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*
$\Omega = \{1, ..., 6\}$
Ereignis "Augenzahl ist gerade": $A = \{2, 4, 6\}$

(ii) *In einem Netzwerk werden Längen (in Byte) der ersten* $n = 10^5$ *Datenpakete beobachtet*
$\Omega = \mathbb{N}^n = \{(\omega_1, ..., \omega_n) | \omega_i \in \mathbb{N} \, \forall \, 1 \leq i \leq n\}$
Interpretation: $\omega_i =$ *Länge des* i*-ten Paketes*
Ereignis "Das größte Paket umfasst maximal 10^7 *Byte"*
$A := \{(\omega_1, ..., \omega_n) \mid \omega_i \leq 10^7 \forall 1 \leq i \leq n\}$

"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 <u>oder</u> B eintritt
- $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ und } \omega \in B\}$
  Ereignis, dass A <u>und</u> 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	$P : \mathcal{A} \to \mathbb{R}$	
	mit Normierung	$P(\emptyset) = 0$	Ereignis, das nie eintritt, hat W. 0
		$P(\Omega) = 1$	Ereignis, das immer eintritt, hat W. 1
		$\rightsquigarrow P : \mathcal{A} \to [0,1]$	

Welche Abb. P sind sinnvoll?

$\to$ 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 disjunkt (d.h. $A \cap B = \emptyset$) $\Rightarrow P_n(A \cup B) = P_n(A) + P_n(B)$

Induktiv folgt: Ereignisse $A_1, ..., A_k$ disjunkt $\Rightarrow P_n(\bigcup\limits_{i=1}^{k} A_i) = \sum\limits_{i=1}^{k} P_n(A_i)$

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 $P : 2^{\Omega} \to [0,1]$ 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*
$\Omega = \{1, ..., 6\}$
$A \subset \Omega : P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6}$
insbesondere $P(\{\omega\}) = \frac{1}{6}$

(ii) *Werfen zweier ununterscheidbarer fairer Würfel*
$\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, ..., 6\}, \omega_1 \leq \omega_2\}$
Hier ist es <u>nicht</u> 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*

$P(A) = \sum\limits_{\omega \in A} P(\{\omega\})$

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

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

(ii) $A_1, ..., A_k \subset \Omega$ *disjunkt* $\Rightarrow P(\bigcup\limits_{i=1}^{k} A_i) = \sum\limits_{i=1}^{k} P(A_i)$

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

(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) $\forall A_1, ..., A_k \subset \Omega \Rightarrow P(\bigcup\limits_{i=1}^{k} A_i) \leq \sum\limits_{i=1}^{k} P(A_i)$ *(Subadditivität)*

Beweis: (i) - (v) übungen

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

$\Rightarrow B_1, ..., B_2$ disjunkt (denn für $i < l$ gilt $B_i \subset A_i, A_i \cap B_l = \emptyset$)

$\bigcup\limits_{i=1}^{k} B_i = \bigcup\limits_{i=1}^{k} A_i$

$\Rightarrow P(\bigcup\limits_{i=1}^{k} A_i) = P(\bigcup\limits_{i=1}^{k} B_i) \overset{(ii)}{=} \sum\limits_{i=1}^{k} P(B_i) \leq \sum\limits_{i=1}^{k} P(A_i)$

(Ist $\omega \in A_i$ für ein i und ist i_ω das <u>kleinste</u> 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 $1, ..., n$ 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 $\Omega_a = \{1, ..., n\}^k = \{(\omega_1, ..., \omega_k) \mid \omega_i \in \{1, ..., n\}\}$
Interpretation: $\omega = (\omega_1, ..., \omega_k) \in \Omega_a$
$\widehat{=}$ im i-ten Zug werde Kugel ω_i gezogen
$|\Omega_a| = n^k$

1.6 (b) **Ziehen ohne Zurücklegen, aber mit Berücksichtigung der Reihenfolge**
$\Omega_b = \{(\omega_1, ..., \omega_n) \in \Omega_a \mid \omega_i \neq \omega_j \forall 1 \leq i < j \leq k\}$
Interpretation wie bei (a), $|\Omega_b| = n \cdot (n-1) \cdot ... \cdot (n-k+1) = \frac{n!}{(n-k)!}$

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 $\omega_1, ..., \omega_{k-1}$

1.6 (c) **Ziehen <u>ohne</u> Zurücklegen und <u>ohne</u> Berücksichtigung der Reihenfolge**
Ordne gezogene Kugelnummern nach Größe

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

Es gibt zu jedem $(\omega_1, ..., \omega_k) \in \Omega$ 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 <u>mit</u> Zurücklegen und <u>ohne</u> Berücksichtigung der Reihenfolge**
$\Omega_d = \{(\omega_1, ..., \omega_k) \in \Omega_a \mid \omega_1 \leq \omega_2 \leq ... \leq \omega_k\}$
Die Abbildung

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

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

ist eine Bijektion.

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

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

Beispiel 1.7. *(i) Lotto 6 aus 49*
6 Kugeln werden ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge aus 49 Kugeln gezogen

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

Alle Möglichkeiten sind gleich wahrscheinlich

$\Rightarrow$ *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 $r \in \{1, \ldots, 10\}$ *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*

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

Ereignis "Erfolg nach genau k Versuchen"

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

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

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

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

Bei 2 richtigen Links erhält man auf ähnliche Weise
$|A| = \frac{8!2}{(8-(k-1))!}$ *(da es nun 2 Möglichkeiten für den k-ten Versuch gibt)*

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

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

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

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

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

Satz & Definition 1.10.

(i) *Sei* (Ω, P) *ein diskreter Wahrscheinlichkeitsraum.*
Dann heißt die Funktion $f : \Omega \to [0,1], f(\omega) = P(\{\omega\})$ *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:*

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

(ii) *Ist umgekehrt* $\Omega \neq \emptyset$ *und* $f : \Omega \to [0,1]$ *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, $P(\Omega_0) = 1$
$\Rightarrow P({\Omega_0}^c) = 0$
$\Rightarrow \forall \omega \in {\Omega_0}^c : f(\omega) = P(\{\omega\}) \leq P({\Omega_0}^c) = 0$
$\Rightarrow \Omega_T \subset \Omega_0 \Rightarrow \Omega_T$ abzählbar, d.h. (a).
(b) ist Spezialfall von (1.1) mit $A = \Omega$
(1.1) folgt aus der σ-Additivität von P, denn

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

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

- $P(\Omega) \overset{(1.1)}{=} \sum_{\omega \in \Omega} f(\omega) \overset{(b)}{=} 1$
- $A_n \subset \Omega, n \in \mathbb{N}$, disjunkt
  $\Rightarrow P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{\omega \in \bigcup_{n \in \mathbb{N}} A_n} f(\omega) = \sum_{n \in \mathbb{N}} \sum_{\omega \in A_n} f(\omega) \overset{(1.1)}{=} \sum_{n \in \mathbb{N}} P(A_n)$

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

Also ist P ein diskretes Wahrscheinlichkeitsmaß mit $P(\{\omega\}) \overset{(1.1)}{=} f(\omega)$ 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* $p \in [0,1]$

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

Dies definiert eine Funktion $f : \{0, ..., n\} =: \Omega \to [0,1]$, *die 1.10(a) und 1.10(b) erfüllt, also die Zähldichte eines diskreten Wahrscheinlichkeitsmaßes auf* Ω *ist.*
Dieses Wahrscheinlichkeitsmaß heißt Binomialverteilung $\mathcal{B}_{(n,p)}$
d.h. $\mathcal{B}_{(n,p)}(\{k\}) = \binom{n}{p} p^k \cdot (1-p)^{n-k} \quad \forall k \in \{0, ..., n\}$

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, <u>bevor</u> das erste Mal 6 gewürfelt wird?*

Bei fairem Würfel:
Wahrscheinlichkeit, k-mal <u>keine</u> 6 zu würfeln $= \left(\frac{5}{6}\right)^k$
Wahrscheinlichkeit, beim $(k+1)$*-ten Mal 6 zu würfeln* $= \left(\frac{1}{6}\right)$

$\Rightarrow gesuchteWahrscheinlichkeit = \left(\frac{5}{6}\right)^k \cdot \frac{1}{6} \qquad \forall k \in \mathbb{N}_0$

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*

$\sum\limits_{k \in \mathbb{N}_0} f(k) = p \cdot \sum\limits_{n \in \mathbb{N}_0} (1-p)^k \overset{geometrischeReihe}{=} p \cdot \frac{1}{1-(1-p)} = 1$

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 Gesetzt 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*
$P(A \mid B) = \frac{P(A\cap B)}{P(B)}$
die bedingte Wahrscheinlichkeit von A gegeben B

Beispiel 2.2. *(vergleiche Präsenzübungsblatt 2, A3)*
Beim Skatspiel hat Spieler 1 keinen Buben erhalten.
Wie groß ist dann die Wahrscheinlichkeit, dass Spieler 2 oder Spieler 3 alle 4 Buben erhalten hat?
Modell: $\Omega = \{(\underbrace{\omega_1, ..., \omega_{10}}_{Karten\,Sp.1}, \underbrace{\omega_{11}, ..., \omega_{20}}_{Karten\ Sp.2}, \underbrace{\omega_{21}, ..., \omega_{30}}_{KartenSp.3}, \underbrace{\omega_{31}, \omega_{32}}_{Stock},) \in K^{32} \mid \omega_i \neq \omega_j \ \forall i \neq j\}$
mit $K = \{KreuzAs, PikAs, ..., Karo7\}$ *Menge aller Karten*
P *Laplace-Verteilung*
$C = \{\omega \in \Omega \mid \{\omega_1, ..., \omega_{10}\} \cap \{KreuzBube, PikBube, HerzBube, KaroBube\} = \emptyset\}$
Ergebnis, dass Spieler 1 keinen Buben erhält.
$B_2 = \{\omega \in \Omega \mid \{KreuzBube, PikBube, HerzBube, KaroBube\} \subset \{\omega_{11}, ..., \omega_{20}\}\}$
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.*

$|B_2|$ *&* $|B_3|$ *in Präsenzaufgabe.*

Gesuchte Wahrscheinlichkeit:

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

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*

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

(ii) *Satz von Bayes*

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

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

Beweis:

(i) $\sum_{i \in I} P(A \mid B_i) \cdot P(B_i) = \sum_{i \in I} \frac{P(A \cap B)}{P(B_i)} \cdot P(B_i) = P(\bigcup_{i \in I} (A \cap B_i)) = P(A)$,
da $A \cap B_i$ disjunkt.

(ii) $P(B_k \mid A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A \mid B_k) \cdot P(B_k)}{P(A)} \Rightarrow$ 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
$P(T \mid D) = 0,99$ *Sensitivität des Tests*
$P(T^c \mid D^c) = 0,98$ *Spezifität des Tests*
$\Rightarrow P(T \mid D^c) = 0,02$

Was bedeutet es, wenn der Test positiv ist?

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

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

25 jährige Schwangere: $P(D) \approx \frac{1}{1250}$ $\Rightarrow P(D \mid T) \approx 0,035$
43 jährige Schwangere: $P(D) \approx \frac{1}{50}$ $\Rightarrow P(D \mid T) \approx 0,503$

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 $P(A \mid B) = P(A)$ 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 $A_1, ..., A_n \subset \Omega$ in einem diskreten Wahrscheinlichkeitsraum (Ω, P) heißen (P-)stochastisch unabhängig, wenn für jede Indexmenge $I \subset \{1, ..., n\}, I \neq \emptyset$, gilt*
$P(\bigcap\limits_{i \in I} A_i) = \prod\limits_{i \in I} P(A_i)$

Bemerkung 2.7.

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

(ii) *Mehr als 2 Ereignisse* $A_1, ..., A_n$ *sind im allgemeinen* <u>*nicht*</u> *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* <u>*nicht*</u> *stochastisch unabhängig*

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

Bsp.: 2 maliges Werfen eines fairen Würfels
$\Omega = \{1, ..., 6\}$*, P Laplace-Verteilung*
$A_1 = \{1, 3, 5\} \times \{1, ..., 6\}$ *(Erste Augenzahl ist ungerade)*
$A_2 = \{1, ..., 6\} \times \{1, 3, 5\}$ *(Zweite Augenzahl ist ungerade)*
$A_3 = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 \text{ ungerade}\}$ *(Summe ist ungerade)*
$\quad = (\{1, 3, 5\} \times \{2, 4, 6\}) \cup (\{2, 4, 6\} \times \{1, 3, 5\})$

A_1, A_2 *sind stochastisch unabhängig*
A_1, A_3 *sind stochastisch unabhängig*
A_2, A_3 *sind stochastisch unabhängig*

z.B. $P(A_2 \cap A_3) = P(\{2, 4, 6\} \times \{1, 3, 5\}) = \dfrac{|\{2, 4, 6\} \times \{1, 3, 5\}|}{|\Omega|} = \frac{9}{36} = \frac{1}{4}$
$\quad = \frac{1}{2} \cdot \frac{1}{2} = P(A_2) \cdot P(A_3)$,
d.h. A_2, A_3 *sind stochastisch unabhängig.*

Aber A_1, A_2, A_3 *sind* <u>*nicht*</u> *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

$\rightsquigarrow$ Abbildung $X : \Omega \to S, S \neq \emptyset$ 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 $X : \Omega \to S$ 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 $(i \in \mathbb{N})$

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

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 interessiere nur die Augensumme

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

$X : \Omega \to \{2, ..., 12\} =: S$
$X(\omega_1, \omega_2) := \omega_1 + \omega_2$

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

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

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

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

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

(i) $X_1, ..., X_n$ *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 $f_{(X_1,...,X_n)}$ von $P^{(X_1,...,X_n)}$ hat die Form

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

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

In dem Fall hat P^{X_i} die Zähldichte

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

Beweis:
(i) $\Rightarrow$ (ii) ist klar, da $\{X_i \in B_i\}, 1 \leq i \leq n$, stochastisch unabhängig.
$\Rightarrow P(\bigcap\limits_{i=1}^{n} \{X_i \in B_i\}) = P\{X_i \in B_i \; \forall 1 \leq i \leq n\} = \prod\limits_{i=1}^{n} P\{X_i \in B_i\}$

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

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

d.h. $f_{(X_1,...,X_n)}(t_1,...,t_n) = P\{(X_i,...,X_n) = (t_1,...,t_n)\} = \prod\limits_{i=1}^{n} P\{X_i = t_i\} = \prod\limits_{i=1}^{n} f_{X_i}(t_i)$
d.h. (iii)

(iii) $\Rightarrow$ (3.1)

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

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

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

$\forall B_i \in S_i$

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

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

Dies folgt sofort aus (ii) mit $B_i := S_i \ \forall i \notin J$, 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
$f(\omega) = P\{(\omega_1, ...\omega_n)\} = p^{\sum[i=1]n\omega_i} \cdot (1 - p)^{\sum_{i=1}^{n}(1-\omega_i)}$
(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 $X_i : \Omega \to \{0, 1\}, X_i(\omega) = \omega_i$, 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.: $X_1, ..., X_n$ sind stochastisch unabhängig.
Denn: Zähldichte von $P^{(X_1,...,X_n)}$

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

hat Produktgestalt.

Betrachte die Zufallsvariable $Y := \sum[i=1]nX_i$*, also die Anzahl der Erfolge in* n *Experimenten.*
P^Y *hat Zähldichte*

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

denn es gibt genau $\binom{n}{k}$ *Möglichkeiten, die* k *Stellen im Vektor* $(\omega_1,...,\omega_n)$ *auszuwählen, an denen eine 1 steht.*
Daher: $P^Y = \mathcal{B}_{(n,p)}$ *: Binomialverteilung (→ Bsp 1.11.(i))*
Speziell $n=1$*:* $\mathcal{B}_{(1,p)}$ *:* <u>*Bernoulli-Verteilung*</u>
Also sind $X_1,...,X_n$ *stochastisch unabhängig mit* $P^{X_i} = \mathcal{B}_{(1,p)}$

Beispiel 3.6. <u>*Capture-Recapture-Verfahren*</u>
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

<u>*Modell:*</u> *n-mal Ziehen* <u>*ohne*</u> *Zurücklegen aus* N *Fischen und* <u>*ohne*</u> *Berücksichtigung der Reihenfolge.*
$\Omega = \{(\omega_1,...,\omega_n) \in \{1,...,N\}^n \mid \omega_1 < \omega_2 < ... < \omega_n\}$
markierte Fische entsprechen den Nummern $1,...,M$
P *Laplace-Verteilung*
$X(\omega) = \sum[i=1]n1_{\{1,...,M\}}(\omega_i)$ *: Anzahl der markierten Fische*

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

mögliche Werte von $X(\omega) : 0 \le X(\omega) \le \min(M,n), n - X(\omega) \le N - M$
$\Rightarrow$ *Wertebereich von* $X : S := \{\max(0, n+M-N), ..., \min(n,M)\}$

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

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.*

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

Die rechte Seite gibt die Zähldichte eines Wahrscheinlichkeitsmaßes auf S *an, der so genannten* <u>*hypergeometrische Verteilung*</u> $\mathcal{H}_{(N,M,n)}$.
Es gilt also $P^X = \mathcal{H}_{(N,M,n)}$

Die hypergeometrische Verteilung beschreibt also die Anzahl der "Erfolge" beim Ziehen <u>*ohne*</u> *Zurücklegen, während die Binomialverteilung die Anzahl beim Ziehen* <u>*mit*</u> *Zurücklegen beschreibt.*

Falls $n \ll N$, *dann ist Ziehen mit oder ohne Zurücklegen fast identisch und daher*
$\mathcal{H}_{(N,M,m)}\{m\} \approx \mathcal{B}_{(n,\frac{M}{N})}\{m\} \; \forall 0 \le m \le n$

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

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* $\lim\limits_{n\to\infty} n \cdot p_n = \lambda > 0$, *so gilt für* $k \in \mathbb{N}_0$
$\lim\limits_{n\to\infty} \mathcal{B}_{(n,p_n)}(\{k\}) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} =: f_\lambda(k)$
$f_\lambda : \mathbb{N}_0 \to [0,1]$ *ist eine Zähldichte.*
Das zugehörige Wahrscheinlichkeitsmaß $\mathcal{P}_\lambda$ *heißt* <u>*Poisson-Verteilung*</u> *mit Parameter* λ.

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

$\mathcal{B}_{(n,p_n)}\{k\} = \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{n\cdot n\cdot\ldots\cdot n} \cdot (n \cdot p_n)^k \cdot (1 - \frac{n\cdot p_n}{n})^n \cdot (1-p_n)^{-k} \cdot \frac{1}{k!}$
$\xrightarrow[n\to\infty]{} \dfrac{\lambda^k \cdot e^{-\lambda}}{k!}$,
da $(n-i)/n \to 1 \; \forall 0 \le i \le k-1$, $(1 - np_n/n)^n \to e^{-\lambda}$ und wegen $np_n \to \lambda \Rightarrow p_n \to 0$ schließlich $(1-p_n)^{-k} \to 1$.

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

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*

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

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

Beispiel 3.11 (Quicksort)**.** *Die Zahlen $x_1, ..., x_n$, 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: $(n-1)+(n-2)+...+1 = \frac{n\cdot(n-1)}{2}$

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:*

$$(2^k-2)+2\cdot(2^{k-1}-2)+...+2^{k-2}\cdot(2^2-2) = (k-2)\cdot 2^k + 2 \approx n\cdot\log_2 n$$

Später: mittlere Zahl der benötigten Vergleiche $\approx \underbrace{\frac{2}{\log 2}}_{\approx 2{,}89}\cdot$ *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*

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

Angenommen k-te kleinste Zahl ausgewählt:

Dann hat X_l *die Länge* $k-1$ *und* X_r *die Länge* $n-k$

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

Dann

$P\{Z(x_1,...,x_n) = l\} = P\{Z(X_l)+Z(X_r) = l-(n-1)\}$

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 $n \in \mathbb{N}, s_0, ..., s_{n+1} \in S$ mit $P\{X_n = s_n, X_{n-1} = s_{n-1}, ..., X_0 = s_0\} > 0$

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

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.

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

Allgemein: $P\{X_i \in A_i \; \forall 0 \leq i \leq n\} = \sum\limits_{\substack{s_i \in A_i \\ (0 \leq i \leq n)}} f_0(s_0) \cdot p_{s_0 s_1} \cdot p_{s_1 s_2} \cdot \ldots \cdot p_{s_{n-1} s_n}$

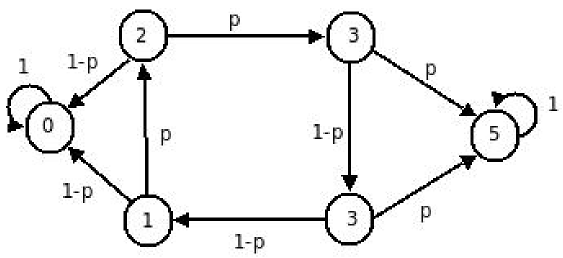
Beispiel 3.13. *Spieler A und B spielen das folgende Glücksspiel: Spieler A bestimmt den Einsatz und wirft eine Münze.*
Falls Kopf fällt, so zahlt B den Einsatz an A, sonst zahlt A den Einsatz an B.

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. $p_{5\,4} = p$, $p_{3\,4} = 1-p$, $p_{i\,4} = 0 \, \forall i \notin \{3, 5\}$
Die Zustände 1 und 5 heißen absorbierend, da sie nicht wieder verlassen werden können.

Mit welcher Wahrscheinlichkeit erhält Spieler A schließlich 5 €?

$p_k = P(X_l = 5$ *für ein* $l > n \mid X_n = k)$ *(hängt nicht von n ab)*
Dann:

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

** = Markov-Eigenschaft*

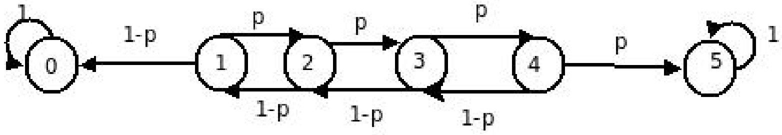
Ebenso erhält man die Beziehungen:
$p_2 = p \cdot p_4$
$p_3 = p + (1-p) \cdot p_1$
$p_4 = p + (1-p) \cdot p_3$
Diese 4 Gleichungen bilden ein lineares Gleichungssystem, das in Matrixschreibweise wie folgt aussieht:

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

Da die Matrix A invertierbar ist, existiert eine eindeutige Lösung $p_1, ..., p_4$

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

Zum Vergleich betrachten wir einen vorsichtigen Spieler $\tilde{A}$, der in der Situation von Spieler A immer nur 1 € 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 $p_1 = \dfrac{4}{1 - 3p + 4p^2 - 2p^3 + p^4}$.

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,...,N
X_m *= Nr der Website, auf der der Surfer nach m Schritten ist.*
S-wertige Zufallsvariable mit $S = \{1, ..., N\}$

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

Dann

$$q_{i\,j} := P(X_{m+1} = j \mid X_m = i) = \begin{cases} 0 & falls j \in L_i \\ \frac{1}{|L_i|} & falls i \in L_i \end{cases}$$

$$= P(X_{m+1} = j \mid X_m = i, X_{m-1} = i_{m-1}, ..., X_0 = i_0)\ \forall (i_0, ..., i_{m-1}, i, j) \in S^{m+2}$$

$$mit\ P\{X_m = i, X_{m-1} = i_{m-1}, ..., X_0 = i_0\} > 0$$

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:

$$p_j^{(m+1)} = P\{X_{m+1} = j\} = \sum_{i=1}^{N} P(X_{m+1} = j \mid X_m = i) \cdot P(X_m = i)$$

$$= \sum_{i=1}^{N} q_{ji} \cdot p_i^{(m)} = (Q \cdot p^{(m)})_j \qquad \forall j \in S$$

$$\Rightarrow p^{m+1} = Q \cdot p^{(m)} = Q \cdot (Q \cdot p^{(m-1)}) = Q^2 \cdot p^{(m-1)} = ... = Q^{m+1} \cdot p^0$$

Frage: Wie verhällt sich $p^(m)$ *für* $m \to \infty$*?*
Erinnerung: λ *Eigenwert der Matrix A mit Eigenvektor* $v :\Leftrightarrow Av = \lambda v$

<u>Satz von Perron</u>:
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 $A = Q^{m_0}$

$$p^{(m)} = Q^{\cdot} \cdot p^{(0)} = A^{\frac{m}{m_0}} \cdot p^{(0)} \xrightarrow[m \to \infty]{} p_s \qquad (m \text{ Vielfaches von } m_0)$$

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 <u>stationäre Verteilung</u> der Markov-Kette.
Ist $P\{X_0 = i\} = (p_s)_i \; \forall i$,

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 $A^k \cdot v \xrightarrow[k \to \infty]{} p_s$ erfolgt exponentiell schnell.*
  $\rightsquigarrow p_s$ relativ effizent berechenbar
- *Annnahme (*) 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* <u>*Erwartungswert*</u> *einer* $\mathbb{R}$*-wertigen Zufallsvariable* X *auf einem diskreten Wahrscheinlichkeitsraum* (Ω, P) *ist def. als* $E_P(X) = E(X) := \sum\limits_{x \in X(\Omega)} x \cdot P\{X = x\}$, *falls* $\sum\limits_{x \in X} |x| \cdot P\{X = x\} < \infty$ *(Andernfalls besitzt* X *keinen Erwartungswert)*

(ii) *Sei* $(\mathbb{R}, Q)$ *ein diskreter Wahrscheinlichkeitsraum, dann heißt* $\mu(Q) := \sum\limits_{x \in \mathbb{R}} x \cdot Q(\{x\})$ <u>*Mittelwert*</u> *von* Q, *falls* $\sum\limits_{x \in X} |x| \cdot Q(\{x\}) < \infty$

Bemerkung 4.2. $Id : \mathbb{R} \to \mathbb{R}, Id(x) := x$ *(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,* $X : \Omega \to S$ *eine Zufallswariable und* $g : S \to \mathbb{R}$. *Dann gilt*
$$E_P(\underbrace{g(X)}_{g \circ X}) = E_{P^X}(g)$$

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

Beispiel 4.4.

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

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

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

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

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

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

(ii) gilt $X(\omega) \leq Y(\omega)\ \forall \omega \in \Omega$*, 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

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

2. Lösung: Bsp 3.5:

X_i *seien unabhängige Zufallsvariablen mit*

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

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

$\Rightarrow E(X) = E(\sum\limits_{n=1}^{n} X_i) = \sum\limits_{n=1}^{n} E(X_i) = \sum\limits_{n=1}^{n} (0 \cdot P\{X_i = 0\} + 1 \cdot P\{X_i = 1\}) = n \cdot p$

Beispiel 4.7. *Erwartete Laufzeit von Quicksort (Fortsetztung von Bsp. 3.11)*
$Z(x_1, ..., x_n)$ = *Zufällige Zahl der Vergleiche, die benötigt werden, um* $x_1, ..., x_n$ *zu sortieren.*
Mit Wahrscheinlichkeit $\frac{1}{n}$ *wähle eine Zahl* $x_j \quad (1 \leq j \leq n)$
Sortiere Zahlen $< x_j$ *in Vektor* X_l *und Zahlen* $> x_j$ *in Vektor* X_r
Dann

$$Z(x_1, ..., x_n) = n - 1 + Z(X_l) + Z(X_r)$$
$$\mu_n = E(Z(x_1, ..., x_n)) \qquad \textit{(hängt nur von n ab!)}$$

Falls x_j *die k kleinste Zahl ist, dann hat* X_l *die länge k-1 also*

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

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

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

$$
\begin{aligned}
\Rightarrow \quad && n\cdot\mu_n &= n\cdot(n-1)+2\cdot\sum_{j=0}^{n-1}\mu_j \\
\Rightarrow \quad && n\cdot\mu_n-(n-1)\cdot\mu_{n-1} &= n\cdot(n-1)-(n-1)\cdot(n-2)+2\mu_{n-1} \\
\Rightarrow \quad && n\cdot\mu_n &= (n+1)\cdot\mu_{n-1}+2(n-1) \\
\Rightarrow \quad && \mu_n &= \frac{n+1}{n}\cdot\mu_{n-1}+2\cdot\frac{n-1}{n} \\
&& &= \frac{n+1}{n}\cdot\left(\frac{n}{n-1}\cdot\mu_{n_2}+2\cdot\frac{n-2}{n-1}\right)+2\cdot n-1n \\
&& =^{Vollst.Induktion} \ldots &= 2(n+1)\cdot\sum_{k=2}^{n}\frac{k-1}{k(k+1)}+\underbrace{\frac{4}{1}\mu_0}_{=0}
\end{aligned}
$$

Umformungen liefern:

$$
\mu_n = 2(n+1)\cdot\sum_{j=1}^{n+1}\frac{1}{j}-4n-2 \qquad \begin{cases} \le 2n\log n \\ \ge 2n\log n-4n \end{cases}
$$

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

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

$$
P\left(\bigcup_{i=1}^{n}A_i\right)
$$
$$
=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j\le n}P(A_j\cap A_i)+\sum_{1\le i<j<k\le n}P(A_i\cap A_j\cap A_k)-...+(-1)^{n-1}P(A_1\cap A_2\cap...\cap A_n)
$$

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

$$\begin{aligned}
1_{\bigcup\limits_{i=1}^{n} A_i} &= 1 - 1_{(\bigcup\limits_{i=1}^{n} A_i)^c} = 1 - 1_{\bigcap\limits_{i=1}^{n} A_i^c} = 1 - \prod_{i=1}^{n} 1_{A_i^c} = 1 - \prod_{i=1}^{n}(1 - 1_{A_i}) \\
&= 1 - (1 - \sum_{i=1}^{n} 1_{A_i} + \sum_{1\leq i<j\leq n} 1_{A_i} \cdot 1_{A_j} - ... + (-1)^{n-1}\prod_{i=1}^{n} 1_{A_i}) \\
&= \sum_{i=1}^{n} 1_{A_i} - \sum_{1\leq i<j\leq n} 1_{A_i \cap A_j} + ... + (-1)^{n-1}\bigcap_{i=1}^{n} A_i
\end{aligned}$$

$$\Rightarrow P(\bigcup_{i=1}^{n} A_1) = E(1_{\bigcup\limits_{i=1}^{n} A_1}) = \sum_{i=1}^{n}\underbrace{E(A_i)}_{P(A_i)} - \sum_{1\leq i<j\leq n}\underbrace{E(1_{A_i\cap A_j})}_{P(A_i\cap A_j)} + ... + (-1)^{n-1}\underbrace{E(1_{\bigcup\limits_{i=1}^{n} A_i})}_{P(\bigcap\limits_{i=1}^{n} A_i)}$$

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

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

die Varianz von X genannt und $\sqrt{Var(X)}$ 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:
$X : \Omega \to \mathbb{R}$
$X^k : \Omega \to \mathbb{R}$
$\omega \to (X(\omega))^k$

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.*

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

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

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

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

$\Rightarrow E(X^2) = E(X\cdot(X-1)) + E(X) = \lambda^2 + \lambda$
$\Rightarrow Var(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

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

(i) $P\{|X| \geq c\} \leq \dfrac{E(|X|)}{c} \quad \forall c > 0 \qquad$ <u>*Markov-Ungleichung*</u>

(ii) $P\{|X - E(X)| \geq c\} \leq \dfrac{Var(X)}{c^2} \quad \forall c > 0 \qquad$ <u>*Chebyshev-Ungleichung*</u>

Beweis:

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

(ii) in den Übungen

Beispiel 4.13.

(i) Sei X $\mathcal{P}_\lambda$-verteilt, $\lambda > 0 \to E(X) = \lambda = Var(X)$
$\Rightarrow P\{X \geq c\} \leq \dfrac{E(X)}{c} = \dfrac{\lambda}{c} \qquad \forall c > 0 \qquad$ *(Markov-Ungleichung)*
und für $c > \lambda$
$P\{X \geq c\} = P\{X - \lambda \geq c - \lambda\} \leq P\{|X - \lambda| \geq c - \lambda\} \leq \dfrac{Var(X)}{(c-\lambda)^2} = \dfrac{\lambda}{(c-\lambda)^2}$
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 größter als $P\{X > c\}$

(ii) Quicksort

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

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

Man kann zeigen: $Var(Z(x_1, ..., x_n)) \leq 3n(n-1)$

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

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

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

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 Erwartungwert <u>nicht</u> linear im Argument.*

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

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

Definition 4.15 (Kovarianz & Korrelation)**.**
Sind X, Y Zufallsvariablen mit Erwartungswerten $E(X)$ bzw. $E(Y)$.

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

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

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

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

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. Y 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 $X_1, ..., X_n$ Zufallsvariablen mit Erwartungswerten $E(X_1), ..., E(X_n)$. Dann gilt:*

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2 \cdot \sum_{1 \leq i < j \leq n} Corr(X_1, X_j) \tag{4.1}$$

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

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

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 $\sum\limits_{i=1}^{n} X_i$, X_i sind stochastisch unabhängig und X_i ist $\mathcal{B}_{(n,p)}$-verteilt, d.h. $P\{X_i = 1\} = p, P\{X_i = 0\} = (1-p)$

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

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.

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

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

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

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

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

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

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

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

Der zugehörige mittlere quadratische Approximationsfehler ist

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

$|Corr(X,Y)| = 1 \Rightarrow$ *kein Vorhersagefehler*
$Corr(X,Y) = 0 \Rightarrow$ *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 vollständigt 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 $n \to \infty$

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

$$P\{\frac{1}{n}|\sum_{i=1}^{n}(X_i - E(X_i))| > \epsilon)\} \xrightarrow[n\to\infty]{} 0$$

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

$$P\{\frac{1}{n}|\sum_{i=1}^{n}(X_i - E(X_i))| \geq c\} = P\{|\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)| \geq n \cdot c\}$$

$$\underset{Chebyshev}{\leq} \frac{Var\left(\sum_{i=1}^{n} X_i\right)}{(n \cdot c)^2} \leq \frac{n \cdot M}{n \cdot c^2} \xrightarrow[n\to 0]{} 0 \Rightarrow Beh.$$

Definition 5.2. *Seien Y, Y_n $\mathbb{R}$-wertige Zufallsvariablen. Gilt für alle $\epsilon > 0$*
$P\{|Y_n - Y| \ \geq 0\} \xrightarrow[n\to\infty]{} 0,$
so sagt man, dass Y_n (P-)Stochastisch gegen Y konvergiert (in Zeichen $Y_n \xrightarrow{P} Y$ oder $Y_n \to Y$ 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 $X_n := \sum_{n=1}^{n} X_i \quad \mathcal{B}_{(n,p)}$
$E(X_i) = p \underset{bzw.5.3}{\overset{5.1}{\Rightarrow}} \frac{Y_n}{n} \xrightarrow{P} p$

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

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

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

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: $Var(X_i) = p \cdot (1-p) \leq \frac{1}{4}$
Der Beweis von Satz 5.1 zeigt:

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

Wähle ϵ so, dass $\frac{1}{4n \cdot \epsilon} = 0,05 \Leftrightarrow \epsilon = \dfrac{1}{\sqrt{4n \cdot 0,05}}$
Dann $P\{|\frac{1}{4}\sum_{i=1}^{n} X_i - p| \geq \epsilon\} \leq 0,05$

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

$$P\{\frac{1}{n}\sum_{i=1}^{n} X_i \notin (0,4;0,6)\} = P\{|\frac{1}{4}\sum_{i=1}^{n} X_1 - \frac{1}{2}| \geq \epsilon = \frac{1}{1}0\} \leq 0,05$$

, d.h. falls bei 500 Würfen 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 $Z = f(Y_1, ..., Y_k)$, 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:

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

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

Konkretes Beispiel:

Seien Y_1, Y_2 gleichverteilt auf $\{1, \frac{1}{L}, \frac{2}{L}, ...1\} =: S_L \quad (L \in \mathbb{N})$
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 = independend and identially distributed) (d.h. $P^{X_i} = P^{X_1} \forall i$) mit Erwartungswert $\mu = E(X_i)$ und Varianz $\sigma^2 = Var(X_i) \in (0, \infty)$ Dann gilt für alle $x \in \mathbb{R}$*

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

Φ heißt Standartnormalverteilungsfunktion
$\varphi = \Phi'$ heißt Standartnormalverteilungsdichte
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:

$$\Phi(-\infty) := \lim_{x \to -\infty} \Phi(x) = 0$$
$$\Phi(\infty) := \lim_{x \to \infty} \Phi(x) = 1$$
$$\varphi(t) = \varphi(-t) \qquad \forall t \in \mathbb{R}$$

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

d.h. $\Phi(-x) = 1 - \Phi(x)$
Insbesondere $\Phi(0) = \frac{1}{2}$