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?

$\rightarrow$ 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 $\Rightarrow P_n(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P_n(A_i)$

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 $P : 2^\Omega \to [0,1]$ 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	Ω_a	$= \{1, ..., n\}^k = \{(\omega_1, ..., \omega_k) \mid \omega_i \in \{1, ..., n\}\}$
Interpretation:	ω	$=(\omega_1, ..., \omega_k) \in \Omega_a$
		$\widehat{=}$ 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 ohne Zurücklegen und ohne 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 mit Zurücklegen und ohne 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 $|\Omega_d| = |\tilde{\Omega}_c| \overset{(c)}{=} \binom{n+k-1}{k}$

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)*

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

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

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 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 <u>nur</u> 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 <u>bedingte Wahrscheinlichkeit von A gegeben B</u>

Beispiel 2.2. *(vergleiche Präsenzübungsblatt 2, A3)*
Beim Skatspiel hat Spieler 1 keinen Buben erhalten.
Wie groß ist <u>dann</u> die Wahrscheinlichkeit, dass Spieler 2 oder Spieler 3 alle 4 Buben erhalten hat?
<u>Modell</u>: $\Omega = \{(\underbrace{\omega_1, ..., \omega_{10}}_{Karten\,Sp.1}, \underbrace{\omega_{11}, ..., \omega_{20}}_{Karten\ Sp.2}, \underbrace{\omega_{21}, ..., \omega_{30}}_{KartenSp.3}, \underbrace{\omega_{31}, \omega_{32}}_{Stock},) \in K^{32} \mid \omega_i \neq \omega_j\ \forall i \neq j\}$
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:

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

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* <u>*(P-)stochastisch unabhängig*</u>*, 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* <u>*(P-)stochastisch unabhängig*</u>*, wenn für jede Indexmenge* $I \subset \{1, ..., n\}, I \neq \emptyset$*, gilt*
$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$

Bemerkung 2.7.

(i) *Definition 2.6 stellt sicher, dass jede beliebige Auswahl* $A_i, i \in I \subset \{1, ..., n\}$ *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* $P(\bigcup\limits_{i=1}^{n} A_i) = \prod\limits_{i=1}^{n} P(A_i)$

Ist z.B. $A_1 = \emptyset$, *so gilt* $P(\bigcap\limits_{i=1}^{n} A_i) = P(\emptyset) = 0 = \prod\limits_{i=1}^{n} P(A_i)$

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

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

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 $X_i : \Omega \to S_i, 1 \leq i \leq n$ (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

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

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

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

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

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 $P^{(X_1,...,X_n)}$

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

hat Produktgestalt.

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

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

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 (*$\rightarrow$ *Bsp 1.11.(i))*
Speziell $n = 1$*:* $\mathcal{B}_{(1,p)}$ *:* *Bernoulli-Verteilung*
Also sind $X_1, ..., X_n$ *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.
$\Omega = \{(\omega_1, ..., \omega_n) \in \{1, ..., N\}^n \mid \omega_1 < \omega_2 < ... < \omega_n\}$
markierte Fische entsprechen den Nummern $1, ..., M$
P *Laplace-Verteilung*
$X(\omega) = \sum_{i=1}^{n} 1_{\{1,...,M\}}(\omega_i)$ *: 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 $X(\omega) : 0 \leq X(\omega) \leq \min(M, n), n - X(\omega) \leq N - M$
$\Rightarrow$ *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* *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* $\mathcal{H}_{(N,M,m)}\{m\} \approx \mathcal{B}_{(n,\frac{M}{N})}\{m\}\ \forall 0 \le m \le n$

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* $\lim_{n\to\infty} n \cdot p_n = \lambda > 0$*, so gilt für* $k \in \mathbb{N}_0$

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

$f_\lambda : \mathbb{N}_0 \to [0,1]$ *ist eine Zähldichte.*
Das zugehörige Wahrscheinlichkeitsmaß $\mathcal{P}_\lambda$ *heißt* *Poisson-Verteilung* *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: $\underbrace{3\ 2\ 6\ 1}_{X_l}|7|\underbrace{13}_{X_r} \quad \rightsquigarrow$ 5 Vergleiche
2. Wahl 3: $2\ 1|3|6|7|13 \quad \rightsquigarrow$ 3 Vergleiche

   $1\ 2|3|6|7|13 \quad \rightsquigarrow$ 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: $P\{X_i \in A_i \; \forall 0 \leq i \leq n\} = \sum\limits_{\substack{s_i \in A_i \\ (0 \leq i \leq n)}} f_0(s_0) \cdot p_{s_0 s_1} \cdot p_{s_1 s_2} \cdot \ldots \cdot p_{s_{n-1} s_n}$

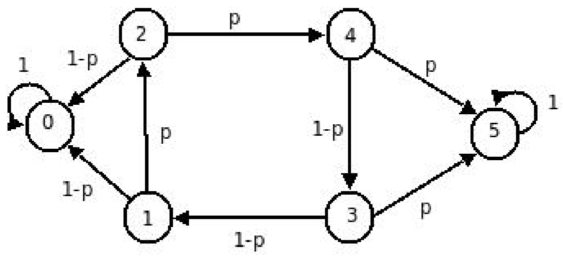
Beispiel 3.13. *Spieler A und B spielen das folgende Glücksspiel: Spieler A bestimmt den Einsatz und wirft eine Münze.*
Falls Kopf fällt, so zahlt B den Einsatz an A, sonst zahlt A den Einsatz an B.

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

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

** = Markov-Eigenschaft*

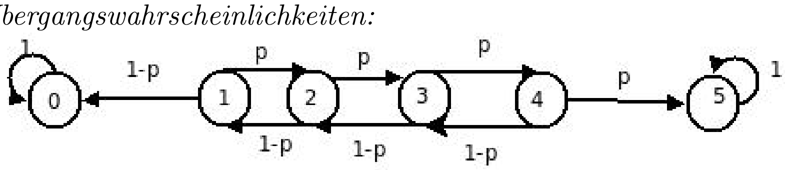
Ebenso erhält man die Beziehungen:
$p_2 = p \cdot p_4$
$p_3 = p + (1-p) \cdot p_1$
$p_4 = p + (1-p) \cdot p_3$
Diese 4 Gleichungen bilden ein lineares Gleichungssystem, das in Matrixschreibweise wie folgt aussieht:

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

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.

<u>*Modell*</u>
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

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

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

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

  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* $E_P(X) = E(X) := \sum\limits_{x \in X(\Omega)} x \cdot P\{X = x\}$*, 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* $\mu(Q) := \sum\limits_{x \in \mathbb{R}} x \cdot Q(\{x\})$ *Mittelwert* *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 Zufallsvariable 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*

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

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

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

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 $(1 \leq j \leq n)$
Sortiere Zahlen $< x_j$ *in Vektor* X_l *und Zahlen* $> x_j$ *in Vektor* X_r
Dann

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

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

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

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

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

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

Umformungen liefern:

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

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

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

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

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

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

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

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: $0 \le Var(X) = E(X^2) - (E(X))^2 \Rightarrow E(X^2) \ge (E(X))^2$

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

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

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

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

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

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

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,* $\lambda > 0 \rightarrow E(X) = \lambda = Var(X)$
$\Rightarrow P\{X \geq c\} \leq \dfrac{E(X)}{c} = \dfrac{\lambda}{c} \qquad \forall c > 0$ *(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. 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 $X_1, ..., X_n$ Zufallsvariablen mit Erwartungswerten $E(X_1), ..., E(X_n)$. Dann gilt:*

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

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 <u>nur</u> Maß für Stärke des <u>linearen</u> 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 <u>nicht</u> 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 $X_n := \sum\limits_{n=1}^{n} X_i \quad \mathcal{B}_{(n,p)}$
$E(X_i) = p \underset{bzw. 5.3}{\overset{5.1}{\Rightarrow}} \frac{Y_n}{n} \xrightarrow{P} p$

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 $P\{|\frac{1}{4}\sum\limits_{i=1}^{n} X_i - p| \geq \epsilon\} \leq 0,05$

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.

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

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 $x \in \mathbb{R}$*

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

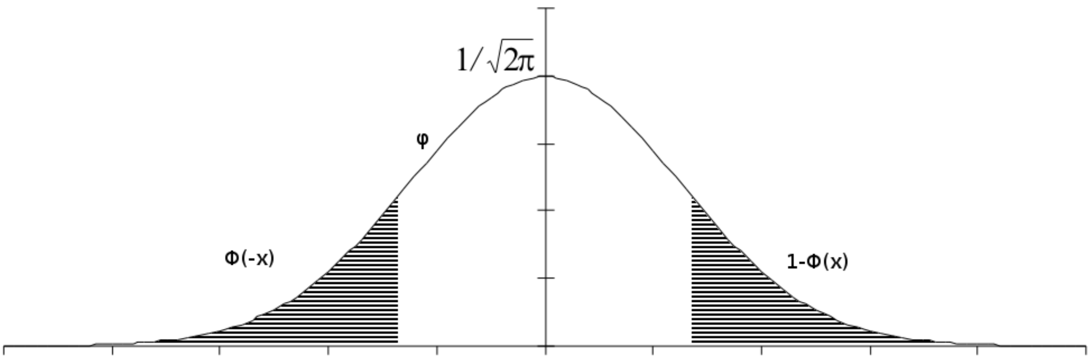
Φ heißt Standardnormalverteilungsfunktion
$\varphi = \Phi'$ heißt Standardnormalverteilungsdichte
d.h. $\varphi(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}, t \in \mathbb{R}$

Beweis: Ist z.B. im Dübgen zu finden.
Es gilt:

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

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

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



Beispiel 5.7. *X_i iid und $\mathcal{P}_\lambda$-verteilt, $\lambda > 0$*
$\Rightarrow \mu = \sigma^2 = \lambda$, $S_n ist \mathcal{P}_{n\lambda}$-verteilt ($\rightarrow$ Bsp. 3.10, 4.4, 4.11)

Für große n

$$P\{\frac{1}{\sqrt{n\lambda}}\underbrace{\sum_{i=1}^{n}(X_i-\lambda)}_{=S_n-n\lambda} \le x\}(\approx \Phi(x))$$
$$= P\{S_n \le n\lambda + \sqrt{n\lambda}x\} = \mathcal{P}_{n\lambda}\{0,...,\lfloor n\lambda + \sqrt{n\lambda}x\rfloor\}$$

$$\Rightarrow P\left\{|\sum_{i=1}^{n}(X_i-\lambda) \ge c\right\} = \underbrace{P\left\{\sum_{i=1}^{n}(X_i-\lambda) \ge c\right\}}_{1-P\{\sum\limits_{i=1}^{n}(X_i-\lambda)<c\}} + P\left\{\sum_{i=1}^{n}(X_i-\lambda) \le -c\right\}$$
$$= 1 - P\left\{\frac{1}{\sqrt{n\lambda}}\sum_{i=1}^{n}(X_i-\lambda) < \frac{c}{\sqrt{n\lambda}}\right\} + P\left\{\frac{1}{\sqrt{n\lambda}}\sum_{i=1}^{n}(x_i-\lambda) \le \frac{c}{\sqrt{n\lambda}}\right\}$$
$$\approx \Phi\left(-\frac{c}{\sqrt{n\lambda}}\right) + 1 - \Phi\left(\frac{c}{\sqrt{n\lambda}}\right)$$
$$= 2\Phi\left(-\frac{c}{\sqrt{n\lambda}}\right)$$

Vergleich mit der Ungleichung von Chebyshev:

$$P\left\{|\sum_{i=1}^{n}(X_i-\lambda)| \ge c\right\} \le \frac{Var\left(\sum\limits_{i=1}^{n}X_i\right)}{c^2} = \frac{n\lambda}{c^2}$$

$n \cdot \lambda = 100$
——————— *Beamervorführung des Vergleiches* ————————————-

Korollar 5.8 (Zentraler Grenzwertsatz von Moivre-Laplace)**.** *Ist Y_n eine $\mathcal{B}_{(n,p)}$-verteilte Zufallsvariable mit $p \in (0,1)$ $\quad(\forall n \in N)$, so gilt*

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

Beweis:
Ist X_i $\mathcal{B}_{(n,p)}$-verteilt und unabhängig,
dann ist $S_n = \sum\limits_{i=1}^{n} X_i$ genauso verteilt wie Y_n ($\rightarrow$ Bsp 3.5)

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

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

Beispiel 5.9. *Das Flugzeug eines Fluges kann $n_0 = 200$ Personen befördern. Mit Wahrscheinlichkeit $p = 0,96$ erscheinen die Ticketinhaber jeweils unabhängig voneinander 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 $P\{X > n_0\} \leq 0,05$

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

$b = \infty$

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

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

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

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

Also dürfen maximal 203 Tickets verkauft werden.

Modifikation: $P\{X > n_0\} = P\{X > t\} \qquad \forall t \in [n_0, n_0 + 1)$
Genauso wie oben ergiebt 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) erhällt man die Bedingung

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

In der Tat ist

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

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)*

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

$\frac{1}{n}\sum\limits_{i=1}^{n} X_i =$ *relative Häufigkeit, mit der Kopf fällt*
Test: Falls $|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \frac{1}{2}| \geq \epsilon$, *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* $P\{|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \frac{1}{2}| \geq \epsilon\} \leq \alpha$, *falls die Münze tatsächlich fair ist.*

Mittels des Zentralengrenzwertsatzes:

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

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

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

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

ist.

Vergleich mit dem Intervall, dass sich aus der Chebyshev'schen Ungleichung ergeben hat (→ 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. Lngen, gewichte, Zeiten) lassen sich nicht auf natürliche Weise durch diskrete Wahrscheinlichkeitsräume bescheiben.

Im Allgemeinen kann man <u>nicht</u> 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 $\quad P : \mathcal{A} \to [0, 1]$
mit $P(\mathbb{R}) = 1$, $A_n \in \mathcal{A}$ disjunkt, dann $P(\bigcup\limits_{n \in N} A_n = \sum\limits_{n \in \mathbb{N}} P(A_n)$

Frage: Wie kann man Wahrscheinlichkeitsmaße <u>einfach</u> 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 disjungt 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 $F : \mathbb{R} \to [0, 1]$, $F(x) := P((-\infty, x])$ definierte <u>Verteilungsfunktion</u> von P*

(i) F ist monoton Steigend

(ii) F ist rechtsseitig stetig

(iii) $F(\infty) := \lim\limits_{x \to \infty} F(x) = 1$, $F(-\infty) := \lim\limits_{x \to -\infty} F(x) = 0$

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 $f : \mathbb{R} \to [0, \infty)$, so dass*

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

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 $f : \mathbb{R}- > [0, \infty)$ 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 $f : \frac{1_{(1,b]}}{b-a}$ als eine Dichte eines Wahrscheinlichkeitsmaßes $\mathcal{U}_{(a,b]}$ auf $\mathbb{R}$, denn $\int_{\mathbb{R}} f(t)dt = \frac{1}{b-a}$ und $\int_{\mathbb{R}} 1_{(a,b]}tdt = \frac{b-a}{b-a} = 1$*

$\mathcal{U}_{(a,b]}$ 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) $f_\lambda(x) := \frac{1}{\lambda} \cdot e^{-\frac{x}{\lambda}} \cdot 1_{[0,\infty]}(x) \qquad (\lambda > 0)$ *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*

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

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

Für $x, y \geq 0$ *gilt*

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

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) $\varphi(t) := \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} \quad , t \in \mathbb{R},$
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, bist 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*

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

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

d.h.

$$c \cdot \int\limits_0^1 \min(x, 1-x) dx = 2 \cdot c \int\limits_0^{\frac{1}{2}} x\, dx = cx^2\Big|_0^{\frac{1}{2}} = \frac{c}{4} \overset{!}{=} 1$$

$\Rightarrow$ *$c = 4$, 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

$$\begin{aligned} P((0, \frac{1}{4}) \cup (\frac{3}{4}, 1)) &= \int 1_{(0,\frac{1}{4}) \cup (\frac{3}{4},1)}(x) \cdot f(x)\, dx \\ &= \int\limits_0^{\frac{1}{4}} 4x\, dx + \int\limits_{\frac{3}{4}}^{1} 4 \cdot (1-x)\, dx \overset{s.o.}{=} 4x^2\Big|_0^{\frac{1}{4}} = \frac{1}{4} \end{aligned}$$

Beim diskreten Wahrscheinlichkeitsraum (Ω, P) ist eine Abbildung $X : \Omega \to S$ 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 $P : \underline{\underline{\mathcal{A}}} \to [0, 1]$. Daher ist $P(X^{-1}(B))$ nur noch definiert, wenn $X^{-1}(B) \in \mathcal{A}$, d.h. "gutartig" ist.

Definiert man $X : \mathbb{R} \to \mathbb{R}$ so, dass $X^{-1}(B) \in \mathcal{A} \qquad \forall B \in \mathcal{A}$,
dann gilt wie oben $P^X(B) := P(\{B \in B\}) \qquad \forall B \in \mathcal{A}$ 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 $P^X(\{x\}) = \int\limits_{\mathbb{R}} f(t) \cdot 1_{\{x\}}(t)\, dt = 0 \quad \forall x \in \mathbb{R}$.
Daher ist ein neuer Ansatz nötig.

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

$$E(X) = \int\limits_0^\infty P\{x > t\}dt - \int\limits_{-\infty}^0 P\{x < t\}dt,$$

falls die rechte Seite wohl definiert ist.

Zur Veranschaulichung Skitzen, falls X nur Werte $0 < x_1 < x_2 < ... < x_n < \infty$ annnimmt.
$p_i = P\{X = x_i\}$
——————- Skitze 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*

$$E(X) = \int\limits_0^\infty 1 - F(t)\, dt - \int\limits_{-\infty}^0 F(t)\, dt,$$

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

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

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

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

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

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*

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

Wie sehen Systeme "gutartiker 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:*
- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- $A_n \in \mathcal{A} \quad \forall n \in \mathbb{N} \Rightarrow \bigcup\limits_{n \in N} A_n \in \mathcal{A}$

$(\Omega, \mathcal{A})$ heißt dann Messraum.
Die Mengen $A \in \mathcal{A}$ heißen dann Ergebnisse.

(ii) *Eine Abbildung $P : \mathcal{A} \to [0,1]$ heißt Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$, falls*
- $P(\Omega) = 1$
- $A_n \in \mathcal{A},\ n \in \mathbb{N},\ disjunkt \Rightarrow P\left(\bigcup\limits_{n\in\mathbb{N}} A_n\right) = \sum\limits_{n \in N} P(A_n)$ *(σ-Additivität)*

$(\Omega, \mathcal{A}, P)$ 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 $P((c,d]) = d - c \qquad \forall 0 \leq c < d \leq 1$.
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, Messbarbeit und Verteilung)**.** *Seien* $(\Omega, \mathcal{A}), (\tilde{\Omega}, \tilde{\mathcal{A}})$ *Messräume und* $X : \Omega \to \tilde{\Omega}$*.* 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 $X : (\Omega, \mathcal{A}) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$ *und sagt, dass* X $\mathcal{A}, \tilde{\mathcal{A}}$*-messbar ist.*

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

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

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*

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

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*.*

$\mathbb{B}$ enthällt 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

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

eindeutig bestimmt ($\to$ Satz 6.1).

Lemma 6.12. *Sei $(\Omega, \mathcal{A}, P)$ 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 $A := \bigcup_{n \in \mathbb{N}} A_n \qquad$ (kurz $A_n \uparrow A$)

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

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

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

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

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

(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.

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

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):

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

(iii) Beh.: $F(\infty) := \underset{x\to\infty}{F}(x) = 1,\ F(-\infty) := \lim_{x\to-\infty} F(x) = 0$
Beweis: Sei $X_n \uparrow \infty$. Dann $(-\infty, X_n) \uparrow \mathbb{R}$ und 6.12(i) liefert

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

Die zweite Aussage folgt analog.

Analog für $\Omega = \mathbb{R}^n$

Satz & Definition 6.13.

(i) *Die* <u>*Borel-σ-Algebra auf $\mathbb{R}^n$*</u> *wird definiert als*

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

(ii) *Ist P ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathbb{B}^n)$, so wird die zugehörige* <u>*multivariante Verteilungsfunktion*</u> *F definiert als*

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

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

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

gilt, so heißt f (multivariante) <u>*Dichte*</u> *von P bzw. von F.*
Es gilt dann für alle Mengen $B \in \mathbb{B}^n$

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

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

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

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

heißt <u>*gemeinsame Verteilungsfunktion*</u> *von $X_1, ..., X_n$*

Gegenstück zu 3.4

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

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

(ii) $\forall B_i \in B_i$, $1 \le i \le n$, *gilt* $P\{X_i \in B_i \; \forall 1 \le i \le n\} = \prod\limits_{i=1}^{n} P\{X_i \in B_i\}$

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

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

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

Besitzen die so genannten <u>*Randverteilungen*</u> P^{X_i}, $1 \le i \le n$, *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*

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

für geeigenete Funktionen $h_i : \mathbb{R} \to [0, \infty)$.
In dem Fall kann man $h_i = f_i$ *wählen*

Satz 6.15. *Besitzt eine* $\mathbb{R}^n$*-wertige Zufallsvariable* $X = (X_1, \dots, X_n)$ *die gemeinsame Dichte* $f : \mathbb{R}^n \to [0, \infty)$, *so hat* $(X_1, \dots, X_m)$ *für* $m < n$ *die Dichte*

$$g(x_1, \dots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_n \cdots dx_{m+1} \qquad (x_1, \dots, x_m) \in \mathbb{R}$$

Beweis: $(X_1, \dots, X_m)$ hat die gemeinsame Verteilungsfunktion

$$\begin{aligned} G(x_1, \dots, x_m) &= P\{X_1 \le x_1, \dots, X_m \le x_m\} \\ &= P\{X_1 \le x_1, \dots, X_m \le x_m, X_{m+1} < \infty, \dots, X_n < \infty\} \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_n \cdots dt_{m+1}}_{g(t_1,\dots,t_m)} dt_m \cdots dt_1 \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} g(t_1, \dots, t_m) dt_m \cdots dt_1 \end{aligned}$$

d.h. g ist die Dichte von G bzw. von $(X_1, \dots, X_m)$

Satz 6.16 (Transformationssatz von Dichten)**.** *Ist $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, besitzt $X : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathbb{B}^n)$ die Dichte f_X und gilt $P\{X \in M\} = 1$ für eine offene , zusammenhängende Menge $M \subset \mathbb{R}^n$, ist schließlich $g : M \to \mathbb{R}^n$ eine injektive, stetige, differenzierbare Abbildung mit Umkehrabbildung $g^{-1} : g(M) \to M$ und invertierbarer Jacobi-Matrix $Dg(x) = \left(\frac{\sigma\ g_i(x)}{\sigma\ x_j} \right)$. Dann besitzt $g(X) : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathbb{B}^n)$ die Dichte*

$$f_{g(x)}(y) = \frac{f_X(g^{-1}(y))}{|\det Dg(g^{-1}(y))|} \cdot 1_{g(M)}(y) \qquad , y \in \mathbb{R}^n$$

wobei $\dfrac{1}{|\det Dg(g^{-1}(y))|} = |\det Dg^{-1}(y)|$

Beispiel 6.17 (Wettschießen)**.** *Der Mittelpunkt einer Zielscheibe sei der Ursprung* $(0, 0)$*. Die Treffer landen zufällig im Punkt (X, Y), wobei X, Y stochastisch unabhängig und $\mathcal{N}_{(0,1)}$-verteilt sind, d.h. sie besitzen jeweils die Dichte*

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

Gesucht: Die Verteilung der Entfernung zum Mittelpunkt, also $\sqrt{X^2 + Y^2}$

Gemäß Satz 6.14 hat (X, Y) die Dichte

$$f(x, y) = \varphi(x) \cdot \varphi(y) = \frac{1}{2\pi} e^{-\frac{(X^2+y^2)}{2}} \qquad , (x, y) \in \mathbb{R}^2$$

Transformation in Polarkoordinaten:

$g : \mathbb{R}^2 \setminus \{0\} \to (0,\infty)x(-\pi,\pi)$

$$g(x,y) = \begin{cases} (\sqrt{x^2+y^2}, \arccos \frac{x}{\sqrt{x^2+y^2}}), \textit{ falls } y \geq 0 \\ (\sqrt{x^2+y^2}, -\arccos \frac{x}{\sqrt{x^2+y^2}}), \textit{ falls } y < 0 \end{cases}$$
$$:= (r,\vartheta)$$

g erfüllt die Vorraussetzung von Satz 6.16
$g^{-1}(r,\vartheta) = (r\cdot\cos\vartheta, r\cdot\sin\vartheta)$
$Dg^{-1}(r,\vartheta) = \begin{pmatrix} \cos\vartheta & -r\cdot\sin\vartheta \\ \sin\vartheta & r\cos\vartheta \end{pmatrix}$
$\Rightarrow |\det Dg^{-1}(r,0)| = r\cdot(\cos^2\vartheta + \sin^2\vartheta) = r$

Mittels Satz 6.16 hat $g(X,Y)$ die Dichte

$$\begin{aligned} f_{g(X,Y)}(r,\vartheta) &= |\det Dg^{-1}(r,\vartheta)| \dot{f}_{(X,Y)}(g^{-1}(r,\vartheta)) \\ &= r\cdot\frac{1}{2\pi}e^{-\frac{r^2}{2}}\cdot 1_{(0,\infty)}(r)\cdot 1_{(-\pi,\pi)}(\vartheta) \\ &= \underbrace{r\cdot e^{-\frac{r^2}{2}}\cdot 1_{(0,\infty)}(r)}_{=:h_1(r)}\cdot\underbrace{\frac{1}{2\pi}\cdot 1_{(-\pi,\pi)}(\vartheta)}_{:=h_2(\vartheta)} \end{aligned}$$

Nach Satz 6.14 sind der Abstand vom mittelpunkt und der Winkel zur X-Achse stochastisch unabhängig, da $f_{g(X,Y)}$ Produktgestalt hat.
Dabei ist h_2 die Dichte des zufälligen Winkels zur X-Achse, d.h. der Winkel ist gleichverteilt auf $(-\pi,\pi]$.
Der Abstand hat die Dichte h_1, die zugehörige Verteilung heißt Rayleight-Verteilung. Der mittlere Abstand beträgt gemäß Satz 6.2

$$\begin{aligned} \int_0^\infty r\cdot h_1(r)\,dr &= \int_0^\infty r\cdot e^{-\frac{r^2}{2}}\,dr = -e^{-\frac{r^2}{2}}\Big|_0^\infty + \int_0^\infty e^{-\frac{r^2}{2}}\,dr \\ &= \sqrt{2\pi}\int_0^\infty \varphi(r)\,dr \overset{sym.\varphi}{=} \sqrt{\frac{\pi}{2}}\cdot\int_{-\infty}^\infty \varphi(r)\,dr = \sqrt{\frac{\pi}{2}} \end{aligned}$$

Satz & Definition 6.18 (Faltung)**.** *Seien X, Y stochastisch unabhängige, $\mathbb{R}$-wertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$. Dann heißt die Verteilung P^{X+Y} von $X+Y$ die Faltung von P^X und P^Y. In Zeichen $P^X * P^Y$.*

Besitzen X und Y Dichten f_x und f_y, so besitzt $P^X * P^Y$ die Dichte

$$(f_X * f_y)(t) := \int_{\mathbb{R}} f_x(x)\cdot f_y(t-x)\,dx \qquad ,t\in\mathbb{R}$$

Beweis: Nach Satz 6.14 hat (X, Y) die Dichte

$$f(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall (x,y) \in \mathbb{R}^2$$

Die Abbildung $g(x,y) = (x+y, x-y)$ erfüllt die Bedingung von Satz 6.16 mit

$$g^{-1}(u,v) = (\frac{u+x}{2}, \frac{u-v}{2})$$

und

$$Dg^{-1}(u,v) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$
$$\Rightarrow |\det Dg^{-1}(u,v)| = \frac{1}{2}$$

Daher hat $g(X,Y) = (X+Y, X-Y)$ die Dichte

$$\begin{aligned} h(u,v) &= |\det Dg^{-1}(u,v)| \cdot f(y^{-1}(u,v)) \\ &= \frac{1}{2} f_X(\frac{u+v}{2}) \cdot f_Y(\frac{u+v}{2}) \end{aligned}$$

Nach Satz 6.15 hat schließlich $X+Y$ die Dichte

$$\begin{aligned} f_{X+Y}(u) = \int\limits_{-\infty}^{\infty} h(u,v)\, dv &= \frac{1}{2} \int\limits_{\infty}^{\infty} f_X(\frac{u+v}{2}) \cdot f_Y(\frac{u-v}{2})\, dv \\ &\underset{\substack{x=\frac{u+v}{2} \\ dv=2dx}}{=} \int\limits_{-\infty}^{\infty} f_X(x) \cdot f_Y(u-x)\, dx \end{aligned}$$

Beispiel 6.19. *Sei X $\mathcal{N}_{(0,1)}$-verteilt, d.h. mit Dichte $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.*
Dann hat $Y = \sigma X + \mu \quad (\mu \in \mathbb{R}, \sigma \neq 0)$ nach Satz 6.16 die Dichte

$$\varphi_{(\mu,\sigma^2)}(y) = \varphi\left(\frac{y-\mu}{\sigma}\right) \cdot \frac{1}{|\sigma|} = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R}$$

Die zugehörige Verteilung P^Y heißt Normalverteilung mit Mittelwert μ und Varianz σ^2; in Zeichen $\mathcal{N}_{(\mu,\sigma^2)}$.

Sei nun Y $\mathcal{N}_{(\mu,\sigma^2)}$-verteilt.
Dann gilt $aY+b = a\cdot(\sigma X+\mu)+b = (a\sigma)X+(a\cdot\mu+b)$ ist $\mathcal{N}_{(a\mu+b,a^2\sigma^2)}$-verteilt, falls $a \neq 0$.

Sind Y_i $\mathcal{N}_{(\mu_i,\sigma_i^2)}$-verteilt $(i = 1,2)$ und stochastisch unabhängig
$\Rightarrow Z_i := Y_i - \mu_i$ sind $\mathcal{N}_{(0,\sigma^2)}$-verteilt und stochastisch unabhängig und $Z_1 + Z_2$ hat gemäß

Satz 6.19 die Dichte

$$
\begin{aligned}
g(z) &= \int\limits_{-\infty}^{\infty} \varphi_{(0,\sigma_1^2)}(x) \cdot \varphi_{(0,\sigma_2^2)}(z-x)\, dx \\
&= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \int\limits_{-\infty}^{\infty} \underbrace{\exp\left(-\left(\frac{x^2}{2\sigma_1^2} + \frac{(z-x)^2}{2\sigma_2^2}\right)\right)}_{*} dx \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \int\limits_{-\infty}^{\infty} \varphi_{\left(\mu_2, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)}(x)\, dx \cdot \exp\left(-\frac{z^2}{2\sigma_2^2}\left(1 - \frac{1}{\frac{\sigma_2^2}{\sigma_1^2}+1}\right)\right) \\
&= \varphi_{(0,\sigma_1^2+\sigma_2^2)}(z)
\end{aligned}
$$

$$
\begin{aligned}
* : \exp\left(-\left(\frac{x^2}{2\sigma_1^2} + \frac{(z-x)^2}{2\sigma_2^2}\right)\right) &= \exp\left(-\frac{1}{2}\left(x^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) - 2\frac{2}{\sigma_2^2}x + \frac{z^2}{\sigma_2^2}\right)\right) \\
&= \exp\left(-\frac{1}{2}\left[\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)\left(x - \frac{\frac{z}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}\right)^2 + \frac{z^2}{\sigma_2^2} - \frac{\frac{z^2}{\sigma_2^4}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}\right]\right)
\end{aligned}
$$

d.h. $Z_1 + Z_2$ *ist* $\mathcal{N}_{(0,\sigma_1^2+\sigma_2^2)}$*-verteilt*
$\Rightarrow Y_1 + Y_2 = Z_1 + Z_2 + \mu_1 + \mu_2$ *ist* $\mathcal{N}_{(\mu_1+\mu_2,\sigma_1^2+\sigma_1^2)}$*-verteilt.*

Fazit:

$$
\mathcal{N}_{(\mu_1,\sigma_1^2)} * \mathcal{N}_{(\mu_2,\sigma_2^2)} = \mathcal{N}_{(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)}
$$

Beweis von 6.5: Hier nur für den Fall, dass X nur endlich viele Werte annimmt. Seien $X_0 = 0 < x_1 < x_2 < \cdots < x_n$ die positiven Werte, die X annehmen kann und seien $y_m < y_{m-1} < \cdots < y_1 < 0 = y_0$ die negativen Werte, die X annehmen kann.

$$E(X) = \sum_{x\in X(\Omega)} x \cdot P\{X = x\} = \sum_{i=1}^{n} x_i \cdot P\{X = x_i\} + \sum_{i=1}^{m} y_i \cdot P\{X = y_i\}$$

$$\begin{aligned}
\sum_{i=1}^{n} x_i \cdot P\{X = x_i\} &= \sum_{i=1}^{n}\sum_{l=1}^{i}(x_l - x_{l-1}) \cdot P\{X = x_i\} \\
&= \sum_{i=1}^{n} l = 1(x_l - x_{l-1}) \cdot \underbrace{\sum_{i=1}^{n} i = lP\{X = x_i\}}_{=P\{X\geq x_l\}=P\{X>x_{l-1}\}} \\
&= \sum_{i=1}^{n} l = 1 \int\limits_{x_{l-1}}^{x_l} \underbrace{P\{x > t\}}_{=P\{X>x_{x_l-1}\}}\ dt \\
&= \int\limits_0^\infty P\{X > t\} dt
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{m} y_i \cdot P\{X = y_i\} &= \sum_{i=1}^{m}\sum_{l=1}^{i}(y_l - y_{l-1} \cdot P\{X = y_i\} \\
&= \sum_{l=1}^{m}(y_l - y_{l-1}) \underbrace{\sum_{i=l}^{m} P\{X = y_i\}}_{=P\{X\leq y_l\}} \\
&= -\sum_{l=1}^{m} \int\limits_{y_l}^{y_{l-1}} P\{X \leq t\} dt \\
&= \sum_{\infty}^{0} P\{X \leq t\} dt
\end{aligned}$$

Analog definiert man den Mittelwert eines Wahrscheinlichkeitsmaßes Q auf $\mathbb{R}$

$$\mu(Q) = E_Q(Id) = \int\limits_0^\infty Q((t,\infty))dt - \int\limits_{-\infty}^0 Q((-\infty,t])dt$$

Die Sätze und Definitionen 4.3, 4.5, 4.9 (bis auf die Summendarstellung), 4.10, 4.12, 4.14-4.17, 4.19, 5.1-5.3 und 5.6 übertragen sich sinngemäß.

Zum Beispiel die Verallgemeinerung von Satz 4.3 (Transformationssatz)

Satz 6.20. *Ist $(\Omega, \mathcal{A}, P)$ eine Wahrscheinlichkeitsraum, $X : (\Omega, \mathcal{A}) \to (\tilde{\Omega}, \tilde{\mathcal{A}})$ eine Zufallsvariable und $g : (\tilde{\Omega}, \tilde{\mathcal{A}}) \to (\mathbb{R}, \mathbb{B})$ eine Abbildung, so gilt*

$$E_P(g(X)) = E_{P^X}(g),$$

falls eine der beiden Seiten wohl definiert ist.

Beweis: Da $P^X(A) = P(X^{-1}(A))$:

$$\begin{aligned}
&\int_0^\infty P\underbrace{\{g(X) > t\}}_{=X^{-1}(\{g>t\})}dt = \int_0^\infty P^X(\{g > t\})dt \\
&- \int_{-\infty}^0 P\{g(X) \leq t\}dt = - \int_{-\infty}^0 P^X(\{g \leq t\})dt \\
\hline
&\overset{6.6}{=} E_P g(x) \qquad \overset{6.6}{=} E_{P^X}(g)
\end{aligned}$$

$\Rightarrow$ *Behauptung*

6.7: Sei X eine $\mathbb{R}$-wertige Zufallsvariable mit Dichte f
$\Rightarrow E(X) = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx$

Verallgemeinerung:

Satz 6.21. *Besitzt $X = (X_1, ..., X_n) : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^n, \mathbb{B}^n)$ eine gemeinsame Dichte f und ist $g : (\mathbb{R}^n, \mathbb{B}^n) \to (\mathbb{R}, \mathbb{B})$ eine Abbildung, so gilt*

$$E(g(X)) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(x_1, \ldots, x_n) \cdot f(x_1, \ldots, x_n)dx_n \cdots dx_1$$

Beweis:

$$\begin{aligned}
E(g(X)) = E_P(g(X)) &\overset{6.20}{=} E_{P^X}(Y) \\
&= \int_0^\infty P^X\{g > t\}dt - \int_{-\infty}^0 P^X\{g \leq t\}dt
\end{aligned}$$

Dabei gilt

$$\begin{aligned}
\int_0^\infty P^X\{g>t\}dt &= \int_0^\infty \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} 1_{\{g(x_1,\dots,x_n)>t\}} f(x_1,\dots,x_n)dx_n \cdots dx_1 \, dt \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_0^\infty 1_{[0,g(x_1,\dots,x_n]}(t)dt \, f(x_1,\dots,x_n)dx_n \cdots dx_1 \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(x_1,\dots,x_n) \cdot 1_{g(x_1,\dots,x_n)\geq 0} f(x_1,\dots,x_n)dx_n \cdots dx_1
\end{aligned}$$

Ebenso

$$\int_{-\infty}^0 P^X\{g \leq t\}dt = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} -g(x_1,\dots,x_n) \cdot 1_{\{g(x_1,\dots,x_n)<0\}} f(x_1,\dots,x_n)dx_n \cdots dx_1$$

Zusammen: $E(g(X)) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(x_1,\dots,x_n) \cdot f(x_1,\dots,x_n)dx_n \cdots dx_1$

Beispiel 6.22. *Sei X $\mathcal{N}_{(0,1)}$-verteilt.*
Dann ist $Y = \sigma X + \mu$ $\mathcal{N}_{(\mu,\sigma^2)}$-Verteilt ($\rightarrow$ Bsp 6.19)
(mit Dichte $x \mapsto \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$, $\sigma > 0, \mu \in \mathbb{R}$)

$$\begin{aligned}
E(Y) &= \sigma \cdot E(X) + \mu = \mu \\
Var(Y) &= Var(\sigma X) = \sigma^2 Var(X) \\
&= \sigma^2 \cdot (E(X^2) - (E(X))^2) = \sigma^2 \cdot E(X^2) \\
&\overset{2.61}{=} \sigma^2 \cdot \int_{-\infty}^\infty x^2 \cdot \varphi(x)dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^\infty x \cdot x e^{-\frac{x^2}{2}} dx \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left[x \cdot (-e^{-\frac{x^2}{2}} \Big|_{-\infty}^\infty + \int_{-\infty}^\infty e^{-\frac{x^2}{2}} dx \right] \\
&= \sigma^2 \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \sigma^2 \int_{-\infty}^\infty \varphi(x)dx = \sigma^2
\end{aligned}$$

Dies erklärt den Namen von $\mathcal{N}_{(\mu,\sigma^2)}$: Normalverteilung mit Mittelwert μ und Varianz σ^2

Weitere Kenngrößen von $\mathbb{R}$-wertigen Zufallsvariablen und deren Verteilungen

Weiteres Maß für den "mittleren" Wert einer Zufallsvariable ist der Median.

Definition 6.23 (Median). *Sei X eine $\mathbb{R}$-wertige Zufallsvariable mit Verteilungsfunktion F.*
Dann heißt jeder Wert $m \in \mathbb{R}$, so dass

$$F(m) = P\{X \leq m\} \geq \frac{1}{2}\ ,\ 1 - F(m-) = P\{X \geq m\} \geq \frac{1}{2}$$

(wobei $F(m-) = \underset{x \uparrow m}{F}(x)$) Median von X bzw. von P^X.

Beispiel 6.24.

(i) Sei X exponentialverteilt mit Mittelwert $\lambda > 0$, d.h. X besitzt die Verteilungsfunktion

$$F_\lambda(x) = P\{X \leq x\} = 1 - e^{-\frac{x}{\lambda}}, x \geq 0$$

Dann gilt

$$F\lambda(m) = 1 - e^{-\frac{m}{\lambda}} \overset{!}{\geq} \frac{1}{2} \Leftrightarrow m \geq -\lambda \cdot \log 2 \approx 0,69 \cdot \lambda$$

$$1 - F_\lambda(m-) = e^{-\frac{m}{\lambda}} \overset{!}{\geq} \frac{1}{2} \Leftrightarrow, \leq \lambda \cdot \log 2$$

$$\Rightarrow \textit{es existiert genau ein Median, nämlich } m = \lambda \cdot \log 2 \approx 0,69 \cdot \lambda$$

$$\Rightarrow \frac{\textit{Median}}{\textit{Erwartungswert}} = \log 2 \approx 0,69$$

Anwendung: X: Zeit bis zum Zerfall eines radioaktiven Isotops. Dann ist Median = Halbwertszeit

(ii) Sei X gleichverteilt auf $\{-1, 1\}$, d.h. $P\{X = -1\} = \frac{1}{2} = P\{X = 1\}$.
Jedes $m \in [-1, 1]$ ist ein Median, denn

$$P\{X \leq m\} \geq P\{X \leq -1\} = \frac{1}{2}, P\{X \geq m\} \leq P\{X \geq 1\} = \frac{1}{2}$$

Es existiert stets wenigstens ein Median.
Der Media ist nicht eindeutig, wenn F auf einem echten Intervall den Wert $\frac{1}{2}$ annimmt.

Analog lässt sich zu jeder Wahrscheinlichkeit $p \in (0,1)$ ein "kleinster" Wert definieren, der wenigstens mit Wahrscheinlichkeit p unterschritten wird.

Definition 6.25 (Quantilfunktion)**.** *Zu einer Verteilungsfunktion F wird die zugehörige Quantilfunktion(verallgemeinerte Inverse) definiert als*

$$F^{\leftarrow} : [0,1] \to [-\infty, \infty]$$
$$F^{\leftarrow}(t) := \begin{cases} \inf\{x \in \mathbb{R} \mid F(x) \geq t\} & , t \in (0,1] \\ \sup\{x \in \mathbb{R} \mid F(x) = 0\} & , t = 0 \end{cases}$$

Dabei wird $\sup \emptyset := -\infty$, $\inf \emptyset := \infty$ definiert.
$F^{\leftarrow}(p)$ wird als p-Quantil bezeichnet.

——————— Zeichnung, um $F^{\leftarrow}$ zu verdeutlichen —————

Gibt es genau ein x mit $F(x) = p$, so gilt $F^{\leftarrow}(p) = x$.
Insbesondere ist $F^{\leftarrow}$ auf $(0,1)$ gleich der Umkehrfunktion von F, wenn F invertierbar ist.

Beispiel 6.26. *X sei exponentialverteilt mit mittelwert $\lambda > 0$, d.h. $F(x) = 1 - e^{-\frac{x}{\lambda}}$*

$$\Rightarrow F^{\leftarrow}(t) = \begin{cases} -\lambda \log(1-t) & , \quad t \in (0,1) \\ 0 & , \quad t = 0 \\ \infty & , \quad t = 1 \end{cases}$$

$F^{\leftarrow}(\frac{1}{2})$: Median

Interquartilsabstand $IQR := F^{\leftarrow}(\frac{3}{4}) - F^{\leftarrow}(\frac{1}{4})$
Maß für die Schwankungsbreite von F bzw. P^X
Hier $IQR = -\lambda \log \frac{1}{4} - (-\lambda \log \frac{3}{4}) = \lambda \log 3 \approx 1,10 \cdot \lambda$

Satz 6.27. *Ist U eine auf $(0,1)$ gleichverteilte Zufallsvariable und F eine beliebige Verteilungsfunktion, so besitzt die Zufallsvariable $X = F^{\leftarrow}(U)$ die Verteilungsfunktion F.*

Beweis: Wir zeigen zunächst

$$F^{\leftarrow}(t) \leq x \Leftrightarrow t \leq F(x) \qquad \forall t \in (0,1)$$

"$\Leftarrow$":

$$F^{\leftarrow}(t) = inf \underbrace{\{x \in \mathbb{R} \mid F(y) \geq t\}}_{x \in \{x \in \mathbb{R} | F(y) \geq t\}, \text{ wenn } t \leq t}$$
$$\Rightarrow F^{\leftarrow}(t) \leq x$$

"$\Rightarrow$": Sei $x \geq F^{\leftarrow}(t)$

Falls $x > F^{\leftarrow}(t)$, dann gilt wegen der Definition von $F^{\leftarrow}$ und Monotonie von F sicherlich $F(x) > t$

Falls $x = F^{\leftarrow}(t)$, dann gilt wegen der rechtsseiten Stetigkeit von F

$$F(x) = \lim_{y \downarrow x} \underbrace{F(y)}_{\geq t,\text{ da } y > F^{\leftarrow}(t)} \geq t$$

Anwendung: Wenn u eine Pseudozufallszahl gemäß Gleichverteilung auf $(0,1)$ ist, so ist $F^{\leftarrow}(u)$ eine Pseudozufallszahl gemäß der Verteilungsfunktion F.

7 Bediensysteme und Netzwerke

Bediensystem: Ein oder mehrere Bediener (Server) fertigen ankommende Kunden (Aufträge) nach vorgegebenem Schema ab.
Zufällig: Die Ankunft der Kunden, die Bedienzeit, ...

Klassifikation von Bediensystemen: $A|B|s|c|R$

A: Art des Ankunftprozesses
B: Art des Bedienprozesses
} dabei insbesondere
M: Markov'sche Prozesse
D: deterministische Prozesse
G: allgemeine Prozesse

s: Anzahl der Bediener (s: server)
c: Anzahl der Warteplätze (c: capacity)
R: Reihenfolge der Bedienung {
FCFS: first come first serve
LCFS: last come first serve
SIRO: service in random order

——— Skizzen ———

Falls $c = \infty$ oder R=FCFS, werden diese Angaben in der Regel weggelassen.

Beispiel 7.1. *Das Bediensystem $M|M|1$ ($= M|M|1|\infty|FCFS$)*

Ankunftsprozess: *Wir diskretisieren die Zeit in Zeittakte der Länge h (h "klein")*
Kunden kommen unabhängig von einander mit der mittleren Rate $\lambda > 0$ an, d.h. die erwartete Anzahl von Kunden, die in <u>einem</u> Zeittakt ankommen, ist gleich $\lambda \cdot h << 1$.
Es wird vereinfachend angenommen, dass maximal ein Kunde pro Takt ankommt (realistisch, falls h hinreichend klein).

$$Z_i := \begin{cases} 1 & \text{, falls ein Kunde ankommt im i-ten Takt} \\ 0 & \text{, falls kein Kunde ankommt im i-ten Takt} \end{cases}$$

Z_i ; $i \in \mathbb{N}$ sind unabhängig und $\mathcal{B}_{(1,\lambda\cdot h)}$-verteilt.

Die Anzahl der Takte, bis der 1. Kunde eintrifft, ist $\mathcal{G}_{\{\lambda h\}}$-verteilt ($\rightarrow$ Bsp. 1.11(i)), wenn man nur die Takte ohne Kundenankunft zählt, d.h. die Wahrscheinlichkeit, dass genau k Takte kein Kunde ankommt ist

$$\mathcal{G}_{\lambda h}\{k\} = (1 - \lambda h)^k \cdot \lambda h$$

Für ein beliebiges $n \in \mathbb{N}$ ist die Anzahl der Takte bis zur 1. Ankunft eines Kunden nach n ebenfalls $\mathcal{G}_{\lambda h}$-verteilt.
Dies gilt auch für zufällige Zeitpunkte N, falls $\{N = n\}$ nur von $Z_1, \ldots, Z_n$ abhängt (und unabhängig von Z_i für $i > n$) für alle $n \in \mathbb{N}$.

Bedienprozess: Der einzige Server bedient Kunden in der Reihenfolge der Ankunft. Die Bedienzeiten (in Anzahl von Takten) seien unabhängig von einander und analog zum Ankunftsprozess $\mathcal{G}_{\mu h}$-verteilt; genauer: Anzahl der Takte zwischen 2 Takten, zu denen Kunden fertig bedient werden, sei $\mathcal{G}_{\mu h}$-verteilt für eine Bedienrate $\mu > 0$.
$\rightsquigarrow Y_i, i \in \mathbb{N}$ i.i.d $\mathcal{B}_{(1,\mu h)}$-verteilt, wobei $\mu h << 1$ (unabhängig von Z_i)

Y_i seind (fiktive) Zeiten, zu denen Bedienung fertig wird, falls ein Kunde im System ist.

Zustandsprozess: X_n = Anzahl der Kunden im System im Takt n, d.h. zur Zeit $n \cdot h$
Dann ist $(X_n)_{n\in\mathbb{N}}$ ein homogener Markov-Prozess (mit Zustandsraum $\mathbb{N}_0$) mit Übergangswahrscheinlichkeiten

$$q_{ji} := P(X_{n+1} = j \mid X_n = i) = \begin{cases} 0 & , |i-j| > 1 \\ P\{Z_n = 1, Y_n = 0\} = P\{Z_n = 1\} \cdot P\{Y_n = 0\} & , j = i+1 \\ = \lambda h \cdot (1 - \mu h) & \\ P\{Y_n = 1, Z_n = 0\} = \lambda h \cdot (1 - \mu h) & , j = i-1 \\ P\{Y_n = 0, Z_n = 0\} + P\{Y_n = 1, Z_n = 1\} & , j = i \neq 0 \\ = (1-\lambda h)(1-\mu h) + \lambda \cdot h \cdot \mu \cdot h & \\ P\{Z_n = 0\} = 1 - \lambda h & , j = i = 0 \end{cases}$$

$\lambda \cdot h, \mu \cdot h << 1 \Rightarrow \lambda \cdot h \cdot \mu \cdot h << \lambda h, \mu h$
$\rightsquigarrow$ vernachlässige Terme $\lambda \cdot h \cdot \mu \cdot h$
$\rightsquigarrow$ approximative "Übergangsmatrix"

$$\widetilde{Q} = \begin{pmatrix} 1-\lambda h & \mu h & 0 & 0 & \ddots \\ \lambda h & 1-\lambda h-\mu h & \mu h & \ddots & \\ 0 & \lambda h & 1-\lambda h-\mu h & \ddots & \\ 0 & 0 & \lambda h & \ddots & \end{pmatrix}$$

$\widetilde{Q}$ ist eine stochastische "Matrix", d.h. Spaltensumme 1.

Im folgenden sei $(X_n)_{n\in\mathbb{N}}$ eine Markovkette mit der "Übergangsmatrix" $\widetilde{Q}$.
Gesucht: Gleichgewichtszustand, d.h. stationäre Verteilung, d.h. das Wahrscheinlichkeitsmaß P_s auf dem Zustandsraum $\mathbb{N}_0$, so dass gilt

$$P^{X_n} = P_s \Rightarrow P^{X_{n+1}} = P_s \ (\Rightarrow P^{X_m} = P_s \ \ \forall m \geq n)$$

Für alle $k \in \mathbb{N}_0$ muss dann gelten

$$
\begin{aligned}
P_s\{0,\dots,k\} = \quad & \underbrace{P\{0 \le X_n \le k\}}_{\substack{=P\{X_n \le k, X_{n+1} \le k\} \\ + P\{X_n \le k, X_{n+1} > k\}}} \quad \overset{!}{=} \quad \underbrace{P\{0 \le X_{n+1} \le k\}}_{\substack{=P\{X_n \le k, X_{n+1} \le k\} \\ + P\{X_n > k, X_{n+1} \le k\}}} \\
\Leftrightarrow & P\{X_n \le k, X_{n+1} > k\} = P\{X_n \ge k, X_{n+1} \le k\} \\
\Leftrightarrow & P\{X_n = k, X_{n+1} = k+1\} = P\{X_n = k+1, X_{n+1} = k\} \\
\Leftrightarrow & P(X_n = k \mid X_n = k+1) \cdot P\{X_n = k+1\} = P(X_{n+1} = k \mid X_n = k+1) \cdot P\{X_n = k+1\} \\
\Leftrightarrow & \dots \\
\Leftrightarrow & \lambda h \cdot P_s\{k\} = \mu h \cdot P_s\{k+1\} \\
\Rightarrow & P_s\{k+1\} = \frac{\lambda}{\mu} \cdot P_s\{k\} \qquad \forall k \in \mathbb{N}_0
\end{aligned}
$$

Induktiv folgt

$$
\begin{aligned}
P_s\{1\} &= \frac{\lambda}{\mu} P_s\{0\} \\
P_s\{2\} &= \frac{\lambda}{\mu} P_s\{1\} = (\frac{\lambda}{\mu})^2 P_s\{0\} \\
P_s\{k\} &= (\frac{\lambda}{\mu})^k P_s\{0\} \qquad \forall k \in \mathbb{N}_0
\end{aligned}
$$

P_s ist ein Wahrscheinlichkeitsmaß auf $\mathbb{N}_0$

$$
\begin{aligned}
\Rightarrow 1 = \sum_{k=0}^{\infty} P_s\{k\} &= P_s\{0\} \sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k \\
&\underset{\frac{\lambda}{\mu}<1}{=} P_s\{0\} \cdot \frac{1}{1-\frac{\lambda}{\mu}}
\end{aligned}
$$

$\Rightarrow$ *einzige mögliche stationäre Verteilung ist gegeben durch*

$$P_S\{k\} = (1-\frac{\lambda}{\mu}) \cdot (\frac{\lambda}{\mu})^k \quad , k \in \mathbb{N}_0 = \mathcal{G}_{1-\frac{\lambda}{\mu}}\{k\} \text{ falls } \frac{\lambda}{\mu} < 1$$

Für jeden festen Zeitpunkt ist die Wahrscheinlichkeit, dass sich genau k Kunden im System befinden gleich $P_s\{k\} = \mathcal{G}_{1-\frac{\lambda}{\mu}}\{k\}$.

Dies gilt auch für einen Zeitpunkt zu dem ein Kunde das System betritt bzw. für den Zustand, den ein Kunde bei Verlassen des Systems hinterlässt, d.h.

$$
\begin{aligned}
P(X_n = k \mid X_{n+1} = X_n + 1) &= P_s\{k\} \qquad \forall k \in \mathbb{N}_0, n \in \mathbb{N} \\
P(X_n = k \mid X_{n+1} = X_n - 1) &= P_s\{k\} \qquad \forall k \in \mathbb{N}_0, n \in \mathbb{N}
\end{aligned}
$$

Es gilt nämlich z.B.

$$
\begin{aligned}
P(X_n = k \mid X_n = X_{n-1} - 1) &= \frac{P\{X_n = k, X_n = X_n = X_{n-1} - 1\}}{P\{X_n = X_{n-1} - 1\}} \\
&= \frac{P\{X_n = k, X_{n-1} = k - 1\}}{\sum\limits_{j=0}^{\infty} P\{X_n = j, X_{n-1} = j + 1\}} \\
&= \frac{P(X_n > k \mid X_{n-1} = k + 1) \cdot P\{X_{n-1} = k + 1\}}{\sum\limits_{j=0}^{\infty} P(X_n = j \mid X_{n-1} = j + 1) \cdot P\{X_{n-1} = j + 1\}} \\
&= \frac{\mu h \cdot P_s\{k + 1\}}{\sum\limits_{j=0}^{\infty} \mu h \cdot P_s\{j + 1\}} = \frac{(1 - \frac{\mu}{\lambda}) \cdot (\frac{\lambda}{\mu})^{k+1}}{\sum\limits_{j=0}^{\infty} (1 - \lambda\mu) \cdot (\frac{\lambda}{\mu})^{k+1}} \\
&= \frac{(\frac{\lambda}{\mu})^k}{\frac{1}{1-\frac{\lambda}{\mu}}} = (1 - \frac{\lambda}{\mu}) \cdot (\frac{\lambda}{\mu})^k = P_s\{k\}
\end{aligned}
$$

$$\textit{Auslastung } \varrho \textit{ des Systems} := \frac{\textit{mittlere nachgefragte Bedienzeit (pro Takt)}}{\textit{maximal mögliche Bedienzeit (pro Takt)}} = \frac{\lambda \cdot \frac{1}{\mu}}{1} = \frac{\lambda}{\mu}$$

(denn im Mittel kommen pro Zeitinheit λ Kunden an, die jeweils im Mittel $\frac{1}{\mu}$ an Bedienzeit in Anspruch nehmen)

Der Gleichgewichtszustand ist möglich g.d.w. $\varrho < 1$ (gilt auch für allgemeine Bediensysteme).

Viele Kenngrößen des Systems hängen von λ nur über ϱ ab.
z.B. die mittlere Kundenzahl im System

$$
\begin{aligned}
E(X_n) = \sum_{k=0}^{\infty} k \cdot P\{X_n = k\} &= \sum_{k=0}^{\infty} k \cdot (1 - \varrho) \cdot \varrho^k = (1 - \varrho) \cdot \varrho \cdot \sum_{k=0}^{\infty} k \cdot \varrho^{k-1} \\
&= (1 - \varrho) \cdot \varrho \cdot \sum_{k=0}^{\infty} \frac{\partial}{\partial \varrho} \varrho^k = (1 - \varrho) \cdot \varrho \frac{\partial}{\partial \varrho} \sum_{k=0}^{\infty} \varrho^k \\
&= (1 - \varrho) \cdot \varrho \frac{\partial}{\partial \varrho} \frac{1}{1 - \varrho} = (1 - \varrho) \cdot \varrho \cdot \frac{1}{(1 - \varrho)^2} = \frac{\varrho}{1 - \varrho}
\end{aligned}
$$

Analog $Var(X_n) = \frac{\varrho}{(1-\varrho)^2}$

Für $\varrho \uparrow 1$ wachsen $E(X_n)$ und $Var(X_n)$ schnell an.
Z.B. wächst $E(X_n)$ um mehr als den Faktor, wenn ϱ von $0,9$ auf $0,95$ erhöht wird.

Weitere Leistungsgröße: mittlere Wartezeit eines Kunden, genauer $E(W)$, wenn W zufällige Zeit ist, die ein Kunde im System zubringt.
Findet der Kunde k Kunden im System vor, die alle im Mittel $\frac{1}{\mu}$ als Bedienzeit benötigen, so verlässt er im MIttel nach $(k+1)\cdot\frac{1}{\mu}$ das System.

$$\Rightarrow E(W) = \sum_{k=0}^{\infty}(k+1)\cdot\frac{1}{\mu}\cdot P_s\{k\} = \frac{1}{\mu}\Bigg(\underbrace{\sum_{k=0}^{\infty}k\cdot P_s\{k\}}_{E(X_n)} + \underbrace{\sum_{k=0}^{\infty}P_s\{k\}}_{=1}\Bigg)$$

$$= \frac{1}{\mu}\cdot(\frac{\varrho}{1-\varrho}+1) = \frac{1}{\mu(1-\varrho)} = \frac{1}{\mu-\lambda}$$

Formel von Little: $E(W) = \frac{E(X_n)}{\overline{\lambda}}$ *(gilt in viel allgemeineren Bediensystemen)*
wobei $\overline{\lambda}$ die mittlere Ankunftsrate ist (hier: $\overline{\lambda}=\lambda$).

Beispiel 7.2. *$|M|M|s|c$-Systeme*
s: Anzahl der Server
c: Anzahl der Warteplätze

X_n: Anzahl der Kunden im System, Zustandsraum: $\{0,\ldots,s+c\}$

Im Fall $c<\infty$ und $s<\infty$ werden Kunden abgewiesen, wenn sich bereits $s+c$ Kunden im System befinden.

$\lambda_k := \begin{cases} \lambda & 0\le k < c+s \\ 0 & k \ge c+s \end{cases}$ *Ankunftsrate, wenn bereits k Kunden im System sind.*

s Server fertigen jeweils unabhängig voneinander mit Wahrscheinlichkeit μh genau einen Kunden in einem Takt ab und mit Wahrscheinlichkeit $1-\mu h$ keinen Kunden.
Befinden sich k Kunden im System, d-h- $\min(k,s)$ Kunden werden gerade bedient, so ist die Zahl der Kunden, die in einem Takt fertig bedient werden, $\mathcal{B}_{(\min(k,s),\mu h))}$-verteilt.

$$\mathcal{B}_{(\min(k,s),\mu h))}\{l\} = \binom{\min(s,k)}{l}\cdot(\mu h)^l\cdot(1-\mu h)^{\min(s,k)-l}$$
$$\underset{\mu h<<1}{\approx} \begin{cases} 1-\min(sk)\mu h & l=0 \\ \min(s,k)\mu h & l=1 \\ 0 & l=2 \end{cases}$$

z.B.

$$(1-\mu h)^{\min(s,k)} = \sum_{i=0}^{\infty}\binom{\min(s,k)}{i}1^i\cdot(-\mu h)^{\min(s,k)-i}$$
$$= 1-\min(s,k)\cdot\mu h + \textit{Terme mit Potenzen von } h \textit{ mit Exponenten } \ge 2$$

$$\mu_k := \begin{cases} k \cdot h & k \leq s \\ s \cdot h & k > s \end{cases} = \min(s,k) \cdot h$$

$\rightsquigarrow$ *approximative "Übergangsmatrix"*

$$\widetilde{Q} = \begin{pmatrix} 1-\lambda_0 \cdot h & \mu_1 h & 0 & \cdots \\ \lambda_0 h & 1-\lambda_1 h - \mu_1 h & \mu_2 h & \cdots \\ 0 & \lambda_2 h & 1-\lambda_2 h-\mu_2 h & \cdots \\ 0 & 0 & \lambda_2 h & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Im fall $s + c < \infty$ ist die letze Zeile $(0 \quad \ldots \quad 0 \quad \lambda_{s+c-1}h \quad 1-\mu_{s+c}h)$.

Zur Bestimmung einer Stationären Verteilung P_s wie in 7.1 erhält man als notwendige Bedingung

$$\begin{aligned} &\mu_{k+1} h P_s\{k+1\} = \tilde{q}_{k,k+1} P_s\{k+1\} = \tilde{q}_{k+1,k} P_s\{k\} = \lambda_k h \cdot P_s\{k\} \\ &\Rightarrow P_s\{k+1\} = \frac{\lambda_k}{\mu_k+1} P_s\{k\} \\ &\overset{induktiv}{\Rightarrow} P_s\{k\} = \frac{\lambda_0}{\mu_1} \cdot \frac{\lambda_1}{\mu_2} \cdot \ \cdots \ \cdot \frac{\lambda_{k-1}}{\mu_k} P_s\{0\} \qquad \forall n \in \mathbb{N}_0 \end{aligned}$$

Z.B. im Fall: $s < \infty, c = 0$

$$\begin{aligned} &\Rightarrow \lambda_i = \begin{cases} \lambda & c < s \\ 0 & sonst \end{cases} \\ &\mu_i = i\mu \qquad \forall 0 \leq i \leq s \\ &\Rightarrow P_s\{k\} = \frac{\lambda^k}{k! \mu^k} \cdot P_s\{0\} \qquad 0 \leq k \leq s \\ &\sum_{k=0}^{s} P_s\{k\} \overset{!}{=} 1 \Rightarrow P_s\{0\} = \frac{1}{\sum_{k=0}^{s} \frac{(\frac{\lambda}{\mu})^k}{k!}} \end{aligned}$$

Es existiert <u>immer</u> eine so definierte stationäre Verteilung.

Die mittlere Ankunftsrate im System ist $\overline{\lambda} = \lambda \cdot (\underbrace{1 - P_s\{s\}}_{*}$

**: Wahrscheinlichkeit, dass das System noch Kunden aufnehmen kann.*

Auslastung: $\varrho = \dfrac{\lambda \cdot \frac{1}{\mu}}{s}$
Zähler: mittlere Zahl Kunden pro Zeiteinheit · mittlere Bedienzeit pro Kunde
Nenner: maximale Bedienzeit aller s Server pro Zeiteinheit pro Kunde

<u>mittlere Verbleibsdauer</u>: $E(W) = \frac{1}{\mu}$
(da vor Bedienung keine Wartezeit anfallen kann)

Formel von Little: $E(W) = \dfrac{E(X_n)}{\overline{\lambda}}$
$\Rightarrow E(X_n) = \overline{\lambda} \cdot E(W) = \lambda \cdot (1 - P_s\{s\} \cdot \frac{1}{\mu} = \varrho \cdot s$

Einfachstes Modell eines Netzwerkes:

Beispiel 7.3. *Zwei hintereinander geschaltete $M|M|1$-Systeme*
Skitze:

$$\underbrace{\textit{Ankunftsrate } \lambda \rightarrow \overline{\underline{\cdots|\;\;|\;\;|\;\;|}} \rightarrow \underset{\textit{Server 1}}{\overset{\textit{Bedienrate}\mu_1}{\overline{\underline{|\;\;|}}}}}_{\textit{Teil1}} \underbrace{\rightarrow\rightarrow \overline{\underline{\cdots|\;\;|\;\;|\;\;|}} \rightarrow \underset{\textit{Server 2}}{\overset{\textit{Bedienrate}\mu_2}{\overline{\underline{|\;\;|}}}} \rightarrow \textit{Abgangsprozess}}_{\textit{Teil2}}$$

Markovkette $(X_{n,1}, X_{n_2})_{n\in\mathbb{N}_0}$
$X_{n,i}$ *= Anzahl Kunden im Teil i z.Z nh,* $\mathbb{N}_0$*-wertig*
Die (approximativen) Übergangswahrscheinlichkeiten sind
$j_1 \hat{=}$ *Anzahl Kunden im ersten Teil z.Z.* $(n+1)h$
$j_2 \hat{=}$ *Anzahl Kunden im zweiten Teil z.Z.* $(n+1)h$
i_1, i_2 *analog z.Z. nh*

$$\begin{aligned}
\tilde{q}_{(j_1,j_2),(i_1,i2)} &= P(X_{n+1,1} = j_1, X_{n+1,2} = j_2 \mid X_{n,1} = i_1, X_{n,2} = i_2) \\
&= \begin{cases} \lambda h & j_1 = i_1 + 1, j_2 = i_2 \\ \mu_1 h & j_1 = i_1 - 1, j_2 = i_2 + 1 \\ \mu_2 h & j_1 = i_1, j_2 = i_2 - 1 \\ 1 - \lambda h - \mu_1 h - \mu_2 h & j_1 = i_1, j_2 = i_2 \end{cases} \quad \textit{falls } i_1, i_2 > 0
\end{aligned}$$

(sonst z.B. im Fall $i_1 = 0$ *enfallen die Terme* $\mu_1 h$*)*

Wann existiert ein Gleichgewichtszustand d.h. eine stationäre Verteilung?
Wie sieht sie gegebenenfalls aus?

Ist das Gesamtsystem im Gleichgewicht, so ist insbesondere Teil 1 auch im Gleichgewicht..
Daher ist notwendigerweise $\lambda < \mu_1$.
Dann muss (anschaulich gesprochen) der Prozess der Kunden, die Teil 1 verlassen, die

gleiche Verteilung haben, wie der Ankunftsprozess.
Da der Abgangsprozess von Teil 1 zugleich der Zugangsprozess von Teil 2 ist, ist Teil 2 genau vom gleichen Typ wie Teil 1, nur mit evtl. anderer Bedienrate μ_2.
Da die zukünftige Entwicklung des ersten Teils unabhängig vom bisherigen Abgangsprozesses ist ("Teil 1 vergisst abgehende Kunden sofort"), sind im Gleichgewichtszustand $X_{n,1}, X_{n_2}$ unabhängig voneinander.
Da nach 7.1 im Gleichgewichtszustand $P^{X_{n,1}} = \mathcal{G}_{1-\frac{\lambda}{\mu_1}}$ und $P^{X_{n,2}} = \mathcal{G}_{1-\frac{\lambda}{\mu_2}}$ folgt für die Verteilung des Gesamtsystems

$$\begin{aligned} P_s\{(k_1,k_2)\} &= P^{X_{n,1}} = \mathcal{G}_{1-\frac{\lambda}{\mu_1}}\{k_1\} P^{X_{n,2}} = \mathcal{G}_{1-\frac{\lambda}{\mu_2}}\{k_2\} \\ &= (\frac{\lambda}{\mu_1})^{k_1}(1-\frac{\lambda}{\mu_1}) \cdot (\frac{\lambda}{\mu_2})^{k_1}(1-\frac{\lambda}{\mu_2}) \end{aligned}$$

falls $\lambda < \mu_1, \lambda < \mu_2$, sonst existiert keine stationäre Verteilung.

Nicht-lineare Netzwerke sind sehr komplex zu analysieren.
Z.B. ——— Skitze eines etwas komplexeren Netzwerkes ———-

Zeitstetige Modelle als Grenzfall $h \to 0$

Im Fall $h > 0$: Zeit genessen in Anzahl der Takten zwischen zwei Ankunften von Kunden sind unabängig $\mathcal{G}_{\lambda h}$-Verteilt.
Nach Hausaufgabenblatt 8, Aufgabe 2(ii) konvergiert die Verteilung der Zeit (= Anzahl Takte $\cdot h$) gegen die Exponentialverteilung mit dem mittelwert $\frac{1}{\lambda}$.

$\rightsquigarrow$ Zeit stetiges Modell: Dann ist die Zeit zwischen 2 Aufträgen von Kunden unabhängig voneinander, exponential Verteilt mit Mittelwert $\frac{1}{\lambda}$.

N_t = Anzahl an Kunden, die bis z-Z. $t \geq 0$ angekommen sind. Poissoinscher Grenzwertsatz zeigt: N_t ist Poisson-Verteilt mit Parameter λ.
$(N_t)_{t\geq 0}$ (homogener) Poisson-Prozess mit Rate λ.
Stationäre Verteilung: $\mathcal{G}_{1-\frac{\lambda}{\mu}}$
(da stationäre Verteilung im Zeitdiskreten Modell gar nicht von h abhängt)

8 Grundkonzepte der Mathematischen Statistik

Gesammelte Daten werden interpretiert als Relation von Zufallsvariablen.:
Gegeben: Daten $x_1, \dots, x_n$
Interpretaion: $x_i = X_i(\omega), 1 \le i \le n$ für die Zufallsvariable $X_i : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathbb{B})$
Gemeinsame Verteilung von $X_1, \dots, X_n$ sei teilweise Unbekannt.
Ziel: Gewinne informationen über diese gemeinsame Verteilung also z.B.

- unbekannte Parameter (z.B. Erwartungswerte) schätzen oder
- vorgegebene Hypothesen über die Verteilung überprüfen (test)

Beispiel 8.1. *In Bediensystemen werden Ankunftszeiten der ersten k Kunden beobachtet.*
$\rightsquigarrow$ *Zwischenankunftszeiten $t_1, \dots, t_k$ (t_i: Ankunftszeit i-ter Kunde - Ankunftszeit (i-1)-ter Kunde)*

Interpretation: $t_i = T_i(\omega)$ für die ZUfallsvariable $T_i : (\Omega, \mathcal{A}, P) \to ((0, \infty), \mathcal{B}_{(0,\infty)}$
Mögliche Fragestellungen:

Schätz-Probleme
- *Wie groß ist die erwartete Zeit $E(T_i)$, die zwischen zwei Kundenankünften vergeht, falls diese für alle i gleich ist.*
- *Was ist die Verteilungsfunktion von T_i, wenn alle T_i die selbe Verteilung haben?*

Test-Probleme
- *Sind die T_i exponential Verteilt?*
- *Haben die T_i alle dieselbe Verteilung?*

In allen Fällen müssen die genauen Modellannahmen spezifiziert werden.

Diese können bei den ersten drei Fragen z.B. lauten:

(*) Die Zufallsvariable $T_1, \dots, T_k$ sind unabhängig voneinander und alle haben dieselbe Verteilung mit $Var(T_i) < \infty$

Insbesondere bei der ersten Frage kann man aber auch sinvolle Schätzungen unter schwächeren Modellannahmen herleiten, z.B.

Die Zufallsvariablen $T_1, \dots, T_k$ sind unkorreliert mit $E(T_i) = E(T_1),\ Var(T_i) \leq M \quad \forall 1 \leq i \leq k$.
Dann liegt für ein großes k nach dem Gesetz der Großen Zahlen der Folgende Schätze vor

$$\hat{\mu} := \frac{1}{n}\sum_{i=1}^{n} T_i$$

Evtl können aber auch starke Modellannahmen sinnvoll sein, z.B.

(**) Die T_i sind unabhängig und exponential verteilt mit unbekanntem Parameter μ

Bei Frage 4 wäre dagegen diese Modellannahme unsinnig.

Definition 8.2 (statistisches Experiment & Verteilungsannahme)**.** *Ein statistisches Experiment ist ein Tripel $(S, \mathcal{D}, (P_\vartheta)_{\vartheta\in\Theta})$ wobei $S \neq \emptyset$ der Stichprobenraum ist, $\mathcal{D}$ eine σ-Algebra auf S und P_ϑ ein Wahrscheinlichkeitsmaß auf $(S, \mathcal{D}) \quad \forall \vartheta \in \Theta$. $(P_\vartheta)_{\vartheta\in\Theta}$: Verteilungsannahme*

Interpretation:

S: Menge aller möglichen Gesamtbetrachtungen (z.B. Beobachtungsvektor $(t_1, \dots, t_k)$)

Anname: Die Gesamtbeobachtung wurde gemäß einer der Wahrscheinlichkeitsmaße P_ϑ erzeugt; der zugehörige Parameter ϑ heißt der wahre Parameter.

Beispiel 8.3. *Im Beispiel 8.1 ist $(S = (0,\infty)^n, \mathbb{B}(0,\infty)^n, (E^n_\lambda)_{\lambda>0})$ eine sinnvolle Wahl.*

1. Annahme: Die Zufallsvariablen T_i sind exponential verteilt.
Formal: $T_i^{(\lambda)} \backsim Exp(\frac{1}{\lambda}) \quad \lambda \in (0,\infty)$
also ist $F_\lambda(x) = 1 - e^{-\frac{x}{\lambda}}$ die Verteilungsfunktion von $T_i^{(\lambda)}$
oder äquivalent $\mathcal{G}_\lambda(x_1, \dots, x_n) = \prod_{i=1}^{n}(1 - e^{-\frac{x_n}{\lambda}})$ ist die Verteilungsfunktion von $(T_1^{(\lambda)}, \dots, T_n^{(\lambda)})$

2. Annahme: Die Zufallswahriablen sind i.i.d.
Formal: $T_1^{(F)} \backsim Q_F \quad F \in \mathcal{F}$ mit Q_F als Wahrscheinlichkeitsmaß mit Verteilungsfunktion $G_F = (x_1, \dots, x_n) = \prod_{i=1}^{n} F(x_i) \quad x_i \geq 0$ und
$\mathcal{F} := \{F \mid F \text{ist Verteilungsfkn. von Zufallsvariable } T \text{ mit } Var(T) < \infty\}$

⇒ *Für die 1. Annahme hat man ein Modell mit einem unbekannten Parameter* $\lambda > 0$. *Solche Modelle (mit endlich dimensionalem Parameter) heißen* *parametrisch.*

Für die 2. Annahme ist der Parameterraum die Menge aller Verteilungsfunktionen mit endlicher Varianz. In solchen Fällen (nicht durch endlich dimensinale Parameter darstellbar) spricht man von nicht parametischen Modellen.

A. Punkt- und Bereichsschätzer

Ziel: Anhand von beobachtungen Kenngrößen der wahren Verteilung zu schätzen.

Definition 8.4. *Sei* $(S, \mathcal{D}, (P_\vartheta)_{\vartheta \in \Theta})$ *ein statistisches Experiment und* $g : \Theta \to \Gamma$ *eine Abbildung. Ist* $\mathcal{C}$ *eine* σ*-Algebra auf* Γ*, so heißt jede* Γ*-wertige Zufallsvariable*
$g : (S, \mathcal{D}) \to (\Gamma, \mathcal{C})$ *ein (Punkt-)Schätzer von* $g(\vartheta)$.

Ist $x \in S$ *die vorliegende Beobachtung, so wird* $\hat{g}(x)$ *auch Schätzwert für* $g(\vartheta)$ *genannt*

Beispiel 8.5. *Radioaktiver Zerfall mit langer Halbwertszeit.*

Anzahl der Zerfälle in n *auf einanderfolgenden Minuten wird beobachtet.*
→ *Modellierung durch stochastisch unabhängige* $Poi(\lambda)$ *Zufallsvariablen* $(\lambda \geq 0)$ *(Poissonscher Grenzwertsatz 3.7)*
$(\mathbb{N}_0^n,\ 2^{\mathbb{N}_0^n},\ (\mathcal{P}_\lambda^n)_{\lambda \geq 0})$,
wobei $\mathcal{P}_\lambda^n$ *das Wahrscheinlichkeitsmaß von* n *stochastisch unabhängigen* $Poi(\lambda)$ *Zufallsvariablen ist.*

→ *geschätzt wird* λ *selbst* $(g(\lambda) := \lambda)$
λ *ist der der Erwartungswert* ⇒ *das arithmetische Mittel ist ein sinnvoller Schätzer ((schwaches) Gesetz der Großen Zahlen)*

$$\hat{\lambda} : \mathbb{N}_0^n \to [0, \infty), \hat{\lambda}(l_1, \ldots, l_n) = \frac{1}{n}\sum_{i=1}^{n} l_i$$

$L_1, \ldots, L_n$ *stochastisch unabhängig* $\backsim Poi(\lambda)\ \lambda \geq 0$
$p(l_1, \ldots, l_n) = \mathcal{P}_\lambda^n$
$\hat{\lambda}(L_1, \ldots, L_n) = \frac{1}{n}\sum_{i=1}^{n} L_i$
$\Rightarrow E_\lambda[\hat{\lambda}(L_1, \ldots, L_n)] = \frac{1}{n}\sum_{i=1}^{n} E_\lambda[l_i] = E_\lambda[L_i] = \lambda$
⇒ *Der Schätzer* $\hat{\lambda}$ *ist erwartungstreu*

Maß für die Güte: MSE (mean squared error)

$$\begin{aligned}
MSE_\lambda(\hat{\lambda}) :&= E_\lambda[(\hat{\lambda} - \lambda)^2] \\
&= E_\lambda\left[\left(\frac{1}{n}\sum_{i=1}^{n} L_i - E_\lambda\left[\frac{1}{n}\sum_{i=1}^{n} L_i\right]\right)^2\right] \\
&= Var_\lambda\left(\frac{1}{n}\sum_{i=1}^{n} L_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var_\lambda(L_i) = \frac{1}{n}\cdot Var_\lambda(L_i) = \frac{\lambda}{n}
\end{aligned}$$

Man kann zeigen, dass es keinen Schätzer λ^* geben kann mit $E_\lambda[\lambda^*] = \lambda$ für $\lambda \geq 0$ und $MSE_\lambda(\lambda^*) < \frac{\lambda}{n}$
$\Rightarrow \hat{\lambda}$ *ist optimal.*

Definition 8.6 (Erwartungstreu & mittlerer quadratischer Fehler)**.** *Ist $(S, \mathcal{D}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Experiment und $g : \Theta \to \mathbb{R}^n$ eine zu schätzende Abbildung. Dann heißt ein Schätzer $\hat{g} : (S, \mathcal{D}) \to (\mathbb{R}^d, \mathbb{B}^d)$ erwartungstreu (oder unbiased / unverzert) für g falls $E_\vartheta[\hat{g}] = g(\vartheta) \quad \forall \vartheta \in \Theta$.*

Für $d = 1$ heißt

$$MSE_\vartheta(\hat{g}) := E[(\hat{g} - g(\vartheta))^2] = Var_\vartheta(\hat{g}) + \underbrace{(E\vartheta(\hat{g}) - g(\vartheta))^2}_{Bias}$$

mittlerer quadratischer Fehler.

Ein erwartungstreuer Schätzer heißt MVUE (minimum variance unbiased estimater), wenn für alle weiteren erwartungstreuen Schätzer $\tilde{g}$ gilt

$$MSE_\vartheta(\tilde{g}) \geq MSE_\vartheta(\hat{g}) \quad \forall \vartheta \in \Theta$$

Bemerkung 8.7. *Der MSE ist nicht bei jeder Anwendung angemessen.
Er wird jedoch oft verwendet, weil er mathematisch gut berechenbar ist*

Definition 8.8. *Sei $(\mathbb{R}^n, \mathcal{B}^n, (P^n_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Experiment, wobei P^n_ϑ die Verteilung von n stochastisch unabhängige Zufallsvariablen $X_1, \ldots, X_n$ bezeichne, die alle die gleiche Verteilung P_ϑ besitzen. Ferner sei für eine messbare Abbildung $f : (\mathbb{R}, \mathbb{B}) \to (\mathbb{R}, \mathbb{B}) g : \Theta \to \mathbb{R}, g(\vartheta) := E_\vartheta[f(X_i)]$ zu schätzen.*

Dann ist der zugehörige verallgemeinerte Momentenschätzer definiert als

$$\hat{g}(x_i, \ldots, x_n) := \frac{1}{n}\sum_{i=1}^{n} f(x_i) \qquad x_1, \ldots, x_n \in \mathbb{R}$$

Beispiel 8.9. *Anzahl der Taxis in einer Stadt soll geschätzt werden.*
Vorgehen: Nummer (Taxis seien durchnummeriert) von n Taxis, die zu einer zufälligen Zeit am Hauptbahnhof warten, werden notiert.

$\rightarrow$ Die Nummern sind stochastisch unabhängig und gleichverteilt auf $\{1, \ldots, M\}$ (wobei M der gesuchte Parameter ist)
$(\mathbb{N}^n, 2^{\mathbb{N}^n}, (\mathcal{U}_M^n)_{M \geq n})$ ist das statistische Experiment.
Zufallsvariablen $K_1, \ldots, K_n \backsim \mathcal{U}_M \qquad M \geq 1$

Schätzer für M?

$$E_M[K_i] = \sum_{k=1}^{M} k \cdot P_M\{K_i = k\} = \sum_{k=1}^{M} k\frac{1}{M} = \frac{M+1}{2}$$

$\Rightarrow M = E_M[f(K_i)]$ *mit* $f(k) = 2k - 1$
$(E_m[f(K_i)] = E_M[2K_i - 1] = 2E_M[K_i] - 1 = M + 1 - 1 = M)$

$\rightarrow$ Momentenschätzer

$$\hat{M}(k_1, \ldots, k_n) = \frac{1}{n}\sum_{i=1}^{n} f(k_i) = \frac{2}{n}k_i - 1$$

$\hat{M}$ ist erwartungstreu:

$$E_M\left[\hat{M}(K_1, \ldots, K_n)\right] = E_M\left[\frac{2}{n}\sum_{i=1}^{n} K_i - 1\right] = \frac{2}{n}\sum_{i=1}^{n} E_M[K_i] - 1 = 2\left(\frac{M+1}{2}\right) - 1 = M$$

und der MSE ist

$$\begin{aligned} MSE_M(\hat{M}) = Var_M(\hat{M}) &= Var_M\left(\frac{2}{n}\sum_{i=1}^{n} k_i - 1\right) \\ &= \frac{4}{n^2}\sum_{i=1}^{n} Var_M(k_i) = \frac{4}{n} Var_M(k_i) \\ &= \frac{4}{n}\left(\frac{M^2 - 1}{12}\right) = \frac{M^2 - 1}{3n} \end{aligned}$$

Satz & Definition 8.10 (Konsistent)**.** *In der Situation von Definition 8.8 ist* $\hat{g} = \hat{g}_n$ *stets erwartungstreu und es gilt*

$$\hat{g}_n \xrightarrow[n\to\infty]{} g(\vartheta) \qquad P_\vartheta^n\text{-}\textit{Stochastisch} \quad \forall \vartheta \in \Theta \tag{8.1}$$

Eine Folge von Schätzern $\hat{g}_n$*, für die die Konvergenz (8.1) gelten heißen (schwach)* <u>*konsistent*</u> *für* g

Beweis:

$$E_\vartheta[\hat{g}(X_1, \dots, X_n)] = E_\vartheta\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right] = \frac{1}{n}\underbrace{\sum}_{E}[f(X_i)] = \frac{1}{n} \cdot n \cdot g(\vartheta) = g(\vartheta)$$

(8.1) folgt aus einer Variante des schwachen Gesetzes der Großen Zahlen, gemäß

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to E_P[X_1] \quad \text{P-stochastisch, falls } X_i \text{ } i.i.d. \text{ mit Erwartungswert } E(X_i)$$

Beispiel 8.11. *(Fortsetzung von Beispiel 3.6)*
N *Anzahl der Fische (unbekannt)*
M *Anzahl der makierten Fische*
n *Anzahl gefangener Fische*
m *Anzahl der makierten gefangen Fische*

$$\mathcal{H}_{(N,M,n)}\{m\} = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

$$\Big(\mathbb{N}_0, 2^{\mathbb{N}_0}), \big(\mathcal{H}_{(N,M,n)}\big)_{N \geq \max(M,n)}\Big)$$

Vorgehen: N *so auswählen, dass die Wahrscheinlichkeit* $\mathcal{H}_{(N,M,n)}\{m\}$ *möglichst groß ist für das tatsächliche* m
$\rightarrow$ *Ziel: von* $\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$ *als Funktion in* N *die Maximalstelle finden.*

Rechung: Für welche N ist

$$\frac{\binom{M}{m}\binom{N+1-M}{n-m}}{\binom{N+1}{n}} \geqq \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

$$\Leftrightarrow \binom{N+1-m}{n-m}\binom{N}{n} \geqq \binom{N-M}{n-m}\binom{N+1}{n}$$

$$\Leftrightarrow \frac{(N+1-M)!}{(n-m)!\cdot(N+1-M-n+m)!}\cdot\frac{N!}{n!(N-n)!}$$

$$\geqq \frac{(N-M)!}{(n-m)!(N-M-n+m)!}\cdot\frac{(N+1)!}{n!(N+1-n)!}$$

$$\Leftrightarrow \frac{(N+1-M)!}{(N-M)!}\cdot\frac{(N+1-n)!}{(N-n)!)} \geqq \frac{(N+1-M-n+m)!}{(N-M-n+m)!}\cdot\frac{(N+1)!}{N!}$$

$$\Leftrightarrow (N+1-M)\cdot(N+1-n) \geqq (N+1-M-n+m)\cdot(N+1)$$

$$\Leftrightarrow N^2+2N-nN+MN+Mn-M-n \geqq N^2+2N-MN-nN+mN-M-n+m$$

$$\Leftrightarrow Mn \geqq mN+m$$

$$\Leftrightarrow N \leqq \frac{M\cdot n}{m}-1$$

<u>*Maximumn-Likelihood-Schätzer*</u>

$$\begin{array}{ll} \hat{N}(m) = \left[\frac{n\cdot M}{m}\right] & \textit{falls } \frac{m\cdot M}{m} \notin \mathbb{N}_0 \\ \hat{N}(m) \in \{\frac{n\cdot M}{m}-1, \frac{n\cdot M}{m}\} & \frac{n\cdot M}{m} \in \mathbb{N}_0 \end{array}$$

Definition 8.12 (Maximumn-Likelihood-Schätzer)**.** *Sei ein statistisches Experiment $(S,\mathcal{D},(P_\vartheta)_{\vartheta\in\Theta})$ gegeben, bei dem entweder alle Wahrscheinlichkeitsmaße P_ϑ diskret sind mit Zähldichte f_ϑ oder $(S,\mathcal{D}) = (\mathbb{R}^n,\mathbb{B}^n)$ und alle Wahrscheinlichkeitsmaße P_ϑ eine Dichte f_ϑ besitzen. Dann heißt ein $\hat{\vartheta}$ für ϑ <u>Maximum-Likelihood-Schätzer</u> (ML-Schätzer) wenn für alle $x \in S$ gilt*

$$f_{\hat{\vartheta}}(x) = \sup_{\vartheta\in\Theta} f_\vartheta(x)$$

d.h. $\hat{\vartheta}(x)$ ist eine Maximalstelle der <u>Likelihoodfunktion</u>

$$L_x : \Theta \to [0,\infty),\ L_x(\vartheta) := f_\vartheta(x) \text{ für alle } x \in S.$$

Ist $\hat{\vartheta}$ ein ML-Schätzer für ϑ und $g : \Theta \to \Gamma$ eine zu schätzende Abbildung, so heißt jeder Schätzer der Form $g(\hat{\vartheta})$ auch Maximum-Likelihood-Schätzer für $g(\vartheta)$ (Likelihood $\hat{=}$ Plausibilität).

Bemerkung 8.13.

- *Es kann mehrere ML-Schätzer geben*
- *Es muss nicht immer ein ML-Schätzer existieren*

Seien $X_1, \dots, X_n \quad X_i \backsim f_{i,\vartheta}$ *stochastisch unabhängige Zufallsvariablen*
$\Rightarrow (X_1, \dots, X_n) \backsim \prod_{i=1}^{n} f_{i,\vartheta}$
In diesem Fall ist es einfacher den logarithmus der Likelihoofunktion zu maximieren.

$$LL_x : \Theta \to [-\infty, \infty)$$
$$LL_x(\vartheta) := \log(f_\vartheta(x))$$

Da der Logarithmus streng monoton steigend ist, stimmen die Maximalstellen von L_x *und* LL_x *überein.*

$$L_x(\vartheta) = \prod_{i=1}^{n} f_{i,\vartheta}(x_i)$$
$$LL_x(\vartheta) = \sum_{i=1}^{n} \log(f_{i,\vartheta}(x_i))$$

Beispiel 8.14. $X_1, \dots, X_n \quad$ *i.i.d.* $\backsim \mathcal{N}_{(\mu,\sigma^2)}$

1. Experiment $\quad \mu$ *unbekannt,* $\sigma^2 > 0$ *bekannt*
2. Experiment $\quad \mu$ *unbekannt,* σ^2 *unbekannt*

$E_1 = \left(\mathbb{R}^n, \mathbb{B}^n, \left(\mathcal{N}^n_{(\mu,\sigma^2)}\right)_{\mu\in\mathbb{R}}\right) \; E_2 = \left(\mathbb{R}^n, \mathbb{B}^n, \left(\mathcal{N}^n_{(\mu,\sigma^2)}\right)_{(\mu,\sigma^2\in\mathbb{R}^n_x}\right)$

ML-Schätzer für μ*:*

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

1. Experiment

$$Lx(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}}$$
$$\begin{aligned} LL_x(\mu) &= \sum_{i=1}^{n} \log\Big(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}}\Big) \\ &= \sum_{i=1}^{n} \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned}$$

$\Rightarrow$ *maximal, genau dann wenn* $-\sum\limits_{i=1}^{n}(x_i-\mu)^2$ *maximal ist*

$$\frac{\partial}{\partial\mu}\left(-\sum_{i=1}^{n}(x_i-\mu)^2\right)=2\sum_{i=1}^{n}(x_i-\mu)\stackrel{!}{=}0\Leftrightarrow\mu=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$\hat{\mu}(x):=\frac{1}{n}\sum\limits_{i=1}^{n}x_i$ *ist der eindeutige ML-Schätzer*

2. Experiment

$$Lx(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

$$LL_x(\mu)=-\frac{n}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\begin{aligned}\frac{\partial}{\partial\mu}LL_x(\mu,\sigma^2)&=\frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\\&=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\stackrel{!}{=}0\Leftrightarrow\mu=\frac{1}{n}\sum_{i=1}^{n}x_i\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\sigma}LL_x(\mu,\sigma^2)&=-\frac{n}{2}\cdot\frac{1}{2\pi\sigma^2}\cdot2\pi+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2\\&=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2\stackrel{!}{=}0\Leftrightarrow\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\end{aligned}$$

$$\begin{aligned}L_x(\mu,\sigma^2)&\xrightarrow[\sigma^2\to\infty]{}0\\L_x(\mu,\sigma^2)&\xrightarrow[\sigma^2\to0]{}0\qquad\textit{falls } x_i-\mu\neq0\ \forall i\end{aligned}$$

$(\hat{\mu}(x),\hat{\sigma}^2(x))=\left(\frac{1}{n}\sum\limits_{i=1}^{n}x_i,\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\hat{\mu}(x))^2\right)$ *ist der ML-Schätzer für* (μ,σ^2).

Gesamtbeobachtung habe eine Dichte oder Zähldichte f_ϑ, falls ϑ der wahre Parameter ist.
Der ML-Schätzer maximiert diese:

$$f_{\hat{\vartheta}(x)}(x)\geq f_\vartheta(x)\qquad\forall\vartheta\in\Theta$$

Beispiel 8.15. *(Fortsetzung von Beispiel 8.9)*
Beobachtet werden n i.i.d. Zufallsvariablen K_i, die gleichverteilt sind auf $\{1,\ldots,M\}$, wobei M zu schätzen ist.

Likelihoodfunktion mit $K = (K_i,\ldots,K_n) \in \mathbb{N}^n$

$$L_K(M) = \prod_{i=1}^{n} \frac{1_{\{1,\ldots,M\}}}{M}(K_i) = \frac{1_{\{1,\ldots,M\}}\left(\max\limits_{1\le i\le n} K_i\right)}{M^N}$$

wird maximal für $M = \max\limits_{1\le i\le n} K_i$
d.h. ML-Schätzer ist eindeutig gegeben durch $M*(K) = \underset{1\le 1\le n_i}{K}$.
Beachte: $M*(K) \le M$ und $M*(K) < M$, wenn nicht das Taxi mit der größten Konzessionsnummer beobachtet wird.
Daher ist $E_M(M*) < M$, d.h. der Schätzer ist <u>nicht</u> erwartungstreu.

$$\begin{aligned}
& P_M\{M*(K)\le m\} &&= P_M\{K_i \le m \; 1\le i\le n\} = \prod_{i=1}^{n} P_M\{K_i\le m\} = \left(\tfrac{m}{M}\right)^n \qquad 1\le m\le M\\
\Rightarrow\;& P_M\{M*(K)=m\} &&= P_M\{M*(K)\le m\} - P_M\{M*(K)\le m-1\}\\
& &&= \left(\tfrac{m}{M}\right)^n - \left(\tfrac{m-1}{M}\right)^n \;\; 1\le m\le M\\
\Rightarrow\;& E_M(M*) &&= \sum_{m=1}^{M} m\cdot\left(\left(\tfrac{m}{M}\right)^n - \left(\tfrac{m-1}{M}\right)^n\right)\\
& &&= M^{-n}\sum_{m=1}^{M}\left(m^{n+1} - (m-1+1)\cdot(m-1)^n\right)\\
& &&= M^{-n}\left[\sum_{m=1}^{M}\left(m^{n+1}-(m-1)^{n+1}\right) - \sum_{m=1}^{M}(m-1)^n\right]\\
& &&= M^{n}\left[M^{n+1} - \sum_{k=1}^{M-1}k^n\right]\\
& &&= M - M^n\sum_{k=1}^{M-1}k^n\\
\Rightarrow\;& \mathit{Bias} &&= E_M(M*) - M = -M^{-n}\sum_{k=1}^{M-1}k^n = b_{n,M}
\end{aligned}$$

d.h. $0 > b_{n,M} > -M^n\int\limits_1^M t^n\,dt = -M^{-n}\frac{t^{n+1}}{n+1}\Big|_1^M > -\frac{M}{n+1}$

In Beispiel 8.9

$$\hat{M}(K) = \frac{2}{n}\sum_{i=1}^{n}K_i - 1$$

$$MSE_M(\hat{M}) = E_M((\hat{M}-M)^2) = \frac{2}{n}\cdot(M^2-1)$$

Frage: Hat der ML-Schätzer $M*$ einen kleineren MSE?
Ähnlich wie oben ergibt sich

$$\begin{aligned}
E_M((M*)^2) &= \sum_{m=1}^{M} m^2 \left(\left(\frac{m}{M}\right)^n - \left(\frac{m-1}{M}\right)^n \right) \\
&= M^{-n} \sum_{m=1}^{M} (m^{n+2} - (m-1)^2 - 2(m-1)^{n+1} - (m-1)^2) \\
&= M^2 - 2M^{-n} \sum_{k=1}^{M-1} k^{n+1} - M^n \sum_{k=1}^{M-1} k^n \\
\Rightarrow MSE_M(M*) &= Var(M*) + b_{b,M}^2 = E((M*)^2)) - (E(M*))^2 + b_{n,M}^2 \\
&= M^2 - 2M^{-n} \sum_{k=1}^{M-1} k^{n+1} - M^{-n} \sum_{k=1}^{M-1} k^n - (M + b_{n,M})^2 + b_{n,M}^2 \\
&= -2M^n \sum_{k=1}^{M-1} k^{n+1} - M^n \sum_{k=1}^{M-1} k^n - 2M \cdot b_{n.M} \\
&= M^{-n} \left[(2M-1) \sum_{k=1}^{M-1} k^n - 2 \sum_{k=1}^{M-1} k^{n-1} \right]
\end{aligned}$$

Für großes M verhält sich dies wie

$$M^{-n} \left[2M \int_0^M t^n \, dt - 2 \int_0^M t^{n+1} \, dt \right] = M^2 \left(\frac{2}{n+1} - \frac{2}{n+2} \right) = \frac{2M^2}{(n+1)(n+2)}$$

Der MSE des ML-Schätzers ist größenordnungsmaäßig um den Faktor n kleiner als der MSE von $\hat{M}$!

Punktschätzer liefern keine Angaben zur Genauigkeit.
Daher werden oft zusätzlich datenabhängige Mengen gesucht, die den wahren Wert wenigstens mit einer vorgegebenen Wahrscheinlichkeit enthalten.
$\rightsquigarrow$ Bereichsschätzer

Beispiel 8.16. *(Fortsetzung von Beispiel 8.14)*
Gegeben ist ein statistisches Experiment $E_1 = (\mathbb{R}^n, \mathbb{B}^n, \left(\mathcal{N}^n_{(\mu,\sigma^2)}\right)_{\mu\in\mathbb{R}})$ ($\sigma^2 > 0$ fest).

Schätzer ist $\hat{\mu}_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad X = (X_1, \ldots, X_n) \in \mathbb{R}^n$

Seien $X_1, \ldots, X_n$ i.i.d $\mathcal{N}_{(\mu,\sigma^2)}$-verteilt.
Beispiel 6.19 zeigt: Y_1, Y_2 seien unabhängige Zufallsvariablen und $P^{Y_i} = \mathcal{N}_{\mu_i,\sigma_i^2)} \Rightarrow P^{Y_1+Y_2} = \mathcal{N}_{(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)}$
Induktiv folgt: $\sum_{i=1}^{n} X_i$ ist $\mathcal{N}_{n\mu,n\sigma^2}$-verteilt.

Beispiel 6.19 zeigt auch: $aY_1 + b$ *ist* $\mathcal{N}_{(a\mu+b,a^2\sigma^2}$*-verteilt*
Hier: $\hat{\mu}(X) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ *ist* $\mathcal{N}_{(\mu,\frac{\sigma^2}{n}}$*-verteilt*
$\Rightarrow \sqrt{\frac{n}{\sigma^2}}(\hat{\mu}_n(X) - \mu$ *ist* $\mathcal{N}_{(0,1)}$*-verteilt.*
$\Rightarrow P_\mu\left\{-t \le \sqrt{\frac{n}{\sigma^2}}\left(\hat{\mu}_n(X) - \mu\right) \le t\right\} = \Phi(t) - \Phi(-t) = 2 \cdot (\Phi(t) - \underbrace{\frac{1}{2}}_{\Phi(0)})$ *für* $t \ge 0$

Zu vorgegebener Wahrscheinlichkeit $\alpha \in (0,1)$ *wähle* t *so, dass*

$$2(\Phi(t) - \frac{1}{2}) = \alpha \Leftrightarrow t = \Phi^{-1}(\frac{1+\alpha}{2} =: c_\alpha$$

$\Rightarrow P_\mu\left\{\mu \in \left[\hat{\mu}_n(X) - \sqrt{\frac{\sigma^2}{n}}c_\alpha, \hat{\mu}(X) + \sqrt{\frac{\sigma^2}{n}}c_\alpha\right]\right\} = P_\mu\left\{-c_\alpha \le \sqrt{\frac{\sigma^2}{n}}(\hat{\mu}_n(X) - \mu) \le c_\alpha\right\} = \alpha$

Fazit: Bereits vor der Beobachtung weiß man wie ein datenabhängiges Intervall zu konstruieren ist, in dem der wahre Wert mit vorgegebener Wahrscheinlichkeit α *liegt.*

Definition 8.17 (Konfidenzintervall)**.** *Sei* $(S, \mathcal{D}, (P_\vartheta)_{\vartheta\in\Theta}$ *ein statistisches Experiment,* $g : \Theta \to \mathbb{R}$ *und* $\alpha \in [0,1]$.
Sind $X_u, X_o : (S,\mathcal{D}) \to (\mathbb{R}, \mathbb{B})$ *mit* $X_u \le X_o$ *so, dass* $P_\vartheta\{g(\vartheta \in [X_u, X_o]\} \ge \alpha \quad \forall \vartheta \in \Theta$ *wird* $[X_u, X_o]$ *zweiseitiges Konfidenzintervall zum Niveau* α *von* $g(\vartheta)$ *genannt.*
Gild $P_\vartheta\{g(\vartheta) \in [X_u, \infty)\} \ge \alpha \quad \forall \vartheta \in \Theta$ *bzw.* $P_\vartheta\{g(\vartheta) \in) - \infty, X_o]\} \ge \alpha \quad \forall\vartheta \in \Theta$*, so heißt* $[X_i, \infty)$ *bzw.* $(-\infty, X_o]$ *einseitiges Konfidenzintervall zum Niveau* α *und* X_u *heißt untere Konfidenzschranke bzw.* X_o *obere Konfidenzschranke*

Einseitige Konfidenzintervalle werden i.d.R. dann verwendet, wenn eine Unterschätzung (bzw. Überschätzung) des Parameters bei weitem kritischer ist, als eine Überschätzung (bzw. Unterschätzung)
$\to [X_u, \infty)$ bzw. $(-\infty, X_o]$

Oft sind nur approximative Konfidenzintervalle berechenbar,
also I, so dass $P_\vartheta\{g(\vartheta) \in I\} \approx \alpha$

Beispiel 8.18. *(*$\to$ *Beispiel 8.5)*
Experiment: $(\mathbb{N}_0^n, 2^{\mathbb{N}_0^n}, (\mathcal{P}_\lambda^n)_{\lambda\ge 0})$
d.h. es werden n *i.i.d.* $\mathcal{P}_\lambda$ *Zufallsvariablen* L_i *beobachtet (*$\mathcal{P}_0 =$ *Wahscheinlichkeitsmaß mit Zähldichte* $f = \{0\}$*)*
In Beispiel 8.5: Schätzer $\hat{\lambda}_n = \frac{1}{n}\sum\limits_{i=1}^{n} L_i$

Mit dem zentralen Grenzwertsatz erhält man (wobei $E(L_i) = \lambda$, $Var(L_i) = \lambda$

$$P\left\{\frac{\sum\limits_{i=1}^{n}(L_i - \lambda)}{\sqrt{n \cdot \lambda}} \leq x\right\} \to n \to \infty \Phi(x) \qquad \forall x \in \mathbb{R}$$

$$\Rightarrow P\{\sqrt{\frac{n}{\lambda}}(\hat{\lambda}_n - \lambda) \leq x\} \to \Phi(x)$$

$$\Rightarrow P\{-t \leq \sqrt{\frac{n}{\lambda}}(\hat{\lambda}_n - \lambda) \leq t\} \to \Phi(t) - \Phi(-t) \qquad \forall t \geq 0$$

Wähle t so, dass $\lambda = \Phi(t) - \Phi(-t) = 2(\Phi(t) - \frac{1}{2}) \Leftrightarrow t = \Phi^{-1}\left(\frac{1+\alpha}{2}\right) =: c_\alpha$
Für einen hinreichend großen Stichprobenumfang n gilt

$$P\{-c_\alpha \leq \sqrt{\frac{n}{\lambda}}(\hat{\lambda}_n - \lambda) \leq c_\lambda\} \approx \alpha$$

$$- c_\alpha \leq \sqrt{\frac{n}{\lambda}}(\hat{\lambda}_n - \lambda) \leq c_\lambda \Leftrightarrow \lambda \in \left[\hat{\lambda}_n - \sqrt{\frac{n}{\lambda}} \cdot c_\lambda, \hat{\lambda}_n + \sqrt{\frac{n}{\lambda}} \cdot c_\lambda\right]$$

Ersetze in dem Intervall den unbekannten Parameter λ durch den Schätzer $\hat{\lambda}_n$
$\rightsquigarrow$ approximatives Konfidenzintervall

$$I_n := \left[\hat{\lambda}_n - \sqrt{\frac{\hat{\lambda}_n}{n}} \cdot c_\alpha, \hat{\lambda}_n + \sqrt{\frac{\hat{\lambda}_n}{n}} \cdot c_\alpha\right]$$

Man kann zeigen:

$$\mathcal{P}_\lambda^n\{\lambda \in I_n\} \to n\alpha \qquad \forall \lambda > 0$$

Analog erhält man einseitige approximative Konfidenzintervalle, wie z.B.

$$P\left\{\sqrt{\frac{n}{\lambda}} \cdot (\hat{\lambda}_n - \lambda) \leq t\right\} \to \Phi(t) \overset{!}{=} \alpha$$

$$\Leftrightarrow t = \Phi^{-1}(\alpha)$$

$$\Rightarrow P\{\lambda \geq \hat{\lambda}_n - \sqrt{\frac{n}{\lambda}} \cdot \Phi^{-1}(\lambda) \approx \alpha$$

d.h. $\tilde{I}_n = \left[\hat{\lambda}_n - \sqrt{\frac{n}{\lambda}}\Phi^{-1}(\alpha), \infty\right]$ ist ein einseitiges approximatives Konfidenzintervall

---

B. <u>Statistische Tests</u>

Ziel: Überprüfe Hypothesen über den zu gundeliegenden Zufallsmechanismus anhand von Daten.
Es treten zwei mögliche Typen von Fehler solcher Tests auf:

Fehler 1. Art : Der Test lehnt fälschlicher WEise die Hypothese ab, obwohl sie zutrifft.

Fehler 2. Art : Der Test lehnt die Hypothese nicht ab, obwohl sie falsch ist.

Bei vielen Anwendungen hat Fehler 1. Art deutlich gravierendere Konsequenzen als Fehler 2. Art.

Daher: Man gibt sich $\alpha[0, 1$ vor und fordert, dass die Wahrscheinlichkeit für Fehler 1. Art $\leq \alpha$ (übliche werte $\alpha = 0,1; 0,05; 0,01; 0,005$). Unter dieser Nebenbedingung soll die Wahrscheinlichkeit für einen Fehler 2.Art möglichst klein werden.

Definition 8.19 (Testproblem, Hypothese, Alternative, Ablehnungsbereich, Signifikantsniveal, statistischer Test & Gütefunktion)**.** *Sei $(S, \mathcal{D}, (P_\vartheta)_{\vartheta\in\Theta}$ ein statistisches Experiment.*
Ein <u>Testproblem</u> ist eine Zerlegung der Parametermenge Θ in 2 disjunkte Mengen $\Theta_o \neq \emptyset$, die sog. <u>Hypothese</u> (oder <u>Null-Hypothese</u> und $\Theta_1 = \Theta \setminus \Theta_0 \neq \emptyset$ die sog. <u>Alternative</u> (oder <u>Gegenhypothese</u>). Ein (nicht-randomisierter) <u>statististischer Test</u> ist eine 0,1-wertive Zufallsvariable $\varphi : (S, \mathcal{D}) \to (\{0, 1\}, 2^{\{0,1\}})$.

$\{\varphi = 1\} = \{x \in S \mid \varphi(x) = 1\}$ heißt <u>Ablehnungsbereich</u> oder <u>Verwerfungsbereich</u> oder <u>kritischer Bereich</u> des Tests.
Der Test hat ein <u>Signifikantsniveau</u> $\alpha \in [0, 1]$, falls $P_\vartheta(\{\varphi = 1\}) \leq \alpha$ für alle $\varphi \in \Theta_0$ gilt.
Die Abbildung $g_\varphi(\vartheta) = P_\vartheta(\{\varphi = 1\})$, $\vartheta \in \Theta_1$, heißt <u>Gütefunktion</u> des Tests.

Interpretation:
Annahme: Für ein $\vartheta \in \Theta$ beschreibt P_ϑ den Zufallsmechanismus.
Hypothese: “Der wahre Parameter liegt in Θ_0” soll überprüft werden.
$\varphi(x) = 1$: Test lehnt auf der Basis der Daten x die Hypothese ab.
$\varphi(x) = 0$: Test lehnt auf der Basis der Daten x die Hypothese nich ab.
$g_\varphi(\vartheta)$: Die wahrscheinlichkeit, dass der Test die Hypothese ablehnt, wenn ϑ der wahre Parameter ist.
Forderung: $g_\varphi \leq \alpha$ auf Θ_0
Ziel: g_φ möglichst groß auf Θ_1
Die Forderungen sind gegenläufig.

Skitze im Fall $\Theta = \mathbb{R}, \Theta_0 = (-\infty, \vartheta_0], \Theta_1 = (\vartheta_0, \infty)$
Typische Gütefunktion: ——————— Skitze ———————

Konseqzenz:

- Wenn der Test die Hypothese ablehnt, kann man mit hoher Wahrscheinlichkeit davon ausgehen, dass die Hypothese falsch ist.
- Wenn der Test die Hypothese nicht ablehnt, darf man nicht davon ausgehen, dass die Hypothese zutrifft, da es meistens Parameterwerte aus der Hypothese gibt, für die die Ablehnungswahrscheinlichkeit wenig größer ist als α.

$\rightsquigarrow$ Möchte man eine Annahme an den Parameter mit hoher Wahrscheinlichkeit nachweisen, so ist die Annahme als Alternative zu wählen (nicht als Hypothese).

Beispiel 8.20. *Spieler A gewinnt doppelten Einsatz, wenn bei einem Münzwurf Kopf fällt.*
Spieler A möchte daher wissen, ob die Wahrscheinlichkeit für Kopf wenigstens $\frac{1}{2}$ ist.

Er wirft die Münze n mal und zählt die Anzahl X der Köpfe $\rightsquigarrow$ X $\mathcal{B}_{(n,p)}$-verteilt
Dazugehörig ist das statistische Experiment $(\{0,\dots,n\}, 2^{\{0,\dots,n\}}, (\mathcal{B}_{n,p})_{p\in[0,1]})$

Hypothese: $\Theta_0 = [0, \frac{1}{2})$
Alternative: $\Theta_1 = [\frac{1}{2}, 1]$
Beachte: Zu überprüfende Annahme ist die Alternative

naheliegender Test:

$$\varphi(x) = 1_{\{c,\dots,n\}}(x)$$

d.h. die Hypothese wird abgelehnt, wenn wenigstens c-mal Kopf fällt.
Dabei soll zu vorgegebenem $\alpha \in (0,1)$ c so gewählt werden, dass

$$\mathcal{B}_{(n,p)}\{\varphi = 1\} \leq \alpha \qquad \forall p \in \Theta_0 \text{ , d.h. } p < \frac{1}{2}$$

(je größer c ist, desto größer ist die Fehlerwahrscheinlichkeit 1. Art und desto kleiner Fehlerwahrscheinlichkeit 2. Art)

$$\begin{aligned}\rightsquigarrow c_\alpha &= min\{c \in \{0,\dots,n\} \mid \mathcal{B}_{(n,p)}\{c,\dots,n\} \leq \alpha \quad \forall p < \frac{1}{2}\} \\ &= min\{c \mid \mathcal{B}_{(n,\frac{1}{n})}\{c,\dots,n\} \leq \alpha\},\end{aligned}$$

da $p \mapsto \mathcal{B}_{(n,p)}\{c,\dots,n\}$ *monoton steigend und stetig ist.*

numerisches Beispiel:
$n = 50,\ \alpha = 0,1$

$$\begin{aligned}\Rightarrow c_\alpha = 30,\ \textit{denn } \mathcal{B}_{(n,p)}\{30,\dots,50\} &\approx 0,0595 \\ = \mathcal{B}_{(n,p)}\{29,\dots,50\} &\approx 0,1013\end{aligned}$$

Die Gütefunktion ist:

$$g_\varphi(p) = \mathcal{B}_{(n,p)}\{c_\alpha, \dots, n\} = \sum_{j=c_\alpha}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$

Lemma 8.21. *Sei $(S, \mathcal{D}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Experiment und $g : \Theta \to \mathbb{R}$ eine nicht-konstante Abbildung.*

(i) Ist I ein (zweiseitiges) Konfidenzintervall für $g(\vartheta)$ zum Niveau $1-\alpha$, so ist $\varphi = 1_{\{}a \notin I\}$ ein Test zum Signifikantsniveau α für die Hypothese $\Theta_0 = \{\varphi \in \Theta \mid g(\vartheta) = a\} = \hat{g}(\{a\}$ (für alle $a \in g(\Theta_1)$)

(ii) Ist $I = [X_u, \infty)$ ein einseitiges Konfidenzintervall zum Niveau $1-\alpha$, so ist $\varphi = 1_{\{1\notin I\}}$ ein Test zum Signifikansniveau α für die Hypothese $\Theta_0 = \{\vartheta \in \Theta \mid g(\vartheta) \leq a\} = \hat{g}((-\infty, a])$ für alle a mit $g^{-1}((-\infty, a]) \notin \{\emptyset, \Theta\}$

(iii) Ist $I = (-\infty, X_o)$ ein einseitiges Konfidenzintervall zum Niveau $1-\alpha$, so ist $\varphi = 1_{\{1\notin I\}}$ ein Test zum Signifikansniveau α für die Hypothese $\Theta_0 = \{\vartheta \in \Theta \mid g(\vartheta) \geq a\} = \hat{g}([a, \infty))$ für alle a mit $g^{-1}([a, \infty)) \notin \{\emptyset, \Theta\}$

Beispiel 8.22. *Wie in Beispiel 8.14 und 8.16 betrachte zwei statistische Experimente*

$$E_1 = (\mathbb{R}^n, \mathbb{B}^n, (\mathcal{N}_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}} \quad (\sigma^2 fest)$$
$$E_2 = (\mathbb{R}^n, \mathbb{B}^n, (\mathcal{N}_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}x[0,\infty)}$$

Testproblem: $\Theta_0 = \{\mu_0\}$, $\Theta_1 = \mathbb{R} \setminus \{\mu_0\}$
Zweiseitiges Konfidenzintervall für μ zum Niveau $1-\alpha$ ist

$$I_{1-\alpha}(x) = \left[\hat{\mu}_n(x) - \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1-\alpha}{2}\right), \hat{\mu}_n(x) + \frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1-\alpha}{2}\right)\right] \qquad \text{(Beispiel 8.16)}$$

wobei $\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$ ein Schätzer und $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ der Beobachtungsvektor ist.
Mit Lemma 8.21(i) erhält man einen Test zum Niveau α

$$\varphi(x) = 1_{\{\mu_0 \notin I(x)\}} = 1_{\underbrace{\left\{\sqrt{\frac{n}{\sigma^2}}|\hat{\mu}_n(x) - \mu_0| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right\}}_{\text{Teststatistik}}}$$

Gütefunktion:

$$g_\varphi(\mu) = \mathcal{N}^n_{(\mu,\sigma^2)}\left\{x \mid \underbrace{\sqrt{\frac{n}{\sigma^2}|\hat\mu_n(x)-\mu_0|}}_{|T_n|} > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right\}$$

$$= 1 - \mathcal{N}^n_{(\mu,\sigma^2)}\left\{-\Phi^{-1}\left(1-\frac{\alpha}{2}\right) - \sqrt{\frac{n}{\sigma^2}} \le T_n - \sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0) \le \Phi^{-1}\left(1-\frac{\alpha}{2}\right) - \sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0)\right\}$$

$$= 1 - \left(\Phi\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right) - \sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0)\right)\right) - \Phi\left(-\Phi^{-1}\left(1-\frac{\alpha}{2}\right) - \sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0)\right)$$

$T_n := \sqrt{\frac{n}{\sigma^2}}(\hat\mu_n - \mu_0)$
Beispiel 8.16: $\hat\mu_n$ *ist* $\mathcal{N}_{(\mu,\frac{\sigma^2}{n})}$*-verteilt, wenn* μ *der wahre Mittelwert ist.*
$\Rightarrow T_n$ *ist* $\mathcal{N}_{(\sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0,\frac{n}{\sigma^2}\frac{\sigma^2}{n})} = \mathcal{N}_{(\sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0,1)}$*-verteilt*
$\Rightarrow T_n - \sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0)$ *ist* $\mathcal{N}_{(0,1)}$*-verteilt.*

Einseitige Test erhält an analog
z.B. $\Theta_0 = (-\infty,\mu_0],\ \Theta_1 = (\mu_0,\infty)$

$$\overline{\varphi}(x) = 1_{\{T_(x) > \overline{c}_\alpha\}} \text{ mit } \overline{c}_alpha = \Phi^{-1}(1-\alpha)$$

Es gilt dann

$$g_{\overline{\varphi}}(\mu) = \mathcal{N}^n_{(\mu,\sigma^2)}\{T_n > \Phi^{-1}(1-\alpha)\} = 1 - \Phi(\Phi^{-1}(1-\alpha) - \sqrt{\frac{n}{\sigma^2}}(\mu-\mu_0)) \quad \forall \mu \in \mathbb{R}$$

Im Experiment E_2 *ist* σ^2 *unbekannt und* φ *(bzw.* $\overline{\varphi}$*) kein zulässiger Test mehr.*
$\rightsquigarrow$ *Ersetze* σ^2 *durch Schätzer, z.B.* $\hat\sigma^2 = \frac{1}{n-1}\sum\limits_{i=1}^n (x_i - \hat\mu_n(x))^2$
In Präsentzblatt 10 Aufgabe A und B wurde gezeigt, dass $\hat\sigma^2$ *ein erwartungstreuer, konsistenter Schätzer für* σ^2 *ist.*
$\rightsquigarrow \tilde T_n := \sqrt{\frac{n}{\hat\sigma^2_n}}(\hat\mu_n - \mu_0)$

$$\varphi_2(x) = 1_{\{|\tilde T_n(x)| c_\alpha\}}$$

Wenn μ_0 *der wahre Mittelwert ist, so hat* $\tilde T_n$ *die sogenante (Student'sche) t-Verteilung mit* $n-1$ *Freiheitsgraden und Dichte*

$$h_{n-1}(t) = \underbrace{\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2}})\sqrt{\pi(n-1)}}_{c_\alpha}(1+\frac{t}{n-1})^{-\frac{n}{2}}, \qquad t \in \mathbb{R}$$

Wähle c_α *so, dass*

$$\mathcal{N}^n_{(\mu_0,\sigma^2)}\{T_n > c_\alpha\} = \int\limits_{c_\alpha}^{\infty} h_{n-1}(t)dt \stackrel{!}{=} \alpha$$

$\Rightarrow \varphi_2$ *hat exakt das Niveau* α
c_α *ist größer als der kritische Wert* $\Phi^{-1}(1-\frac{\alpha}{2})$ *bei bekanntem* σ^2

$\alpha = 0,08$

$n=$	10	20	51	101
$c_\alpha=$	2,262	2,093	2,009	1,984

$\Phi^{-1}(1-\frac{\alpha}{2} = 1,9600$

Der Test $\tilde{\varphi}_2(x) = 1_{\{\tilde{T}_n > \Phi^{-1}(1-\frac{\alpha}{2}\}}$ *hat Niveau* $> 0,08$ *bei 10*

Manchmal ist zu überprüfen, ob eine bestimmte Verteilung oder ein bestimmter Verteilungstyp vorliegt (und nicht, ob ein Parameter in einem bestimmten Bereich liegt).
$\rightsquigarrow$ Anpassungstests (goodness of fit test)

Beispiel 8.23. *Beobachtet werden i.i.d. Zufallsvariablen mit Werten* $\{0,\ldots,k\}$
Ziel: Überprüfe, ob die Zufallsvariablen eine vorgegebene Zähldichte haben (z.B. Zufallsvariablen sind gleichverteilt auf $\{0,\ldots,k\}$*).*

Beobachtungsvektor $x = (x_1,\ldots,x_n)$

$$\hat{f}_n(i) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{i\}}(x_j) \text{ : relative Häufigkeit mit der } i \text{ auftritt } (0 \le i \le k)$$

Wenn f_0 *die wahre Zähldichte ist, dann* $\hat{f}_n(i) \approx f_0(i)$.
Dann ist $n \cdot \hat{f}_n(i)$ $\quad \mathcal{B}_{(n,f_0(i))}$*-verteilt*
Der Zentrale Grenzwertsatz liefert

$$P\left\{a < \frac{n\cdot\hat{f}_n(i) - n\cdot f_0(i)}{\sqrt{n\cdot f_0(i)(1-f_0(i)}} \le b\right\} \xrightarrow{n\to\infty} \Phi(b) - \Phi(a)$$

Ist $\frac{|n\hat{f}_n(i) - n\cdot f_0(i)|}{\sqrt{n\cdot f_0(i)(1-f_0(i))}}$ *"groß", dann ist die Hypothese vermutlich falsch.*
Test statistik $\sum_{i=0}^{k} \frac{(n\cdot\hat{f}_n(i) - n\cdot f_0(i))^2}{n\cdot f_0(i)} =: T_n$

χ^2-Test: $\varphi = 1_{\{T_n > c_\alpha\}}$ *mit* c_α *so, dass* $P\{T_n > c_\alpha\} \approx \alpha$ *wenn* f_0 *die wahre Zähldichte ist.*

Man kann zeigen: T_n *hat näherungsweise χ^2-Verteilung mit k Freiheitsgraden, d.h. die Dichte*

$$g_k(t) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} t^{\frac{k}{2}-1} e^{-\frac{t}{2}} 1_{(0,\infty)}(t)$$

$\rightsquigarrow$ *Wähle* c_α *so, dass* $\int_{c_\alpha}^{\infty} g_k(t)dt = \alpha \Leftrightarrow \int_0^{c_\alpha} g_k(t)dt = 1-\alpha$

Faustformel: Approximation ist gut, wenn $n\cdot f_0(i) \ge 5 \quad \forall i$ *(meist reicht* ≥ 1*).*
Sonst: Fasse mögliche Werte in Gruppe zusammen, bis erwartete Anzahl der Beobachtungen in jeder Gruppe hinreichen groß ist (≥ 5 *bzw.* ≥ 1*)*

Beispiel aus Übungsblatt 12, Aufgabe 2(ii)

Anzahl	*Häufigkeit*	*Test auf* $\mathcal{P}_{0,61}$ $(nf_0(i))$
0	*109*	*108,67*
1	*65*	*66,29*
2	*22*	*20,22*
≥ 3	*4*	*4,82*

$\rightsquigarrow T_n \approx 0,3235, \quad \alpha = 0,1 \quad c_\alpha \approx 6,25$ *bei 4 Gruppe* $\hat{=}$ *3 Freiheitsgrade*

Faustregel: Da ein Parameter ($\lambda = 0,61$) *aus den Daten geschätzt wurde, sollte die Anzahl der Freiheitsgrade um 1 verringert werden, d-h-* c_α *wird aus* χ^2*-Verteilung mit 2 Freiheitsgraden berechnet*
$\rightsquigarrow c_\alpha = 4,6$

Fazit: Der Test lehnt die Hypothese nicht ab.

χ^2-Tests können auch zur Überprüfung von stetigen Verteilungen verwendet werden, indem man Wertebereiche der Beobachtungen in Gruppen einteilt (also diskretisiert).

Beispiel 8.24. *Beobachtungen: Lebenszeiten/Funktionsdauer* T_i, $1 \leq i \leq n$, *gleichartiger elektronischer Bauteile.*
Modellannahme: T_i*i.i.d.*
Hypothese: T_i *sind exponential verteilt mit unbekanntem Mittelwert* $\lambda > 0$
Auf Übunbungsblatt 11 Aufgabe 2 wurde gezeigt, dass $\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$ *ein sinnvoller Schätzer ist für* λ.

Der Wertebereich $(0, \infty)$ *wird in Gruppen eingeteilt, so dass die erwartete Anzahl von Beobachtungen in jeder Gruppe* ≈ 5 *ist.*
$\rightsquigarrow$ *Definiere* t_j *so, dass* $m \cdot P\{T_1 \in (t_{j-1}, t_j]\} \approx 5$ *(falls Hypothese richtig ist).*
Für eine Exponentialverteilung E_λ *mit* $\lambda = \hat{\lambda}_n$ *wähle*

$$t_j = E_\lambda^{\leftarrow}\left(\frac{5j}{n}\right), \quad 0 \leq j \leq \left[\frac{n}{5}\right]$$

wobei $E_\lambda^{\leftarrow}$ *die Quantilfunktion von* E_λ *ist, d.h.*

$$E_\lambda^{\leftarrow}(t) = -\lambda \log(1-t) \qquad \textit{(Umkehrfunktion der Verteilungsfunktion } E_\lambda$$

Dann

$$P\{T_1 \in (t_{j-1}, t_j]\} = (1 - e^{-\frac{t_j}{\lambda}}) - (1 - e^{-\frac{t_{j-1}}{\lambda}}) = \frac{5j}{n} - \frac{5 \cdot (j-1)}{n} = \frac{5}{n}$$

Für alle $1 \leq j \leq m$

$$\hat{c}_j = \sum_{i=1}^{n} 1_{(t_{j-1},t_j]}(t_i) \text{ : Anzahl der Beobachtungen im } j\text{-ten Intervall } (t_{j-1}, t_j]$$

sollte ≈ 5 *sein, wenn tatsächlich Exponentialverteilung vorliergt. Teststatistik:* $\sum_{i=1}^{n} \frac{(\hat{c}_j - 5)^2}{5}$

Näherungsweise χ^2*-Verteilung mit* $\underbrace{m-1}_{AnzahlGruppen-1} \underbrace{-1}_{AnzahlParameter} = m-2$

9 Pseudozufallszahlen

Erzeugung auf einer endlichen Menge oder auf [0,1] gleichverteilter Pseudozahlen

Häufigstes Verfahren: lineare Konvergenzgeneratoren

Definition 9.1 (linearer Konvergenzgenerator). *Seien $l, n, m \in \mathbb{N}_0$. Wir schreiben $l \equiv n(\mod n)$, falls $l - n$ durch m teilbar ist.*

Zu vorgegebenen $a, b, m \in \mathbb{N}_0$ erzeugt ein linearer Konvergenzgenerator Pseudozufallszahlen $n_i \in \{0, \ldots, m-1\}$ ausgehend von einem Startwert n_0 nach der Regel $n_{j+1} = a \cdot n_j + b(\mod m)$, d.h. n_{j+1} ist der Rest von $a \cdot j_j + b$ bei Division durch m.

Beachte: Bei gegebenem n_0 ist folge n (n_j) rein deterministisch und wiederholt sich nach spätestens m Folgegliedern.
$\rightsquigarrow$ wähle m sehr groß und a, b, m so, dass Wiederholungen tatsälich erst wieder nach m Gliedern eintritt.

Dies ist der Fall, wenn

(i) a, b keinen gemeinsamen Teiler haben

(ii) Jede Primzahl, die m teilt, auch $a - 1$ teilt

(iii) Wenn 4 m teilt, dann auch $a - 1$

Das Problem bei den linearen Konvergenzgeneratoren ist:
$(n_j, n_{j+1}, \ldots, n_{j+k-1})$ liegt bei ungünstiger Wahr auf wenigen $(k-1)$-dimensionalen Hyperräumen in $\mathbb{R}^k$.

$\rightsquigarrow$ auf $[0, 1]$ "gleichverteilte" Pseudozufallszahlen

$$u_j = \frac{n_j}{m} \in [0, 1)$$

Wenn m groß ist, dann gilt für einen linearen Konvergenzgenerator mit maximaler Periode m

$$\frac{1}{m} \sum_{i=1}^{m} 1_{[0,1]}(U_i) = \frac{[mt] + 1}{m} \qquad \leftarrow (t, t + \frac{1}{m}]$$

$$u_i = \frac{k}{m}$$

d.h. Pseudozufallszahlen $u_i \leq t$ treten ungefähr mit der relativen Häufigkeit t auf.

Erzeugung von Pseudozufallszahlen gemäß weiteren Wahrscheinlichkeitsmaßen auf $\mathbb{R}$

a.) Quantiltransformation
Sei $F^{\leftarrow}$ die Quantilfunktin von F.
Satz 6.27 besagt: Ist U gleichverteilt auf $(0,1)$
$\Rightarrow F^{\leftarrow}(U)$ ist eine Zufallsvariable mit Verteilungsfunktion F.

Also sind $F^{\leftarrow}(u_j)$ Pseudozufallszahlen gemäß der Verteilungsfunktion F.

Speziell: Sollen Pseudozufallszahlen gemä0 einer diskreten Verteilung mit Werten $x_1 < x_2 < \cdots << x_l$ und zugehörigen Wahrscheinlichkeiten $p_1, p_2, \ldots, p_l$ erzeugt werden ($\sum\limits_{i=1}^{l} p_i = 1$) so kann dies wie folgt geschehen:
Definiere $q_0 = 0,\ q_j := \sum\limits_{i=1}^{j} p_i,\ 1 \leq j \leq \mathbb{R}$
Ist u_j eine Pseudozufallszahl gemäß Gleichverteilung $[0,1]$, so definiere $y_i := x_j$ falls $u_j \in (q_{i-1}, q_i]$.

b.) Verwerfungsmethode (rejection sampling)
Besitze F eine Dichte f (f sei gut berechenbar).
Wähle weitere Dichte g, so dass für ein $\alpha > 1$

$$f(x) \leq \alpha g(x) \qquad \forall x \in \mathbb{R}$$

Pseudozufallszahlen gemäß der Verteilung mit Dichte g seien leicht zu erzeugen.
Dann liefert folgender Algorithmus Pseudozufallszahlen gemäß der Dichte f.

1. Schritt Erzeuge Pseudozufallszahl z gemäß Dichte g und unabhängig davon Pseudzufallszahl u gemäß gleichverteilung auf $(0,1)$
2. Schritt Falls $u \leq \frac{f(z)}{\alpha g(z)}$, dann verwende Pseudozufallszahl z, anderenfalls gehe zurück zum 1. Schritt.

Die so erhaltenen Pseudozufallszahlen verhalten sich wie Realisation einer Zufallsvariable mit Dichte f, den die Zufallsvariable Z mit Dichte g und U auf (0,1) gleichverteilt als unabhängige Zufallsvariablen

$$P\left(Z \leq y \mid U \leq \frac{f(Z)}{\alpha g(Z)}\right) = \frac{P\{Z \leq y, U \leq \frac{f(Z)}{\alpha g(Z)}\}}{P\{U \leq \frac{f(Z)}{\alpha g(Z)}\}} = \int\limits_{-\infty}^{y} f(z)dz = F(y)$$

wobei

$$P\{Z \leq y, Z \leq \frac{f(Z)}{\alpha g(Z)}\} = \int\limits_{-\infty}^{y} P\{U \leq \frac{f(Z)}{\alpha g(Z)}\} g(z)\, dz = \int\limits_{-\infty}^{y} \frac{f(Z)}{\alpha g(Z)} g(z)\, dz = \frac{1}{\alpha} \int\limits_{-\infty}^{y} f(z)\, dz$$

ebenso für den nenner (mit $y = \alpha$)

$$P\{U \leq \frac{f(Z)}{\alpha g(Z)}\} = \frac{1}{\alpha} \int\limits_{-\infty}^{\infty} = \frac{1}{\alpha}, \text{ da f Dichte ist}$$

Die Rechnung zeigt: $\frac{1}{\alpha}$ = Wahrscheinlichkeit, dass die Pseudozufallszahl akzeptiert wird
$\rightsquigarrow \alpha$ soll möglichst nahe bei 1 sein.

Im Mittel werden $2 \cdot \alpha$ Pseudozufallszahlen (gemäß g bzw. Gleichverteilung) benötigt, um eine Pseudozufallszahl gemäß F zu erzeugen.

Beispiel 9.2. *Sei $F = \Phi$, d.h. standartnormalverteilte Pseudozufallszahlen sollen erzeugt werden.*

$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad , \ x \in \mathbb{R}$$

$$g(x) = \frac{1}{2} e^{-|x|} \quad , \ x \in \mathbb{R} \quad : \textit{Dichte der zweiseitigen Exponentialverteilung}$$

zugehörige Verteilungsfunktion:

$$x \leq 0 : G(x) = \int\limits_{-\infty}^{x} g(t) dt = \frac{1}{2} \int\limits_{-\infty}^{x} e^t dt = \frac{1}{2} e^t \ |_{-\infty}^{x} = \frac{1}{2} e^x$$

$$x > 0 : G(x) = \int\limits_{-\infty}^{0} g(t) dt + \int\limits_{0}^{x} g(t) dt = \frac{1}{2} + \frac{1}{2} \int\limits_{0}^{x} e^{-t} dt = 1 - \frac{1}{2} e^{-x}$$

G ist invertierbar mit der Inversen

$$G^{\leftarrow}(t) = \begin{cases} \log(2t) & t \leq \frac{1}{2} \\ -\log(2(1-t)) & t > \frac{1}{2} \end{cases}$$

$$\frac{\varphi(x)}{g(x)} = \sqrt{\frac{2}{\pi}} e^{|x| - \frac{x^2}{2}} \ \textit{ist symmetrisch um 0 und für } \lambda \geq 0 \textit{ gilt}$$

$$\left(\frac{\varphi(x)}{g(x)}\right)' = \sqrt{\frac{2}{\pi}} (1-x) e^{x - \frac{x^2}{2}} \stackrel{!}{=} 0$$

$$\Leftrightarrow x = 1 \ \textit{und} \ \lim_{x \to \infty} \frac{\varphi(x)}{g(x)} = 0 \ \textit{und} \ \frac{\varphi(0)}{g(0)} = \sqrt{\frac{2}{\pi}}$$

$\Rightarrow x = 1$ *ist Maximalstelle von* $\frac{\varphi}{g}$ *mit zugehörigem Maximum* $:\sqrt{\frac{2e}{\pi}} =: \alpha \leq 1,32$

Also $f(x) \leq \alpha g(x) \qquad \forall x \in \mathbb{R}$
Erzeuge Pseudozufallszahlen gemäß $\mathcal{N}_{(0,1)}$ *wie folgt:*

1. Schritt *Erzeuge unabhängige auf (0,1) gleichverteilte Pseudozufallszahlen* u, v *und* $z := G^{\leftarrow}(v)$

2. Schritt *Akzeptiere* z *als Pseudozufallszahl, falls* $u \leq \frac{f(z)}{\alpha g(z)} = e^{|z| - \frac{z^2}{2} - \frac{1}{2}}$, *sonst zurück zum 1. Schritt.*