{L}, \dots, 1\} =: S_L$ ($L \in \mathbb{N}$)

und Y_1, Y_2 seien unabhängig.

$\Rightarrow (Y_1, Y_2)$ gleichverteilt auf S_L^2

Falls L sehr groß ist, so ist (Y_1, Y_2) näherungsweise gleichverteilt auf $[0, 1]^2$

----- Zeichnung eines Viertels des Einheitskreises -----

$Z = 1_{\{X_1^2 + Y_2^2 \leq 1\}}$: Indikator, Y_1, Y_2 im Viertelkreis liegt.

$$E(Z) = P\{Y_1^2 + Y_2^2 \leq 1\} \approx \frac{\text{Fläche Viertelkreis}}{\text{Fläche Einheitsquadrat}} = \frac{\pi}{4}$$
$$\Rightarrow \pi = 4 \cdot E(Z)$$

Bestimme Approximation von π durch Simulation, in dem n Paare von auf S_L gleichverteilten Pseudozufallszahlen $(Y_1^{(i)}, Y_2^{(i)})$ gewählt werden und daraus $z_i := 1_{\{(Y_1^{(i)})^2 + (Y_2^{(i)})^2 \leq 1\}}$ berechnet.

$$4 \cdot \underbrace{\frac{1}{n} \sum_{i=1}^n z_i}_{\approx E(Z)} \approx 4 \cdot E(Z) = \pi$$

Satz & Definition 5.6 (Zentraler Grenzwertsatz). Seien $X_i, i \in \mathbb{N}$ unabhängig und identisch verteilte Zufallsvariablen (iid = independent and identically distributed) (d.h. $P^{X_i} = P^{X_1 \forall i}$) mit Erwartungswert $\mu = E(X_i)$ und Varianz $\sigma^2 = \text{Var}(X_i) \in (0, \infty)$. Dann gilt für alle $x \in \mathbb{R}$

$$P \left\{ \frac{1}{\sqrt{n \cdot \sigma^2}} \sum_{i=1}^n (X_i - \mu) \leq x \right\} \xrightarrow{n \rightarrow \infty} \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

Φ heißt Standardnormalverteilungsfunktion
 $\varphi = \Phi'$ heißt Standardnormalverteilungsdichte

d.h. $\varphi(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}, t \in \mathbb{R}$

Beweis: Ist z.B. im Dübgen zu finden.

Es gilt:

$$\Phi(-\infty) := \lim_{x \rightarrow -\infty} \Phi(x) = 0$$

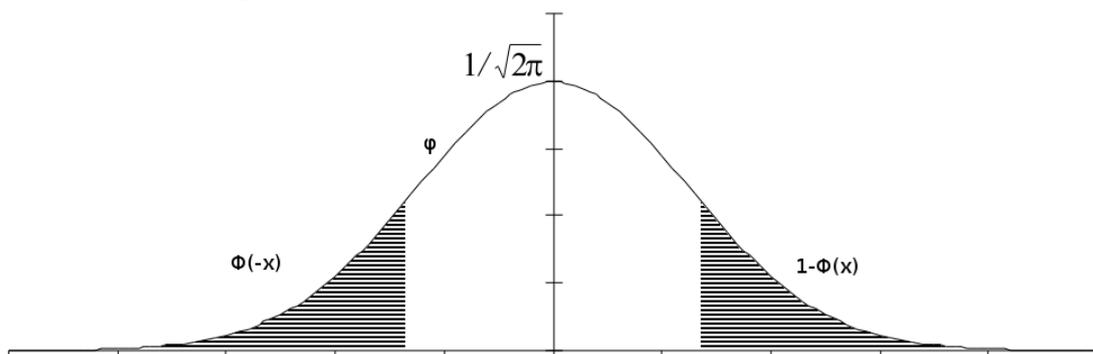
$$\Phi(\infty) := \lim_{x \rightarrow \infty} \Phi(x) = 1$$

$$\varphi(t) = \varphi(-t) \quad \forall t \in \mathbb{R}$$

$$\stackrel{x \geq 0}{\Rightarrow} \Phi(-x) = \int_{-\infty}^{-x} \varphi(t) dt = \int_x^{\infty} \underbrace{\varphi(-t)}_{=\varphi(t)} dt = \underbrace{\int_{-\infty}^{\infty} \varphi(t) dt}_{=\Phi(\infty)=1} - \underbrace{\int_{-\infty}^x \varphi(t) dt}_{=\Phi(x)}$$

d.h. $\Phi(-x) = 1 - \Phi(x)$

Insbesondere $\Phi(0) = \frac{1}{2}$



Außerdem $E(X_i) = p, \text{Var}(X_i) = p \cdot (1 - p)$ (Bsp 4.18)

$$\Rightarrow P \left\{ a < \frac{Y_n - np}{\sqrt{np(1-p)}} \leq b \right\} \xrightarrow{n \rightarrow \infty} \Phi(b) - \Phi(a) \quad \forall -\infty \leq a \leq b \leq \infty$$

Beispiel 5.9. Das Flugzeug eines Fluges kann $n_0 = 200$ Personen befördern.
Mit Wahrscheinlichkeit $p = 0,96$ erscheinen die Ticketinhaber jeweils unabhängig von einander zum Flug.

Frage: Wie viele Tickets darf das Unternehmen verkaufen, damit die Wahrscheinlichkeit einer Überbuchung $\leq 0,05$ ist?

$X =$ Anzahl der Personen, die zum Flug erscheinen

$n =$ Anzahl der verkauften Tickets

$\Rightarrow X$ ist $\mathcal{B}_{(n,p)}$ -verteilt

Wähle n maximal, so dass $P\{X > n_0\} \leq 0,05$

$$P\{X > n_0\} = P \left\{ \frac{X - np}{\sqrt{np(1-p)}} > \frac{n_0 - np}{\sqrt{np(1-p)}} \right\} \approx 1 - \Phi \left(\frac{n_0 - np}{\sqrt{np(1-p)}} \right) \leq 0,05$$

$$b = \infty$$

$$a = \frac{n_0 - np}{\sqrt{np(1-p)}}$$

$$\Leftrightarrow \frac{n_0 - np}{\sqrt{np(1-p)}} \geq \Phi^{-1}(0,95) = Z_{0,95} \approx 1,645$$

$$\Leftrightarrow (n_0 - np)^2 \geq np(1-p)Z_{0,95}^2 \quad n \geq n_0$$

$$\Leftrightarrow \dots \Leftrightarrow n_0 \leq n \leq \frac{n_0 + \frac{1}{2}(1-p)Z_{0,95}^2}{p} - \sqrt{\left(\frac{n_0 + \frac{1}{2}(1-p)Z_{0,95}^2}{p} \right)^2 - \left(\frac{n_0}{p} \right)^2} \approx 203,55$$

Also dürfen maximal 203 Tickets verkauft werden.

Modifikation: $P\{X > n_0\} = P\{X > t\} \quad \forall t \in [n_0, n_0 + 1)$

Genauso wie oben ergibt sich:

$$P\{X > t\} \approx 1 - \Phi \left(\frac{t - np}{\sqrt{np(1-p)}} \right)$$

Oft ist die Approximation besonders gut für den mittleren Wert $t = n_0 + \frac{1}{2}$

Dies nennt man Stetigkeitskorrektur

Bei der Wahr $t = n_0 + \frac{1}{2}$ (statt wie oben n_0) erhält man die Bedingung

$$n \leq 204,06 \rightsquigarrow \text{Der Verkauf von 204 Tickets ist erlaubt}$$

In der Tat ist

$$P\{X > n_0\} = \mathcal{B}_{(n,p)}\{n_0 + 1, \dots, n\} \approx \begin{cases} 0,048 & \text{für } n = 204 \\ 0,094 & \text{für } n = 205 \end{cases}$$

d.h. die Bedingung ist tatsächlich erfüllt für $n = 204$

Beispiel 5.10 (Fortsetzung von Beispiel 5.4). Eine Münze wird $n = 500$ mal geworfen. p ist die Wahrscheinlichkeit für Kopf.

Gesucht: Ein Test, ob $p = \frac{1}{2}$ (d.h. die Münze ist fair)

$$X_i = \begin{cases} 1 & \text{falls im } i\text{-ten Wurf Kopf fällt} \\ 0 & \text{falls im } i\text{-ten Wurf Zahl fällt} \end{cases}$$

$\frac{1}{n} \sum_{i=1}^n X_i =$ relative Häufigkeit, mit der Kopf fällt

Test: Falls $|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2}| \geq \epsilon$, dann wird die Münze als unfair angesehen.

Frage: Wie ist ϵ zu wählen?

Wähle α (z.B. $\alpha = 0,05$) und dann ϵ so, dass die Wahrscheinlichkeit $P\{|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2}| \geq \epsilon\} \leq \alpha$, falls die Münze tatsächlich fair ist.

Mittels des Zentralengrenzwertsatzes:

$$\begin{aligned} P\{|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2}| \geq \epsilon\} &= P\{1n \sum_{i=1}^n X_i - \frac{1}{2} \geq \epsilon\} + P\{1n \sum_{i=1}^n X_i - \frac{1}{2} \leq -\epsilon\} \\ &= 1 - P\{1n \sum_{i=1}^n X_i - \frac{1}{2} < \epsilon\} + P\{1n \sum_{i=1}^n X_i - \frac{1}{2} \leq -\epsilon\} \\ &= 1 - P\left\{ \frac{\sum_{i=1}^n X_i - \frac{n}{2}}{\sqrt{n \frac{1}{2} (1 - \frac{1}{2})}} < \frac{n\epsilon}{\sqrt{n \frac{1}{4}}} \right\} + P\left\{ \frac{\sum_{i=1}^n X_i - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \leq \frac{n\epsilon}{\sqrt{\frac{n}{4}}} \right\} \\ &\approx 1 - \Phi(\sqrt{n} \cdot 2\epsilon) + \Phi(-\sqrt{n} \cdot 2\epsilon) = 2\Phi(-2\epsilon\sqrt{n}) \stackrel{!}{=} \alpha \end{aligned}$$

$$\Rightarrow \epsilon = \frac{\Phi^{-1}\left(\frac{\alpha}{2}\right)}{2\sqrt{n}} \underset{\substack{\alpha=0,05 \\ n=500}}{\approx} 0,0438$$

d.h. die Münze kann als unfair angesehen werden, wenn die relative Häufigkeit mit der Kopf fällt nicht in dem Intervall

$$(0,4562; 0,5438)$$

ist.

Vergleich mit dem Intervall, dass sich aus der Chebyshev'schen Ungleichung ergeben hat (\rightarrow Beispiel 5.4):

$$(0, 4; 0, 6)$$

Bei der Approximation mit dem Zentralengrenzwertsatz werden unfaire Münzen viel eher erkannt.

6 Wahrscheinlichkeitsmaße auf \mathbb{R}

Kontinuierliche Ergebnisse eines Zufallsexperiments (z.B. Längen, gewichte, Zeiten) lassen sich nicht auf natürliche Weise durch diskrete Wahrscheinlichkeitsräume beschreiben.

Im Allgemeinen kann man nicht mehr jeder Untermenge des Grundraumes eine Wahrscheinlichkeit zuweisen.

Wahrscheinlichkeiten erhalten nur “gutartige” Mengen.

- Intervalle sind gutartig
- Komplemente gutartiger Mengen sind gutartig
- abzählbare Vereinigungen gutartiger Mengen sind gutartig

(formale Definition später)

\mathcal{A} : System aller gutartigen Mengen.

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

mit $P(\mathbb{R}) = 1$, $A_n \in \mathcal{A}$ disjunkt, dann $P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$

Frage: Wie kann man Wahrscheinlichkeitsmaße einfach beschreiben?

Es reicht $P(A)$ nur für alle Intervalle anzugeben, genauer $P((a, b])$ für alle $-\infty < a < b < \infty$.

Wegen $P((a, b]) = P((-\infty, b]) - P((-\infty, a])$

denn $(a, b], (-\infty, a]$ sind disjunkt und so $(a, b] \cup (-\infty, a] = (-\infty, b]$

Also reicht es $P((-\infty, x])$ für alle $x \in \mathbb{R}$ festzulegen.

Satz & Definition 6.1 (Verteilungsfunktion). *Ist P ein Wahrscheinlichkeitsmaß auf \mathbb{R} , so gilt für die durch $F : \mathbb{R} \rightarrow [0, 1]$, $F(x) := P((-\infty, x])$ definierte Verteilungsfunktion von P*

(i) F ist monoton Steigend

(ii) F ist rechtsseitig stetig

(iii) $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$, $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$

Ist umgekehrt F eine Funktion von \mathbb{R} nach $[0, 1)$, die die Eigenschaften (i) - (iii) besitzt, so existiert genau ein Wahrscheinlichkeitsmaß auf \mathbb{R} , das F als Verteilungsfunktion besitzt.

Beweis von (i):

$$x \leq y \Rightarrow (-\infty, x] \subset (-\infty, y] \Rightarrow F(x) = P((-\infty, x]) \leq P((-\infty, y]) = F(y)$$

Satz & Definition 6.2 (Dichte). Sei P ein Wahrscheinlichkeitsmaß auf \mathbb{R} . Existiert eine integrierbare Funktion $f : \mathbb{R} \rightarrow [0, \infty)$, so dass

$$P((-\infty, x]) = \int_{-\infty}^x f(t) dt \quad (6.1)$$

so heißt f Dichte von P .

Ist $A \subset \mathbb{R}$ so, dass $f \cdot 1_A$ integrierbar ist, so gilt

$$P(A) = \int f(x) \cdot 1_A(x) dx =: \int_A f(x) dx$$

Umgekehrt ist jede integrierbare Funktion $f : \mathbb{R} \rightarrow [0, \infty)$ mit $\int_{\mathbb{R}} f(t) dt = 1$ Dichte eines Wahrscheinlichkeitsmaßes auf \mathbb{R} , das durch (6.1) eindeutig festgelegt ist.

Beispiel 6.3.

(i) Sei $-\infty < a < b < \infty$ und definiere $f : \frac{1_{(a,b]}}{b-a}$ als eine Dichte eines Wahrscheinlichkeitsmaßes $\mathcal{U}_{(a,b]}$ auf \mathbb{R} , denn $\int_{\mathbb{R}} f(t) dt = \frac{1}{b-a}$ und $\int_{\mathbb{R}} 1_{(a,b]} t dt = \frac{b-a}{b-a} = 1$

$\mathcal{U}_{(a,b]}$ heißt Gleichverteilung auf $(a, b]$

Es gilt $a \leq c < d \leq b$

$$\begin{aligned} \mathcal{U}_{(a,b]}((c, d]) &= \int f(x) \cdot 1_{c,d}(x) dx \\ &= \int \frac{1_{(a,b]}(x)}{b-a} \cdot 1_{(c,d]}(x) dx \\ &= \frac{d-c}{b-a} \end{aligned}$$

d.h. die Wahrscheinlichkeit von $(a, d]$ hängt nur von der Länge des Intervalls ab, nicht von seiner Lage innerhalb von $(a, b]$

(ii) $f_\lambda(x) := \frac{1}{\lambda} \cdot e^{-\frac{x}{\lambda}} \cdot 1_{[0, \infty)}(x)$ ($\lambda > 0$) ist Dichte eines Wahrscheinlichkeitsmaßes auf \mathbb{R} , denn

$$\int_{\mathbb{R}} f_\lambda(x) dx = \frac{1}{\lambda} \int_0^\infty e^{-\frac{x}{\lambda}} dx = -e^{-\frac{x}{\lambda}} \Big|_0^{-\infty} = 1$$

Das zugehörige Wahrscheinlichkeitsmaß E_λ heißt Exponentialverteilung mit Mittelwert $\lambda > 0$. Sie besitzt die Verteilungsfunktion

$$F_\lambda(x) = \int_{-\infty}^x f_\lambda(t) dt \stackrel{\text{s.o.}}{=} -e^{-\frac{t}{\lambda}} \Big|_0^x = 1 - e^{-\frac{x}{\lambda}} \quad \forall x \geq 0$$

und $F_\lambda(x) = 0 \quad \forall x < 0$

Für $x, y \geq 0$ gilt

$$\begin{aligned} E_\lambda((x+y, \infty) \mid (x, \infty)) &= \frac{E_\lambda((x+y, \infty) \cap (x, \infty))}{E_\lambda((x, \infty))} \\ &= \frac{E_\lambda((x+y, \infty))}{E_\lambda((x, \infty))} = \frac{1 - E_\lambda((-\infty, x+y))}{1 - E_\lambda((-\infty, x))} \\ &= \frac{1 - F_\lambda(x+y)}{1 - F_\lambda(x)} = \frac{e^{-\frac{x+y}{\lambda}}}{e^{-\frac{x}{\lambda}}} \\ &= \frac{e^{-\frac{x}{\lambda}} \cdot e^{-\frac{y}{\lambda}}}{e^{-\frac{x}{\lambda}}} = e^{-\frac{y}{\lambda}} \end{aligned}$$

Fazit: Exponentialverteilungen beschreiben Lebensdauern von Dingen, die nicht altern, d.h. die Wahrscheinlichkeit noch y Jahre zu überleben, gegeben dass bereits x Jahre überlebt wurden, hängt nicht von x ab.

(iii) $\varphi(t) := \frac{1}{\sqrt{2\phi}} \cdot e^{-\frac{t^2}{2}}$, $t \in \mathbb{R}$,
ist die Dichte der sogenannten Standardnormalverteilung $\mathcal{N}_{(0,1)}$, denn die zugehörige Verteilungsfunktion $\Phi(x) = \int_{-\infty}^x \varphi(x) dt$ erfüllt $\Phi(\infty) = 1$ (\rightarrow Zentralergrenzwertsatz)

Beispiel 6.4. Ein Seil der Länge 1 wird so lange an beiden Enden gezogen, bis es reißt.

Die Wahrscheinlichkeit, dass das Seil in der Umgebung der Stelle $x \in (0, 1)$ reißt, sei proportional zum Abstand zum näher gelegenen Ende, also proportional zu $\min(x, 1-x)$

Modell: P sei das Wahrscheinlichkeitsmaß mit Dichte

$$f(x) = \begin{cases} c \cdot \min(x, 1-x) & , x \in (0, 1) \\ 0 & , \text{sonst} \end{cases}$$

Wähle $c > 0$ so, dass $\int_{\mathbb{R}} f(x) dx \stackrel{!}{=} 1$

d.h.

$$c \cdot \int_0^1 \min(x, 1-x) dx = 2 \cdot c \int_0^{\frac{1}{2}} x dx = cx^2 \Big|_0^{\frac{1}{2}} = \frac{c}{4} \stackrel{!}{=} 1$$

$\Rightarrow c = 4$, d.h. $f(x) = 4 \cdot \min(x, 1-x) \cdot 1_{(0,1)}(x)$

Die Wahrscheinlichkeit, dass das längere Teilstück $\geq \frac{3}{4}$, ist

$$\begin{aligned} P\left(\left(0, \frac{1}{4}\right) \cup \left(\frac{3}{4}, 1\right)\right) &= \int 1_{(0, \frac{1}{4}) \cup (\frac{3}{4}, 1)}(x) \cdot f(x) dx \\ &= \int_0^{\frac{1}{4}} 4x dx + \int_{\frac{3}{4}}^1 4 \cdot (1-x) dx \stackrel{s.o.}{=} 4x^2 \Big|_0^{\frac{1}{4}} = \frac{1}{4} \end{aligned}$$

Beim diskreten Wahrscheinlichkeitsraum (Ω, P) ist eine Abbildung $X : \Omega \rightarrow S$ eine Zufallsvariable mit der Verteilung von X : $P^X(B) = P(X^{-1}(B)) = P\{X \in B\} \quad \forall B \subset S$. Dies ist dann ein Wahrscheinlichkeitsmaß auf S .

Jetzt ist $P : \underline{\mathcal{A}} \rightarrow [0, 1]$. Daher ist $P(X^{-1}(B))$ nur noch definiert, wenn $X^{-1}(B) \in \mathcal{A}$, d.h. "gutartig" ist.

Definiert man $X : \mathbb{R} \rightarrow \mathbb{R}$ so, dass $X^{-1}(B) \in \mathcal{A} \quad \forall B \in \mathcal{A}$, dann gilt wie oben $P^X(B) := P(\{B \in B\}) \quad \forall B \in \mathcal{A}$ definiert ein Wahrscheinlichkeitsmaß P^X auf \mathbb{R} .

Im diskreten Wahrscheinlichkeitsraum gilt:

$$E(X) = \sum_{x \in X(\Omega)} x \cdot \underbrace{P\{X = x\}}_{P^X(\{x\})}$$

Das ist hier unsinnig:

Wenn z.B. P^X eine Dichte f besitzt, so ist $P^X(\{x\}) = \int_{\mathbb{R}} f(t) \cdot 1_{\{x\}}(t) dt = 0 \quad \forall x \in \mathbb{R}$.

Daher ist ein neuer Ansatz nötig.

Lemma 6.5. *Ist X eine \mathbb{R} -wertige diskrete Zufallsvariable, so gilt*

$$E(X) = \int_0^{\infty} P\{x > t\} dt - \int_{-\infty}^0 P\{x < t\} dt,$$

falls die rechte Seite wohl definiert ist.

Zur Veranschaulichung Skizzen, falls X nur Werte $0 < x_1 < x_2 < \dots < x_n < \infty$ annimmt.

$$p_i = P\{X = x_i\}$$

————— Skizze zur Veranschaulichung —————

Definition 6.6 (Erwartungswert). *Ist X eine \mathbb{R} -wertige Zufallsvariable mit der Verteilungsfunktion F (d.h. P^X hat Verteilungsfunktion F), so wird der Erwartungswert von X definiert als*

$$E(X) = \int_0^{\infty} 1 - F(t) dt - \int_{-\infty}^0 F(t) dt,$$

falls die rechte Seite wohl definiert ist (d.h. wenigstens ein Integral endlich ist).

Satz 6.7. *Besitzt X eine Dichte f , so gilt*

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

falls das Integral wohl definiert ist.

Beispiel 6.8.

(i) *Sei X eine Zufallsvariable mit der Verteilungsfunktion Φ (d.h. $P^X = \mathcal{N}_{(0,1)}$) bzw. mit Dichte φ . Dann gilt*

$$E(X) = \int_{-\infty}^{\infty} x \varphi(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx = 0,$$

da es das Intervall einer ungeraden Funktion ist.

(ii) Ist X exponential verteilt, d.h. $P^X = E_\lambda$ für ein $\lambda > 0$, mit der Verteilungsfunktion $F_\lambda(X) = (1 - e^{-\frac{x}{\lambda}}) \cdot 1_{(0,\infty)}(x)$, so folgt

$$E(X) \stackrel{\text{def.}}{=} \int_0^\infty \cdot e^{-\frac{x}{\lambda}} dx = -\lambda e^{-\frac{x}{\lambda}} \Big|_0^\infty = \lambda$$

Wie sehen Systeme “gutartiger Mengen” aus?

Definition 6.9 (σ -Algebra, Messraum, Ergebnis, Wahrscheinlichkeitsmaß & -raum).
Sei $\Omega \neq \emptyset$ eine beliebiger Grundraum.

(i) Eine Menge $\mathcal{A} \subset 2^\Omega$ von Untermengen von Ω heißt σ -Algebra auf Ω , falls folgende Bedingungen erfüllt sind:

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- $A_n \in \mathcal{A} \quad \forall n \in \mathbb{N} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

(Ω, \mathcal{A}) heißt dann Messraum.

Die Mengen $A \in \mathcal{A}$ heißen dann Ergebnisse.

(ii) Eine Abbildung $P : \mathcal{A} \rightarrow [0, 1]$ heißt Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) , falls

- $P(\Omega) = 1$
- $A_n \in \mathcal{A}, n \in \mathbb{N}, \text{ disjunkt} \Rightarrow P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n) \quad (\sigma\text{-Additivität})$

(Ω, \mathcal{A}, P) heißt dann Wahrscheinlichkeitsraum.

Diskrete Wahrscheinlichkeitsräume sind Spezialfälle mit $\mathcal{A} = 2^\Omega$.

Oft ist σ -Algebra 2^Ω zu groß.

Man kann zeigen: Es gibt kein Wahrscheinlichkeitsmaß P auf $([0, 1], 2^{[0,1]})$ so dass $P((c, d]) = d - c \quad \forall 0 \leq c < d \leq 1$.

D.h. es gibt keine Gleichverteilung auf dem Einheitsintervall, die jeder Untermenge eine Wahrscheinlichkeit zuweist.

Die Sätze und Definitionen 1.4, 1.9, 2.1, 2.3, 2.5, 2.6 übertragen sich sinngemäß auf allemeine Wahrscheinlichkeitsräume, wobei stets nur Mengen aus \mathcal{A} (statt beliebige Untermengen von Ω) betrachtet werden.

Satz & Definition 6.10 (Zufallsvariable, Messbarkeit und Verteilung). Seien $(\Omega, \mathcal{A}), (\tilde{\Omega}, \tilde{\mathcal{A}})$ Messräume und $X : \Omega \rightarrow \tilde{\Omega}$. X heißt $\tilde{\Omega}$ -wertige Zufallsvariable, falls $X^{-1}(\tilde{A}) \in \mathcal{A} \quad \forall \tilde{A} \in \tilde{\mathcal{A}}$.

Man schreibt dann $X : (\Omega, \mathcal{A}) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{A}})$ und sagt, dass X $\mathcal{A}, \tilde{\mathcal{A}}$ -messbar ist.

Sei P ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) . Dann wird durch

$$P^X(\tilde{A}) := P(X^{-1}(\tilde{A})) = P\{X \in \tilde{A}\}, \quad \tilde{A} \in \tilde{\mathcal{A}}$$

ein Wahrscheinlichkeitsmaß P^X auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ definiert; die so genannte Verteilung von X unter P .

Beweis: wie 3.1

Satz & Definition 6.11.

(i) Ist $\Omega \neq \emptyset$ und $\mathcal{C} \subset 2^\Omega$, so ist

$$\sigma(\mathcal{C}) := \bigcup_{\substack{\mathcal{A} \text{ } \sigma\text{-Algebra auf } \Omega \\ \mathcal{C} \in \mathcal{A}}} \mathcal{A}$$

die kleinste σ -Algebra auf Ω , die alle Mengen aus \mathcal{C} enthält. Sie heißt die durch von \mathcal{C} erzeugte σ -Algebra.

(ii) Speziell für $\Omega = \mathbb{R}$ heißt

$$\mathbb{B} := \sigma(\{(a, b] \mid -\infty < a < b < \infty\})$$

Borel- σ -Algebra auf \mathbb{R} und allgemeiner für Intervalle I

$$\mathcal{B}(I) := \sigma(\{(a, b] \mid a < b, \quad a, b \in I\})$$

die Borel- σ -Algebra auf I .

\mathbb{B} enthält alle “üblichen” Untermengen von \mathbb{R} , aber man kann zeigen, dass $\mathbb{B} \neq 2^\mathbb{R}$. Im folgenden sollen alle (nicht-diskrete) Wahrscheinlichkeitsmaße auf \mathbb{R} stets auf (\mathbb{R}, \mathbb{B}) definiert sein.

Jedes Wahrscheinlichkeitsmaß P auf (\mathbb{R}, \mathbb{B}) ist durch seine Verteilungsfunktion

$$F(x) = P((-\infty, x]), \quad x \in \mathbb{R}$$

eindeutig bestimmt (\rightarrow Satz 6.1).

Lemma 6.12. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

(i) Sei $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ eine monoton steigende Folge (d.h. $A_n \subset A_{n+1} \quad \forall n \in \mathbb{N}$).
Dann gilt für $A := \bigcup_{n \in \mathbb{N}} A_n$ (kurz $A_n \uparrow A$)

$$P(A) = \lim_{n \rightarrow \infty} P(A_n)$$

(ii) Ist $B_n \in \mathcal{A}$, $n \in \mathbb{N}$ eine monoton fallende Folge (d.h. $B_n \supset B_{n+1} \quad \forall n \in \mathbb{N}$) und
 $B := \bigcap_{n \in \mathbb{N}} B_n$ (kurz $B_n \downarrow B$), so gilt

$$P(B) = \lim_{n \rightarrow \infty} P(B_n)$$

Beweis:

(i) Definiere $C_1 := A_1$, $C_n := A_n \setminus A_{n-1} \quad \forall n \geq 2$

$\Rightarrow C_n, n \in \mathbb{N}$, disjunkt, $\bigcup_{i=1}^n C_i = A_n$ und insbesondere $\bigcup_{i \in \mathbb{N}} C_i = A$

$$\Rightarrow P(A) = \sum_{i \in \mathbb{N}} P(C_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(C_i) = \lim_{i \rightarrow \infty} P(A_n)$$

(ii) wende (i) auf $A_n := B_n^c \uparrow \bigcup_{k \in \mathbb{N}} B_k^c = \left(\bigcap_{k \in \mathbb{N}} B_k \right)^c = B^c =: A$ an.

$$P(B) = 1 - P(A) \stackrel{(i)}{=} 1 - \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} (1 - P(A_n)) = \lim_{n \rightarrow \infty} P(B_n)$$

Beweis von 6.1 (ii) & (iii):

(ii) Beh.: F ist rechtsseitig stetig

Beweis: Sei $X_n \downarrow X$. Dann $(-\infty, X_n] \downarrow (-\infty, X]$. Mit 6.12(ii):

$$F(X_n) = P((-\infty, X_n]) \rightarrow P((-\infty, X]) = F(X)$$

(iii) Beh.: $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$, $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$

Beweis: Sei $X_n \uparrow \infty$. Dann $(-\infty, X_n] \uparrow \mathbb{R}$ und 6.12(i) liefert

$$F(\infty) = \lim_{n \rightarrow \infty} P((-\infty, X_n]) = P(\mathbb{R}) = 1$$

Die zweite Aussage folgt analog.

Analog für $\Omega = \mathbb{R}^n$

Satz & Definition 6.13.

(i) Die Borel- σ -Algebra auf \mathbb{R}^n wird definiert als

$$\mathbb{B}^n = \sigma(\{ \prod_{i=1}^n (a_i, b_i] \mid -\infty < a_i < b_i < \infty \})$$

(ii) Ist P ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathbb{B}^n)$, so wird die zugehörige multivariate Verteilungsfunktion F definiert als

$$F(x_1, \dots, x_n) := P(\prod_{i=1}^n (-\infty, x_i]), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

(iii) Existiert eine Abbildung $f : \mathbb{R}^n \rightarrow [0, \infty)$, so dass

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_n \cdots dt_2 dt_1 \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

gilt, so heißt f (multivariate) Dichte von P bzw. von F .
Es gilt dann für alle Mengen $B \in \mathbb{B}^n$

$$P(B) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} 1_B(t_1, \dots, t_n) \cdot f(t_1, \dots, t_n) dt_n \cdots dt_1$$

(iv) Ist (Ω, \mathcal{A}) ein Messraum und $X_i : \Omega \rightarrow \mathbb{R}$, $1 \leq i \leq n$, so ist $X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ $\mathcal{A}, \mathbb{B}^n$ -messbar genau dann, wenn X_i \mathcal{A}, \mathbb{B} -messbar ist $\forall 1 \leq i \leq n$.

Die multivariate Verteilungsfunktion von P^X , d.h.

$$F(x_1, \dots, x_n) := P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\} \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

heißt gemeinsame Verteilungsfunktion von X_1, \dots, X_n

Gegenstück zu 3.4

Satz 6.14. Für Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathbb{B})$, $1 \leq i \leq n$, sind äquivalent:

(i) X_1, \dots, X_n sind stochastisch unabhängig (d.h. $\{X_i \in B_i\}$, $1 \leq i \leq n$ sind stochastisch unabhängig $\forall B_i \in \mathbb{B}$)

(ii) $\forall B_i \in \mathbb{B}$, $1 \leq i \leq n$, gilt $P\{X_i \in B_i \forall 1 \leq i \leq n\} = \prod_{i=1}^n P\{X_i \in B_i\}$

(iii) Es gibt Funktionen $H_i : \mathbb{R} \rightarrow [0, \infty)$, $1 \leq i \leq n$, so dass für die gemeinsame Verteilungsfunktion von X_1, \dots, X_n gilt

$$F(x_1, \dots, x_n) = \prod_{i=1}^n H_i(x_i) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

In dem Fall kann man in (iii) H_i gleich der Verteilungsfunktion von X_i wählen.

Besitzen die so genannten Randverteilungen P^{X_i} , $1 \leq i \leq n$, von $P^{(X_1, \dots, X_n)}$ jeweils eine Dichte f_i , so ist (i)-(iii) äquivalent zu

(iv) $P^{(X_1, \dots, X_n)}$ besitzt eine Dichte f der Form

$$f(x_1, \dots, x_n) := \prod_{i=1}^n h_i(x_i) \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

für geeignete Funktionen $h_i : \mathbb{R} \rightarrow [0, \infty)$.

In dem Fall kann man $h_i = f_i$ wählen