Einführung in die Stochastik

Mitschrift zur Vorlesung von Prof. Kamps

an der RWTH-Aachen

Sommersemester 2004

von Cristian Wente

Aachen, 8. Februar 2006

Hinweis:
Die Nummerierungen der Sätze, Definitione, etc.
stimmen nicht immer mit der Vorlesung überein.

Auch enthält dieses Skript Ergänzungen und Beispiele, die nicht angesprochen wurden.

Die einzige Verpflichtung die der Autor gegenüber Fehlern annimmt ist die,
diese sofort zu korrigieren, wenn sie ihm bekannt gemacht werden.

Dieses Skript ist NICHT fehlerfrei.
Es hat noch KEINE Revision durch Prof. Kamps stattgefunden

Vielen Dank an folgende Personen, die mir geholfen haben dieses Skript zu gestalten:

- Sascha Beckers
- Michael Arens
- Annika Günther
- Jörn Wübker
- Thorsten Wessling

Inhaltsverzeichnis

0 Ziele und Aufgaben der Stochastik

- Modelle für reale zufallsabhängige Vorgänge
  $\hookrightarrow$ Aussagen im Modell
  $\hookrightarrow$ Entscheidungshilfe durch „Rückübersetzung“ in die Realität

- Datenanalyse (beschreibende/schließende Statistik)

Beispiel 0.1 *Netzwerkanalyse, Warteschlangen, Analyse von Algorithmen, Simulation von Systemen*

zufallsabhängige Vorgänge $\longleftrightarrow$ deterministische Vorgänge (Ursache$\rightarrow$Wirkung)
$\hookrightarrow$ prinzipiell ein Ereignis nicht vorhersagbar, weil „Zufall“ eingreift:

- Ergebnisse von Glücksspielen

- Lebensdauer von Systemen

- falsch übertragene Bits

oder
Vorhersage prinzipiell möglich, aber zu komplex:

- Zahl der Einflußgrößen zu hoch

- Probleme bei der Quatifizierung der Einflußgrößen

Ziel:

> Sicherheit über die Unsicherheit gewinnen
> im folgenden Sinn:
> Einzelversuch ist nicht vorhersagbar, aber Aussagen bei häufiger Versuchswiederholung
> $\hookrightarrow$ Datenerhebung, Simulation

Beispiel 0.2 (Würfelwurf) *mögliche Ergebnisse kodiert durch* $1, \ldots, 6$
Stichprobenraum $\Omega = \{1, \ldots, 6\}$
Ergebnis: $\omega \in \Omega$ *(*Elementarereignis*)*
„Klar“: gleiche Chance für jede Zahl:

$$P(\{\omega\}) = \frac{1}{6} \qquad , \forall \omega \in \Omega$$

Andere Fragestellung: „Es fällt eine gerade Zahl“
$\hookrightarrow$ *beschreibbar durch Teilmenge* $A := \{2, 4, 6\} \subset \Omega$
„Klar“: $P(A) = \frac{1}{2}$
$\hookrightarrow$ *Modell:*

$$\begin{aligned} \Omega, \mathfrak{P}(\Omega) &= \{A | A \subseteq \Omega\} \\ P : \mathfrak{P}(\Omega) &\rightarrow [0, 1] \\ \textit{mit} \quad P(A) = \tfrac{|A|}{6} & \qquad , \forall A \subseteq \Omega \end{aligned}$$

Zu überprüfen: Übereinstimmung von Modell und Realität durch Experiment (Statistik)

1 Diskrete Wahrscheinlichkeitstheorie

1.1 Diskrete Wahrscheinlichkeitsräume und deren Erweiterung

Definition 1.1 *Sei Ω eine höchstens abzählbar große Menge,*
$\mathfrak{A} := \mathfrak{P}(\Omega) \ := \{A | A \subseteq \Omega\}$ *und*
$p : \Omega \to [0,1]$ *eine Abbildung mit* $\sum\limits_{\omega \in \Omega} p(\omega) = 1$
Die durch $P(A) = \sum\limits_{\omega \in A} p(\omega) \qquad , A \subseteq \Omega,$
definierte Abbildung heißt diskrete WAHRSCHEINLICHKEITSVERTEILUNG *über* Ω
Die Funktion p heißt ZÄHLDICHTE.
Das Tripel $(\Omega, \mathfrak{A}, P)$ *heißt diskreter* WAHRSCHEINLICHKEITRAUM.
$(\Omega, \mathfrak{A})$ *heißt diskreter* MESSBARER RAUM

Bezeichnung 1.2 Ω *heißt* GRUNDMENGE, ERGEBNISRAUM, STICHPROBENRAUM
$A \subseteq \Omega$ *heißt* EREIGNIS *und speziell* ELEMENTAREREIGNIS *falls* $|A| = 1$
`Kurzschreibweise`*:*

$$P(\{\omega\}) \overset{Abk.}{=} P(\omega) = p(\omega) \qquad , \omega \in \Omega$$

Lemma 1.3

1. *Gegeben sei die Situation aus (1.1). Dann gilt:*
   (a) $0 \leqslant P(A) \leqslant 1 \qquad , A \in \mathfrak{P}(\Omega)$
   (b) $P(\Omega) = 1$
   (c) *P ist σ-additiv, d.h. für paarweise disjunkte* $A_1, A_2, \ldots \in \mathfrak{P}(\Omega)$
   (mit $A_i \cap A_j = \emptyset$ für $i \neq j$) gilt:

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

   insbesonders ist $P(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i) \qquad , \forall n \in \mathbb{N}$
   und $P(\Omega) = \sum\limits_{\omega \in \Omega} p(\omega) = 1$

2. *Sei $P : \mathfrak{P}(\Omega) \ \to \mathbb{R}$ eine Abbildung mit obigen Eigenschaften, dann gibt es genau eine Funktion* $p : \Omega \to [0,1]$ *mit* $P(A) = \sum\limits_{\omega \in A} p(\omega) \qquad , \forall A \subseteq \Omega$

Beispiel 1.1 *Der Laplace-raum*
Sei $\Omega = \{1, \ldots, n\}, p(\omega) = \frac{1}{n} \qquad , \forall \omega \in \Omega$
P heißt dann diskrete GLEICHVERTEILUNG *oder* LAPLACE-VERTEILUNG.
Es ist $P(A) = \frac{|A|}{|\Omega|} = \frac{1}{n}|A| \qquad , \forall A \subseteq \Omega$
$\hookrightarrow$ *„Anzahl der günstigen Fälle dividiert durch die Anzahl aller Möglichen"*

Beispiel 1.2 *Problem des Chevalier de Néré 17.Jhdt.*
Was ist bei 3 Würfelwürfen wahrscheinlicher, Summe 11 oder 12 ?
Von Interesse:

Ereignis A *Summe der Augen ist 11*

Ereignis B *Summe der Augen ist 12*

Informal:

A				B			
6	4	1	(6)	6	5	1	(6)
6	3	2	(6)	6	4	2	(6)
5	5	1	(3)	6	3	3	(3)
5	4	2	(6)	5	5	2	(3)
5	3	3	(3)	5	4	3	(6)
4	4	3	(3)	4	4	4	(1)
$\sum = 11$			(27)	$\sum = 12$			(25)

jeweils 6 Fälle

Modell: $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3) | \omega_i \in \{1, \dots, 6\}, i \in \{1, 2, 3\}\}$

$$\begin{aligned} A &= \{\omega \in \Omega | \omega_1 + \omega_2 + \omega_3 = 11\} \\ B &= \{\omega \in \Omega | \omega_1 + \omega_2 + \omega_3 = 12\} \end{aligned}$$

$$p(\omega) = \frac{1}{|\Omega|} = \frac{1}{6^3} \qquad , \forall \omega \in \Omega$$

Also: Ω *ist eine Menge von Tripeln, d.h. Würfel sind unterscheidbar*
Abzählen liefert: $|A| = 27, |B| = 25$
Also: $P(A) = \sum\limits_{\omega \in \Omega} p(\omega) = \frac{27}{6^3} > \frac{25}{6^3} = P(B)$

Viele Argumente der Wahrscheinlichkeitrechnung auf Laplace-räumen reduzieren sich auf Kombinatorik und „geschicktes“ Abzählen
$\hookrightarrow$ Wahl des „richtigen“ Modells und Zählweise
(siehe Kapitel 2)

Erweiterung und Ausblicke Ausgangspunkt: σ-Additivität in (1.3)
zur Modellierung gewünscht:

$$\begin{aligned} \Omega = &\ [0,1] && \text{(Zinssatz)} \\ \Omega = &\ \mathbb{R} && \text{(Abweichung vom „Sollwert“)} \\ \Omega = &\ \mathbb{R}^+ && \text{(Lebensdauer eines Systems)} \end{aligned}$$

$\hookrightarrow$ Probleme, falls Ω überabzählbar !
Es gibt keine „Gleichverteilung“ über der Potenzmenge von $\Omega = [0, 1]$, d.h.
es gibt keine Wahrscheinlichkeitsverteilungmit der Eigenschaft

$$P(A + h) = P(A) \qquad , \forall A, A + h \in \mathfrak{P}(\Omega)$$

wobei $A + h := \{a + h | a \in A\}, h \in \mathbb{R}$

Konsequenz aus der Mathematik:
Beschränkung auf ein geeignetes Mengensystem, das echt enthalten ist in der Potenzmenge $\mathfrak{P}(\Omega)$ falls σ-Additivität weiter gefordert ist.
Geeignete Struktur: σ-Algebra

1.2 σ-Algebra

Definition 1.4 *Sei $\Omega \neq \emptyset$ und $\mathfrak{A} \subset \mathfrak{P}(\Omega)$ ein System von Teilmengen von Ω. $\mathfrak{A}$ heißt σ-*Algebra *von Ereignissen über Ω, falls gilt:*

1. $\Omega \in \mathfrak{A}$
2. $A \in \mathfrak{A} \Rightarrow A^C \in \mathfrak{A} \qquad , \forall A \in \mathfrak{A}$
3. $(A_n)_{n \in \mathbb{N}} \subseteq \mathfrak{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$

D.h. eine σ-Algebra ist abgeschlossen gegenüber der Bildung von Komplementen und abzählbaren Vereinigungen.
Bemerkung: Als Elementarereignis (vgl (1.2)) bezeichnet man in diesem Zusammenhang eine Menge aus $\mathfrak{A}$, die keine echte Vereinigung anderer Ereignisse ist.

Lemma 1.5

1. *Sei $\mathfrak{A}$ eine σ-Algebra, so gilt:*
   (a) $\emptyset \in \mathfrak{A} \qquad \emptyset = \Omega^C \in \mathfrak{A}$
   (b) $A, B \in \mathfrak{A} \Rightarrow A \cap B \in \mathfrak{A} \qquad A \cap B = (A^C \cup B^C)^C$
   (c) $(A_n)_{n \in \mathbb{N}} \subseteq \mathfrak{A} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathfrak{A}$ *Induktion mit 2*
2. *Seien $B \subset \Omega$ und $\mathfrak{A}$ eine σ-Algebra über Ω, dann gilt:*
   $B \cap \mathfrak{A} := \{B \cap A | A \in \mathfrak{A}\}$
   ist eine σ-Algebra über B (Spur-σ-Algebra)
3. *Sei $\Omega \neq \emptyset$*
   $$\mathfrak{A} = \{A \subset \Omega | A \text{ oder } A^C \text{ ist höchstens abzählbar über } \Omega\}$$
   ist eine σ-Algebra über Ω

Beweis zu 2,3 in der Übung
Bemerkung:

$\mathfrak{P}(\Omega)$ ist eine σ-Algebra über Ω
$\mathfrak{P}(\Omega)$ ist die feinste, $\mathfrak{A} = \{\emptyset, \Omega\}$ die gröbste σ-Algebra über Ω.

Man kann zeigen:

Es gibt stets eine kleinste σ-Algebra, die eine gegebenes System $\mathcal{F}$ von „einfachen“ Mengen enthält.

Satz 1.6 *Seien $\Omega \neq \emptyset, \mathcal{F} \subset \mathfrak{P}(\Omega)$*
Die kleinste σ-Algebra, die $\mathcal{F}$ enthält, ist gegeben durch

$$\mathfrak{A}(\mathcal{F}) := \{A \in \mathfrak{P}(\Omega) \mid \textit{für jede } \sigma\textit{-Algebra } \mathfrak{A} \textit{ mit } \mathcal{F} \subset \mathfrak{A} \textit{ gilt } A \in \mathfrak{A}\}$$

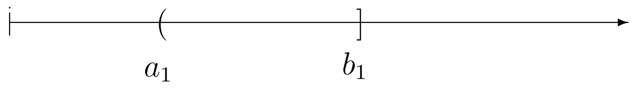
$\mathfrak{A}(\mathcal{F})$ heißt die von $\mathcal{F}$ erzeugte σ-Algebra.

Oft zur Modellierung benötigt: $\Omega = \mathbb{R}^n$
Wähle $\mathcal{F}$ als Menge aller nach links halboffenen Intervalle

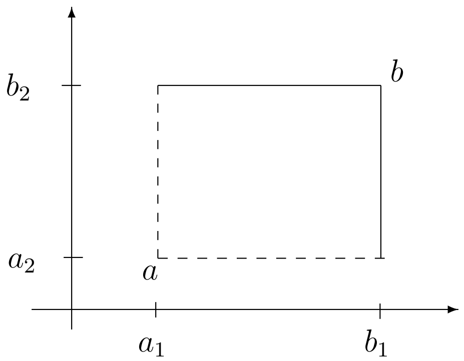
$$(a, b] = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n | a_i < x_i \leqslant b_i, 1 \leqslant i \leqslant n\}$$

für $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n) \in \mathbb{R}^n$

n=1:



n=2:



Zugehörige σ-Algebra: $\mathfrak{A}(\mathcal{F}) =: \mathfrak{B}^n$ (Borel'sche σ-Algebra)
Bemerkung:

Alle offenen und abgeschlossenen Mengen liegen in $\mathfrak{B}^n$, aber es gilt $\mathfrak{B}^n \neq \mathfrak{P}(\mathbb{R}^n)$

1.3 Wahrscheinlichkeitsraum und -verteilung

Definition 1.7 (Kolmogorov 1933) *Sei $\mathfrak{A}$ eine σ-Algebra über $\Omega \neq \emptyset$ und eine Abbildung $P : \mathfrak{A} \to [0, 1]$ mit*

1. $P(A) \geqslant 0, \quad \forall A \in \mathfrak{A}$
2. $P(\Omega) = 1$
3. $P(\bigcup A_n) = \sum P(A_n)$ *für alle paarweise disjunkten Mengen $A_i \in \mathfrak{A}$ (σ-Additivität)*

heißt Wahrscheinlichkeitsverteilung *oder* Wahrscheinlichkeitmass *auf $\mathfrak{A}$.*
$(\Omega, \mathfrak{A}, P)$ heißt Wahrscheinlichkeitraum*, $(\Omega, \mathfrak{A})$ heißt* messbarer Raum*.*

Bemerkung:

Vergleiche mit diskretem Wahrscheinlichkeitraum:
Wahrscheinlichkeitsverteilung ist festgelegt durch $P(\omega)$
der Elementarereignisse ω.
Zugang zu allgemeinen σ-Algebren ist in folgendem Sinne nicht schwierig:
Ist $P(A)$ festegelgt $\forall A \in \mathcal{F}$ mit Eigenschaften 1-3 so folgt (nach Maßtheorie)
$P(B)$ ist eindeutig bestimmt für alle $B \subset \mathfrak{A}(\mathcal{F})$

Zur Vereinfachung zunächst:
Ω höchstens abzählbar und $\mathfrak{A} = \mathfrak{P}(\Omega)$, die in Kapitel 3 hergeleiteten Regeln zum Rechnen mit Wahrscheinlichkeiten gelten allgemein.

1.4 Einpunktverteilung

Definition 1.8 *Sei $\Omega \neq \emptyset$ abzählbar, $\omega \in \Omega$ fest.*
Die durch $\varepsilon_\omega(A) = \begin{cases} 1 & \text{, falls } \omega \in A \\ 0 & \text{, sonst} \end{cases}$
festgelegte Wahrscheinlichkeitsverteilung $\varepsilon_\omega : \mathfrak{P}(\Omega) \to \mathbb{R}$ *heißt* DIRAC-VERTEILUNG *oder* EINPUNKTVERTEILUNG *in ω.*

Bemerkung: ε_ω ist Wahrscheinlichkeitsverteilung, denn:

1. $P(A) \geqslant 0$ (Klar)
2. $P(\Omega) = 1$ (denn $\omega \in \Omega$)
3. $(A_n)_n \subset \mathfrak{A}$ paarweise disjunkt
   $\Rightarrow$ es gibt höchstens ein $i \in \mathbb{N} : \omega \in A_i$

$$\begin{aligned} \varepsilon_\omega(\bigcup A_n) &= \begin{cases} 1 & \text{, falls } \omega \in \bigcup A_n \\ 0 & \text{, sonst} \end{cases} \\ &= \begin{cases} 1 & \text{, falls es ein } i \text{ gibt mit } \omega \in A_i \\ 0 & \text{, sonst} \end{cases} \\ &= \sum \varepsilon_\omega(A_n) \end{aligned}$$

1.5 Träger von Wahrscheinlichkeitsverteilung

Definition 1.9 *Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitraum.*

$$T := supp(P) := \{\omega \in \Omega | P(\omega) > 0\}$$

heißt TRÄGER *von P*

Lemma 1.10 *Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitraum und T Träger von P. Dann gilt*

$$P(A) = \sum_{\omega \in T} P(\omega)\varepsilon_\omega(A) \quad , A \in \mathfrak{A}$$

d.h. P ist darstellbar als gewichtete Summe von Einpunktverteilungen.

Beweis:

Sei $A \in \mathfrak{A}$

$$\begin{aligned}
\sum_{\omega \in T} P(\omega)\varepsilon_\omega(A) &= \sum_{\omega \in \Omega} P(\omega)\varepsilon_\omega(A) \\
&= \sum_{\omega \in A} P(\omega)\underbrace{\varepsilon_\omega(A)}_{=1} + \sum_{\omega \in A^C} P(\omega)\underbrace{\varepsilon_\omega(A)}_{=0} \\
&= \sum_{\omega \in A} P(\omega) \\
&= P(A)
\end{aligned}$$

2 Grundformeln der Kombinatorik

Kombinatorik: „Kunst des Zählens“
in diesem Fall die Bestimmung der Mächtigkeit von Mengen

Beispiel 2.1 *Speisekarte: 4 Vorspeisen, 3 Hauptgerichte, 3 Nachspeisen*
Anzahl der unterschiedlichen Menus: $3 * 4 * 3 = 36$

Beispiel 2.2 *4 Kneipen: Besuchsplan für die nächsten 3 Tage (1 Kneipe/Tag)*

1. *mehrfacher Besuch möglich:* $4^3 = 64$ *Pläne*
2. *jede Kneipe höchstens einmal:* $4 * 3 * 2 = 24$ *Pläne*

Bemerkung:

Urnenmodell als Hilfsmittel zum Laplace-Experiment:
Urne mit nummerierten Kugeln $(1, \dots, n)$
$\hookrightarrow$ sukzessives, zufälliges Ziehen einer Kugel, dabei
Ziehen

1. mit Zurücklegen (mit Wiederholung)
2. ohne Zurücklegen (ohne Wiederholung)

Stichprobe

1. mit zeitlicher Reihenfolge (Tupel)
2. ohne Reihenfolge (lexikographische Ordnung)

Satz 2.1 *Ziehen mit Zurücklegen in Reihenfolge*
Realisierung durch k-Tupel $\omega = (\omega_1, \dots, \omega_k)$, *wobei*
ω_j: *Die Nummer der j-ten gezogenen Kugel*
Aus Symmetriegründen ist jedes ω *gleichwahrscheinlich.*

$$\begin{aligned} \Omega_1 &= \{\omega = (\omega_1, \dots, \omega_k) | \omega_i \in \{1, \dots, n\}, i \in \{1, \dots, k\}\} \\ &=: \{1, \dots, n\}^k \end{aligned}$$

mit $|\Omega_1| = n^k$ *(vgl. Bsp 2.2.1)*

Satz 2.2 *Ziehen ohne Zurücklegen in Reihenfolge*
Realisierung durch k-Tupel mit verschiedenen Einträgen

$$\begin{aligned} \Omega_2 &= \{\omega \in \{1, \dots, n\}^k | \omega_i \neq \omega_j \text{ für } 1 \leqslant i \neq j \leqslant k\} \\ |\Omega_2| &= n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!} =: (n)_k \end{aligned}$$

Speziell: $n = k \Rightarrow |\Omega_2| = n!$ *und* Ω_2 *ist Menge aller Permutationen von* $1, \dots, n$

Satz 2.3 *Ziehen ohne Zurücklegen ohne Reihenfolge*
Lexikographische Ordnung:

$$\begin{aligned} \Omega_3 &= \{\omega \in \{1,\ldots,n\}^k | \omega_1 < \omega_2 < \ldots < \omega_k\} \\ &\quad \textit{oder alternativ} \\ \Omega_3' &= \{A \subset \{1,\ldots,n\} | \quad |A| = k\} \end{aligned}$$

Bijektion zwischen Ω_3 und Ω_3':

$$\omega \mapsto \{\omega_1,\ldots,\omega_k\}$$

$$|\Omega_3| = \binom{n}{k} = \frac{n!}{k!(n-k)!} = |\Omega_3'|$$

Betrachte Abbildung $f : \Omega_2 \to \Omega_3'$ mit

$$f : (\omega_1,\ldots,\omega_k) \mapsto \{\omega_1,\ldots,\omega_k\}$$

Jedes Urbild $f^{-1}(y) = \{\omega \in \Omega_2 | \ f(\omega) = y\}$ von $y \in \Omega_3'$ hat genau $k!$ Elemente

$$|\Omega_3'| = \binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{|\Omega_2|}{k!}$$

Merkregel:
Die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge ist $\binom{n}{k}$

Satz 2.4 *Ziehen mit Zurücklegen ohne Reihenfolge*
Realisierung ist k-Tupel mit aufsteigend geordneten Koordinaten:

$$\begin{aligned} \Omega_4 &= \{\omega \in \{1,\ldots,n\}^k | \omega_1 \leqslant \omega_2 \leqslant \ldots \leqslant \omega_k\} \\ |\Omega_4| &= \binom{n+k-1}{k} \end{aligned}$$

Betrachte folgende Abbildung:

$$(\omega_1,\ldots,\omega_k) \mapsto (\omega_1, \omega_2+1, \ldots, \omega_k+k-1)$$

Diese ist bijektive von Ω_4 in einen neuen Stichprobenraum $\widehat{\Omega}_3$ eines Modells mit $n+k-1$ nummerierten Kugeln, von denen k Kugeln gezogen wurden:

$$\begin{aligned} \widehat{\Omega}_3 &= \{(\omega_1',\ldots,\omega_k') \in B^k | \omega_1' < \omega_2' < \ldots < \omega_k'\} \\ \textit{mit } B &= \{1,\ldots,n+k-1\} \textit{ und } \omega_i' = \omega_i + i - 1 \end{aligned}$$

$\Rightarrow |\Omega_4| = |\widehat{\Omega}_3| = \binom{n+k-1}{k}$

Bemerkung: andere Bezeichnungen:

1. (n,k)-Permutationen aus Ω mit Wiederholung
2. (n,k)-Permutationen aus Ω ohne Wiederholung
3. (n,k)-Kombinationen aus Ω ohne Wiederholung
4. (n,k)-Kombinationen aus Ω mit Wiederholung

Satz 2.5 *Die Urnenmodelle 1-3 liefern Laplace-Räume, aber 4 nicht !*
Siehe Beispiel 1.2

2.1 Übersicht

	k-mal Ziehen aus n Kugeln	mit Zurücklegen	ohne Zurücklegen	
Permutation	mit Reihenfolge	2.1 $\lvert\Omega_1\rvert = n^k$	2.2 $\lvert\Omega_2\rvert = (n)_k$	unterscheidbare Murmeln
Kombination	ohne Reihenfolge	2.4 $\lvert\Omega_4\rvert = \binom{n+k-1}{k}$	2.3 $\lvert\Omega_3\rvert = \binom{n}{k}$	un-unterscheidbare Murmeln
		mit Mehrfachbelegung	ohne Mehrfachbelegung	k Murmeln auf n Urnen verteilen

Satz 2.6 *Interpretation über das Verteilen von Murmeln*
Anzahl der Möglichkeiten k unterscheidbare Murmeln auf n Plätze zu verteilen
i-te Murmel in Zelle ω_i, $1 \leqslant i \leqslant k, \omega_i \in \{1, \dots, n\}$

Urnenmodell		*Murmelmodell*
Nummer der Ziehung	$\widehat{=}$	*Nummer der Murmel*
Nummer der Kugel	$\widehat{=}$	*Nummer der Zelle*

$\hookrightarrow$ *Fall Ω_1 bei Mehrfachbelegung, usw.*

Beispiel 2.3

1. **(Geburtstagsproblem)**
   Gesucht:

   Wahrscheinlichkeit mit der von k zufällig gewählten Personen mindestens 2 am selben Tag des Jahres Geburtstag haben ($= p_k$)

   Modell:

   Jahr mit 365 Tagen,
   Geburten gleichwahrscheinlich an allen Tagen des Jahres
   $\hookrightarrow$ Ω_1 *mit $n = 365$ mit n^k Elementen*
   Ziehen von k Daten mit Zurücklegen

   Ereignis E_k kein Geburtstag doppelt
   Bekannt: $P(A) + P(A^C) = 1$

   $$P(E_k) = \frac{(n)_k}{n^k} = \frac{(365)_k}{365^k} = q_k = 1 - p_k$$

   Speziell: $p_{10} \approx 0,12, \quad p_{23} \approx 0,51, \quad p_{50} \approx 0,97$

2. **(Hashing)**
   Problem:

   - *Teilmenge eines Ganzen mit Hilfe geeigneter Hash-Tafeln abspeichern*

- *Zugriff auf derartige Teilmengen*

Vorgehen

Zufälliges Ablegen von k Daten in einem Feld der Länge n ($\geqslant k$)
Mehrfachbelegungen führen zur Kollision

Sei $A_{k,n} \widehat{=}$ Kollision findet statt
$\hookrightarrow A_{k,n}^C \rightarrow$ *Unterscheidbare Murmeln ohne Mehrfachbelegung*

$$P(A_{k,n}^C) = \frac{(n)_k}{n^k} = \prod_{i=0}^{k-1}(1-\frac{i}{n})$$

Beispiel 2.4 (Lotto)
$k = 6$ *Kugeln aus* $n = 49$ *ohne Zurücklegen*

$$\left.\begin{array}{rcl} \textit{mit Reihenfolge} & \rightarrow & \Omega_2 \\ \textit{ohne Reihenfolge} & \rightarrow & \Omega_3 \end{array}\right\} \text{„}\textit{Freiheit}\text{“ } \textit{der Modellwahl}$$

Fragestellung „3te gezogene Kugel ist die 49“ läßt sich nur in Ω_2 behandeln!
Sei $p_k := P(k$ Richtige$)$

$$p_6 = \frac{1}{\binom{49}{6}} = \frac{1}{13.983.816}$$

p_4 *in* Ω_3:

Sei $\omega' = \{\omega'_1, \ldots, \omega'_6\}$ die Menge der geratenen Zahlen.

$$E_4 = \{\omega \in \Omega_3 | \quad |\{\omega_1, \ldots, \omega_6\} \cap \{\omega'_1, \ldots, \omega'_6\}| = 4\}$$

Dann sind 4 Kugeln von ω' fest (die „richtigen“ Kugeln), $\binom{6}{4}$ als Anzahl der Möglichkeiten und dann 2 Kugeln von $\{1, \ldots, 49\}\backslash\omega'$, dies sind $\binom{43}{2}$
$\hookrightarrow$ *alle Kombinationen sind möglich, also $\binom{6}{4}\cdot\binom{43}{2}$ ist die Anzahl der günstigen Möglichkeiten*

$$p_4 = \frac{\binom{6}{4}\cdot\binom{43}{2}}{\binom{49}{6}}$$

(Beweis siehe Übung)

2.2 Die hypergeometrische Verteilung

Beispiel 2.5 (Schwarze und weiße Kugeln) *Das vorige Beispiel war ein Spezialfall der hypergeometrischen Verteilung.*
Betrachte eine Urne mit S schwarzen Kugeln und W weißen Kugeln, $n = W + S$
Ziehe $k(\leqslant n)$ Kugeln ohne Zurücklegen.
Gesucht ist die Wahrscheinlichkeit, dass die Stichprobe genau s schwarze und $k - s = w$ weiße Kugeln enthält.

$$h(\underbrace{s}_{\textit{Variable}}; \underbrace{k, n, S}_{\textit{Parameter}}) := \frac{\binom{S}{s}\cdot\binom{W}{w}}{\binom{S+W}{k}} = \frac{\binom{S}{s}\cdot\binom{n-S}{k-s}}{\binom{n}{k}}, \quad s \leqslant k$$

Definition 2.7 *Die (obige) durch h bestimmte Zähldichte*

$$(h(0) + h(1) + \ldots + h(k) = 1)$$

definiert die HYPERGEOMETRISCHE VERTEILUNG

Anwendung: Qualitätskontrolle

Warenstichprobe bei Gut-Schlecht-Prüfungen. Lieferung von n Teilen (intakt oder defekt) mit S defekten und $n - S$ intakten Teilen.
Kontrolle: Stichprobe vom Umfang $k \leqslant n$ ohne Zurücklegen
Gesucht ist die Wahrscheinlichkeitfür s defekte Teile in der Stichprobe: $h(s; k, n, S)$

Beispiel 2.6 *Fische zählen*
Frage: Anzahl der Fische in einem Teich
Fange S Fische, markiere diese und setze wieder ein.
Warten (d.h. „mischen")
Fange k Fische
(in der Biologie als capture-recapture-Verfahren bezeichnet)
Intuitiv ergibt sich das Verhältnis $\frac{S}{n} \approx \frac{s}{k}$
d.h. „schätze" für $n = k \cdot \frac{S}{s}$
Im Rahmen der mathematischen Statistik bedeutet $h(s; k, n, S)$ die Wahrscheinlichkeit für das Fangen von s markierten Fischen.

Prinzip der Maximum-Likelihood-Schätzung:
Wähle n so, dass bei Ergebnis s die Wahrscheinlichkeit $h(s; k, n, S)$ für diese Realisation maximal wird
Ergebnis: $\widehat{n} = k \cdot \frac{S}{s}$

Fragen in der Statistik:

$\widehat{n}$: „Schätzung" in welchem Sinn ?
Eigenschaften der Schätzung ?

Bemerkung:

Intervall angeben, so dass „wahrer" Wert mit vorgegebener Vertrauenswahrscheinlichkeit in diesem Intervall liegt ($\hookrightarrow$ Konfidenzintervall)

2.3 Binomialverteilung

Beispiel 2.7 *Anwendung wie bei Lotto, aber mit Zurücklegen, mit Reihenfolge.*
Anzahl der möglichen Stichproben: n^k
Gesucht ist die Anzahl der Stichproben, die s defekte Teile enthalten.
Es gibt S^s *Möglichkeiten,* s *defekte Teile aus* S *auszuwählen,* $(n-S)^{k-s}$ *Möglichkeiten* $k-s$ *intakte Teile aus* $n-S$ *auszuwählen.*
Die Anzahl der Möglichkeiten, s *defekte Teile auf* k *Plätze zu verteilen ist* $\binom{k}{s}$
Damit ergibt sich die Wahrscheinlichkeit für s *defekte Teile* ($s \in \{0, \ldots, k\}$)*:*

$$\begin{aligned} b\left(s; k, \frac{S}{n}\right) &= \frac{1}{n^k} \cdot \binom{k}{s} \cdot S^s \cdot (n-S)^{k-s} \\ &= \binom{k}{s} \left(\frac{S}{n}\right)^s \left(1 - \frac{S}{n}\right)^{k-s} \\ &= \binom{k}{s} p^s (1-p)^{k-s} \\ & \quad \textit{wobei } p := \frac{S}{n} \textit{ (Schlecht-Anteil)} \end{aligned}$$

Definition 2.8 *Die durch* b *bestimmte Zähldichte (als Funktion von* s*) definiert die sogenannte* Binomialverteilung, *kurz:* $b(s; k, p)$ *oder* $b(k, p)$
Probe:

$$\sum_{s=0}^{k} \binom{k}{s} p^s (1-p)^{k-s} \stackrel{\textit{Binomialsumme}}{=} p \cdot (1-p)^k = 1$$

Intuitiv folgt, dass für große n *kaum Unterschied zwischen Ziehen mit und ohne Zurücklegen besteht*

Lemma 2.9 *Sei* $(S_n)_{n \in \mathbb{N}}$ *mit* $\lim\limits_{n \to \infty} \frac{S_n}{n} = p \in (0, 1)$.
Dann gilt:

$$\lim_{n \to \infty} h(s; k, n, S_n) = b(s; k, p)$$

Eine Produktion enthalte (wie oben) den Anteil p defekter Teile.
Dann ist die Wahrscheinlichkeit für s defekte Teile in einer Stichprobe vom Umfang k (mit Zurücklegen)

$$b(s; k, p) = \binom{k}{s} p^s (1-p)^{k-s}$$

Was passiert mit der Binomialverteilung bei $k \to \infty$ und $p \to 0$?

Lemma 2.10 *Gegeben sei eine Folge von Binomialverteilungen* $b(s; k, p_k) k \in \mathbb{N}$ *mit* $k \cdot p_k = \lambda > 0 \quad \forall k \in \mathbb{N}$.
Dann gilt:

$$\lim_{k \to \infty} b(s; k, p_k) = \frac{\lambda^s}{s!} e^{-\lambda} \qquad \textit{für } s \in \mathbb{N}_0$$

Beweis:

$$\begin{aligned}
\lim_{k\to\infty} \tbinom{k}{s} p^s (1-p)^{k-s} &= \lim_{k\to\infty} \frac{k!}{s!\cdot(k-s)!}\left(\frac{\lambda}{k}\right)^s \left(1-\frac{\lambda}{k}\right)^{k-s} \\
&= \lim_{k\to\infty} \frac{\lambda^s}{s!} \frac{k\cdot(k-1)\cdot\ldots\cdot(k-s+1)}{k^s} \cdot \left(1-\frac{\lambda}{k}\right)^k \cdot \left(1-\frac{\lambda}{k}\right)^{-s} \\
&= \lim_{k\to\infty} \frac{\lambda^s}{s!} \cdot 1 \cdot \underbrace{\frac{k-1}{k}}_{\to 1} \cdot\ldots\cdot \underbrace{\frac{k-s+1}{k}}_{\to 1} \cdot \underbrace{\left(1-\frac{\lambda}{k}\right)^k}_{\to e^{-\lambda}} \cdot \underbrace{\left(1-\frac{\lambda}{k}\right)^{-s}}_{\to 1} \\
&= \frac{\lambda^s}{s!} \cdot e^{-\lambda}
\end{aligned}$$

Offensichtlich definiert $p(s) := \frac{\lambda^s}{s!} \cdot e^{-\lambda}$ eine Zähldichte auf $\Omega = \mathbb{N}_0$
Sei $\lambda > 0$ beliebig

$$\sum_{s=0}^{\infty} p(s) = e^{-\lambda} \underbrace{\sum_{s=0}^{\infty} \frac{\lambda^s}{s!}}_{=e^{\lambda}} = 1$$

2.4 Poisson-verteilung

Bezeichnung 2.11 *Die Wahrscheinlichkeitsverteilung auf* $\mathbb{N}_0$ *definiert durch die Zähldichte*

$$p(s) := \frac{\lambda^s}{s!} \cdot e^{-\lambda}, s \in \mathbb{N}_0, \lambda > 0$$

heißt POISSON-VERTEILUNG *mit Parameter* λ
kurz: $po(s;\lambda)$ *bzw* $po(\lambda)$

Bemerkung: Lemma 2.9 heißt auch Gesetz der seltenen Ereignisse wegen $p_k \to\ 0, k \to \infty$
Beispiele sind radioaktiver Zerfall, Häufigkeiten von Fehlern in Systemen, Anzahl von Telefonaten,etc.
Dies gilt auch unter der Voraussetzung $k \cdot p_k \overset{k\to\infty}{\longrightarrow} \lambda (> 0)$

Beispiel 2.8 [1] *Ein Kunstsammler fügt seiner Sammlung jedes Jahr zwei neue Bilder hinzu. Seine erste Anschaffung enthielt eine Fälschung, alle weiteren sind Originale.*
Im k-ten Jahr nach der Eröffnung der Sammlung prüfen k Gutachter unabhängig voneinander je ein Bild der Sammlung zutreffend auf Echtheit.
$\Rightarrow$ *die Wahrscheinlichkeit für s entdeckte Fälschungen ist*

$$b\left(s;k,\frac{1}{2k}\right) \text{ mit } k\cdot\frac{1}{2k} = \frac{1}{2} =: \lambda$$

Würde die Sammlung nun unendlich (über Generationen) so weitergeführt, so ist die Wahrscheinlichkeit für s entdeckte Fälschungen

$$\lim_{k\to\infty} b\left(s;k,\frac{1}{2k}\right) = \frac{\lambda^s}{s!}\cdot e^{-\lambda} = \frac{1}{2^s s!} e^{-\frac{1}{2}}$$

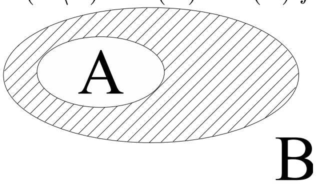
[1] Mit Dank an Annika Günther

3 Eigenschaften eines Wahrscheinlichkeitsraums

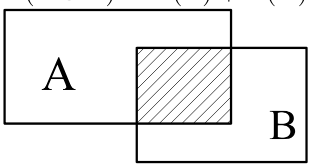
In diesem Abschnitt sei $(\Omega, \mathfrak{A}, P)$ ein allgemeiner Wahrscheinlichkeitraum

Lemma 3.1 *Es gelten für $A, B \in \mathfrak{A}$*

1. $P(A \cup B) = P(A) + P(B)$ *falls* $A \cap B = \emptyset$
2. $P(B \backslash A) = P(B) - P(A)$ *falls* $A \subset B$ *(Subtraktivität)*



3. $P(A^C) = 1 - P(A)$
4. $A \subset B \Rightarrow P(A) \leqslant P(B)$ *Monotonie von* P
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



6. $P(\bigcup_{i=1}^{n} A_i) \leqslant \sum_{i=1}^{n} P(A_i) \quad , A_i \in \mathfrak{A}$

Beweis:

1. Klar aus σ-Additivität
2. $B \backslash A = B \cap A^C \quad , A \subset B$

$$\begin{aligned} P(B) &= P(B \cap \Omega) \\ &= P(B \cap (A \cup A^C)) \\ &= P(\underbrace{(B \cap A)}_{=A} \cup (B \cap A^C)) \text{ (disjunkt)} \\ &\stackrel{1}{=} P(A) + P(B \cap A^C) \end{aligned}$$

3. ergibt sich aus 2 mit $B = \Omega$
4. $P(A) \stackrel{2}{=} P(B) - P(B \backslash A) \leqslant P(B)$ mit $A \subset B$
5. Wegen σ-Additivität gilt:

$$\begin{aligned} P(A^C \cap B) + P(A \cap B) &\stackrel{2}{=} P(B) \\ P(B^C \cap A) + P(A \cap B) &\stackrel{2}{=} P(A) \qquad (*) \end{aligned}$$

$$\text{Nun ist } A \cup B = \underbrace{(A \cap B^C) \cup (A \cap B) \cup (A^C \cap B)}_{\text{alle disjunkt}}$$

$$\begin{aligned}
\Rightarrow P(A \cup B) &\overset{1}{=} P(A \cap B^C) + P(A \cap B) + P(A^C \cap B) \\
&\overset{*}{=} P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\
&= P(A) + P(B) - P(A \cap B)
\end{aligned}$$

6. Trick: Ereignis disjunkt aufschreiben und Additivität ausnutzen.

$$\begin{aligned}
\text{Es ist } \bigcup_{i=1}^{n} A_i &= A_1 \cup (A_1^C \cap A_2) \cup (A_1^C \cap A_2^C \cap A_3) \cup \ldots \cup (\bigcap_{j=1}^{n-1} A_j^C \cap A_n) \\
&= A_1 \cup \bigcup_{i=2}^{n} ((\bigcap_{j=1}^{i-1} A_j^C) \cap A_i) \\
\Rightarrow P(\bigcup_{i=1}^{n} A_i) &= P(A_1) + \sum_{i=2}^{n} P((\bigcap_{j=1}^{i-1} A_j^C) \\
&\overset{4}{\leqslant} P(A_1) + \sum_{i=2}^{n} P(A_i) = \sum_{i=1}^{n} P(A_i)
\end{aligned}$$

3.1 Limes superior und inferior

Definition 3.2 *Sei* $(A_n)_{n\in\mathbb{N}} \subseteq \mathfrak{A}$, $\mathfrak{A}$ *sei eine* σ*-Algebra über* $\Omega \neq \emptyset$

$$(A_n)_n \text{ heißt MONOTON} \begin{cases} \text{WACHSEND} & \text{, falls } A_n \subseteq A_{n+1} \quad \forall n \in \mathbb{N} \\ \text{FALLEND} & \text{, falls } A_n \supseteq A_{n+1} \quad \forall n \in \mathbb{N} \end{cases}$$

Kurz: $(A_n)_n \uparrow$ *bzw.* $(A_n)_n \downarrow$
Für monoton wachsende bzw. fallende Ereignisfolgen heißt jeweils

$$\lim_{n\to\infty} A_n \uparrow = \bigcup_{n=1}^{\infty} A_n \text{ bzw. } \lim_{n\to\infty} A_n \downarrow = \bigcap_{n=1}^{\infty} A_n$$

der GRENZWERT *von* $(A_n)_{n\in\mathbb{N}}$.

Für eine beliebige Ereignisfolge $(A_n)_n$ *heißen*

$$\limsup_{n\to\infty} A_n = \lim_{n\to\infty} (\bigcup_{k=n}^{\infty} A_k) = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{k=n}^{\infty} A_k}_{\textit{fallend mit } n\to\infty}$$

LIMES SUPERIOR *und*

$$\liminf_{n\to\infty} A_n = \lim_{n\to\infty} (\bigcap_{k=n}^{\infty} A_k) = \bigcup_{n=1}^{\infty} \underbrace{\bigcap_{k=n}^{\infty} A_k}_{\textit{wachsend mit } n\to\infty}$$

LIMES INFERIOR *von* $(A_n)_{n\in\mathbb{N}}$

Bemerkung:

Es ist mit $(A_n)_n \subset \mathfrak{A}$:

$$\limsup_{n\to\infty} A_n \in \mathfrak{A}, \liminf_{n\to\infty} A_n \in \mathfrak{A}$$

nach Definition der σ-Algebra und

$$\begin{aligned}
\limsup_{n\to\infty} A_n &= \{\omega \in \Omega | \omega \text{ liegt in unendlich vielen der } A_i\} \\
&\mathrel{\widehat{=}} \text{ unendlich viele der } A_i \text{ treten ein} \\
\liminf_{n\to\infty} A_n &= \{\omega \in \Omega | \omega \text{ liegt in allen } A_i \text{ bis auf endlich viele}\} \\
&\mathrel{\widehat{=}} \text{ alle, bis auf endlich viele,(fast alle) der } A_i \text{ treten ein}
\end{aligned}$$

Lemma 3.3 *Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitraum, $(A_n)_n \subseteq \mathfrak{A}$, dann gilt*

1. $P(\bigcup\limits_{n=1}^{\infty} A_n) = P(\limsup\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(A_n)$ *,falls* $(A_n)_n \uparrow$
   (Stetigkeit von P von Unten)

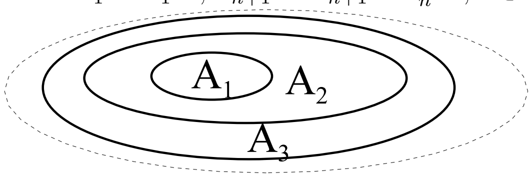
2. $P(\bigcap\limits_{n=1}^{\infty} A_n) = P(\liminf\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(A_n)$ *,falls* $(A_n)_n \downarrow$
   (Stetigkeit von P von Oben)

3. $P(\limsup\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(\bigcup\limits_{k=n}^{\infty} A_k)$
   $P(\liminf\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(\bigcap\limits_{k=n}^{\infty} A_k)$

4. $P(\bigcup\limits_{n=1}^{\infty} A_n) \leqslant \sum\limits_{n=1}^{\infty} P(A_n)$ *(Sub-σ-Additivität)*

Beweis:

1. Seien $B_1 = A_1 \quad , B_{n+1} = A_{n+1} \cap A_n^C \quad , n \in \mathbb{N}$ und $A_{n+1} \supset A_n$



$\Rightarrow (B_n)$ paarweise disjunkt und

$$\bigcup_{k=1}^{\infty} B_n = B_1 \cup \bigcup_{k=2}^{\infty} \underbrace{(A_n \cap A_{n-1}^C)}_{\subseteq A_n} \overset{!}{\subseteq} \bigcup_{n=1}^{\infty} A_n$$

und

$$\begin{aligned}
\omega \in \bigcup_{n=1}^{\infty} A_n \quad &\Rightarrow \quad \exists i : \omega \in A_i \wedge \omega \in A_j \forall j < i \\
&\Rightarrow \quad \omega \in B_i \\
&\Rightarrow \quad \omega \in \bigcup_{n=1}^{\infty} B_n \\
&\Longrightarrow \quad \bigcup_{n=1}^{\infty} B_n \overset{!}{\supseteq} \bigcup_{n=1}^{\infty} A_n
\end{aligned}$$

Also folgt: $\bigcup\limits_{n=1}^{\infty} B_n \stackrel{!}{=} \bigcup\limits_{n=1}^{\infty} A_n$

Damit:

$$\begin{aligned}
P(\bigcup_{n=1}^{\infty} A_n) &= P\Big(\bigcup_{n=1}^{\infty} B_n\Big) \\
&\stackrel{\sigma\text{-Add}}{=} \sum_{n=1}^{\infty} P(B_n) \\
&= P(A_1) + \sum_{n=2}^{\infty} P(B_n) \\
&= P(A_1) + \lim_{m\to\infty} \sum_{n=2}^{m} P(B_n) \\
&= P(A_1) + \lim_{m\to\infty} \sum_{n=1}^{m} \underbrace{P(B_{n+1})}_{P(A_{n+1})-P(A_n)} \qquad \text{(Teleskopsumme)} \\
&= P(A_1) + \lim_{m\to\infty} (P(A_{m+1}) - P(A_1)) \\
&= \lim_{m\to\infty} P(A_m)
\end{aligned}$$

2. Mit de-Morgan aus 1:

$$\bigcap_{n=1}^{\infty} A_n = (\bigcup_{n=1}^{\infty} A_n^C) \quad , (A_n^C) \uparrow$$

$$\begin{aligned}
\Rightarrow P(\bigcap_{n=1}^{\infty} A_n) &= 1 - P(\bigcup_{n=1}^{\infty} A_n^C) \\
&\stackrel{1}{=} 1 - \lim_{m\to\infty} \underbrace{P(A_m^C)}_{1-P(A_m)} \\
&= \lim_{m\to\infty} P(A_m)
\end{aligned}$$

3.

$$\begin{aligned}
P(\limsup_{n\to\infty} A_n) &= P(\lim_{n\to\infty} \underbrace{\bigcup_{k=n}^{\infty} A_k)}_{\text{fällt mit n}} \\
&\stackrel{2}{=} \lim_{m\to\infty} P(\bigcup_{k=n}^{\infty} A_k) \qquad \square
\end{aligned}$$

$$\begin{aligned}
P(\liminf_{n\to\infty} A_n) &= P(\lim_{n\to\infty} \underbrace{\bigcap_{k=n}^{\infty} A_k)}_{\text{wächst mit n}} \\
&\stackrel{2}{=} \lim_{m\to\infty} P(\bigcap_{k=n}^{\infty} A_k) \qquad \square
\end{aligned}$$

4. Analog zu Lemma 3.1(6) mit Verwendung der σ-Additivität (in 3.1 für endliche Ereignisse)

3.2 Siebformel von Sylvester-Poincaré

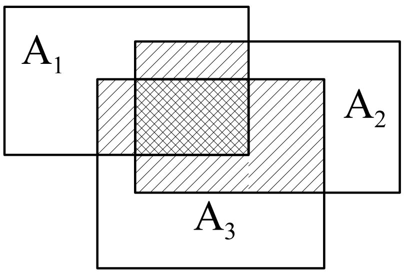
Lemma 3.4 *Für Ereignisse $(A_n)_{n\in\mathbb{N}}$ in einem Wahrscheinlichkeitraum $(\Omega, \mathfrak{A}, P)$ gilt:*

$$\begin{aligned} P(\bigcup_{k=1}^{n} A_k) &= \sum_{k=1}^{n} P(A_k) - \sum_{1\leqslant i_1<i_2\leqslant n} P\big(A_{i_1}\cap A_{i_2}\big) \\ &\quad + \sum_{1\leqslant i_1<i_2<i_3\leqslant n} P\big(A_{i_1}\cap A_{i_2}\cap A_{i_3}\big) \\ &\quad \pm \ldots + (-1)^{n+1} P\left(\bigcap_{k=1}^{n} A_k\right) \end{aligned}$$

Speziell:

$$n = 2 \qquad P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Struktur aus der Kombinatorik als Einschluss-Ausschlussprinzip *bekannt*



$n = 3$:

Beispiel 3.1 *Sei $\Omega_n := \{\pi|\ \pi : \{1,\ldots,n\} \mapsto \{1,\ldots,n\}$ bijektiv$\}$, d.h. Ω_n ist Menge aller Permutationen aus n Zahlen:*
$\Rightarrow |\Omega_n| = (n)_n = n!$
Seien $A := \{\pi \in \Omega_n|\ \pi(i) \neq i \quad ,1 \leqslant i \leqslant n\}$ die Menge der fixpunktfreien Permutationen und P sei Laplaceverteilung auf Ω_n
Weiterhin:

$$A_i := \{\pi \in \Omega_n|\ \pi(i) = i \quad ,1 \leqslant i \leqslant n\} \text{ Fixpunkt mindestens an Stelle } i$$

$\Rightarrow A = (\bigcup_{i=1}^{n} A_i)^C$

$$\begin{aligned} \Rightarrow P(A) &= 1 - P(\bigcup_{i=1}^{n} A_i) \\ &\overset{3.4}{=} 1 - \sum_{k=1}^{n} P(A_k) + \sum_{1\leqslant i_1<i_2\leqslant n} P(A_{i_1}\cap A_{i_2}) \\ &\quad - \sum_{1\leqslant i_1<i_2<i_3\leqslant n} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}) \\ &\quad \pm \ldots + (-1)^{n+1} P\left(\bigcap_{k=1}^{n} A_k\right) \end{aligned}$$

Nun ist $|A_i| = (n-1)!$ *(genau eine Stelle fest)*
$i_1 < i_2, \quad |A_{i_1} \cap A_{i_2}| = (n-2)!$ *(zwei Stellen fest)*
d.h. $|A_{i_1} \cap \ldots \cap A_{i_k}| = (n-k)! \quad , i_1 < \ldots < i_k$

$$\begin{aligned}
\Rightarrow P(A) &= 1 - \frac{1}{n!}\sum_{k=1}^{n}\left((-1)^{k+1} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} (n-k)!\right) \\
&= 1 - \sum_{k=1}^{n}(-1)^{k+1}\frac{(n-k)!}{n!}\tbinom{n}{k} \leftarrow \textit{Anzahl der Summanden} \\
&= 1 - \sum_{k=1}^{n}(-1)^{k+1}\frac{(n-k)!}{n!}\cdot\frac{n!}{k!(n-k)!} \\
&= \sum_{k=0}^{n}(-1)^{k}\frac{1}{k!}
\end{aligned}$$

Weiterhin sei:

$$\begin{aligned}
B_{n,m} &:= \{\pi \in \Omega_n |\ \pi \textit{ hat genau } m \textit{ Fixpunkte}\} \\
&= \{\pi \in \Omega_n |\ \pi(i_j) = i_j \quad , 1 \leqslant j \leqslant m \quad \wedge \quad \pi(i_k) \neq i_k \quad \forall i_k, k > m\} \\
\Rightarrow P(B_{n,m}) &= \underbrace{\frac{1}{n!}}_{|\Omega_n|}\tbinom{n}{m}(n-m)!\sum_{k=0}^{n-m}(-1)^k\frac{1}{k!}
\end{aligned}$$

d.h., dass es für m Fixpunkte genau $\binom{n}{m}$ *Möglichkeiten der Auswahl gibt. Unter den übrigen* $n-m$ *Stellen gibt es keine Fixpunkte.*

$$\Rightarrow \qquad P(B_{n,m}) = \frac{1}{m!}\sum_{k=0}^{n-m}(-1)^k\frac{1}{k!} \qquad \textit{und } A = B_{n,0}$$

Bemerkung: Interpretation über Sortierproblemen

Gegeben: Feld der Länge n

Gesucht: Wahrscheinlichkeit, dass mindestens $k \leqslant n$ Elemente des Feldes schon an der richtigen Stelle stehen, wenn die Elemente bezüglich eines ordinalen Merkmals sortiert werden sollen.

Voraussetzung: Laplaceraum, wo jede der $n!$ Anordnungen die Wahrscheinlichkeit $\frac{1}{n!}$ hat

Literatur: Knuth 73 Band 1, S.178

- Wahrscheinlichkeit, dass mindestens ein Element an der richtigen Stelle: $1 - P(A)$
- Wahrscheinlichkeit, dass in einem Feld der Länge n bereits k Elemente richtig sortiert sind: $B_{n,k}$
- $A_i \widehat{=}$ Menge aller Eingabefelder mit der richtigen Sortierung des i-ten Elements

Bemerkung: Es gilt:

$$\begin{aligned}\left|\sum_{k=0}^{n-m}(-1)^k\frac{1}{k!}-e^{-1}\right| &= \left|\sum_{k=0}^{n-m}(-1)^k\frac{1}{k!}-\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\right|\\ &= \left|\sum_{k=n-m+1}^{\infty}(-1)^k\frac{1}{k!}\right|\\ &\leqslant \frac{1}{(n-m+1)!} \qquad \text{alternierende Reihe}\end{aligned}$$

Unterscheidung:

- $n \geqslant 8$ und $m \geqslant 5 \;\rightarrow\; \frac{1}{m!} \leqslant \frac{1}{5!}$
- $n \geqslant 8$ und $m < 5 \;\rightarrow\; \frac{1}{(n-m+1)!} \leqslant \frac{1}{5!}$

$\Rightarrow$ Ist $n \geqslant 8$ und m beliebig, dann gilt:

$$\left|P(B_{n,m})-\frac{e^{-1}}{m!}\right| \leqslant \frac{1}{m!}\frac{1}{(n-m+1)!} \overset{\text{grob!}}{=} \frac{1}{5!} < 0,01 \qquad \star$$

Beispiel 3.2 *n Personen geben ihren Hut ab und Rückgabe erfolgt zufällig: Wie groß ist Wahrscheinlichkeit, dass niemand den eigenen Hut erhält ?*

$$\hookrightarrow P(B_{n,0}) = P(A) = \frac{1}{0!}\sum_{k=0}^{n}(-1)^k\frac{1}{k!} \overset{\star}{\approx} \frac{e^{-1}}{0!} \approx 0,37$$

mit einem Fehler von 0,01 unabhängig *von n !*

Korrolar 3.5 (Bonferroni-Ungleichungen) *Aus der Sylvester-Pointcaré-Siebformel: Seien $A_1,\dots,A_n$ Ereignisse aus $\mathfrak{A}$ im Wahrscheinlichkeitsraum $(\Omega,\mathfrak{A},P)$, dann gilt:*

$$\sum_{k=1}^{n}P(A_k)-\sum_{1\leqslant i_1<i_2\leqslant n}P\Big(A_{i_1}\cap A_{i_2}\Big) \quad\leqslant\quad P\left(\bigcup_{k=1}^{n}A_k\right) \quad\overset{3.1}{\leqslant}\quad \sum_{k=1}^{n}P(A_k)$$

Weitere Ungleichungen entstehen durch Abbruch der Siebformel nach gerader bzw. ungerader Ordnung

3.3 Zusammenfassung

Bezeichnung 3.6

Mathematisches Objekt	*Interpretation*
Ω	*Grundraum, Ergebnisraum*
ω	*(mögliches) Ergebnis*
$A \in \mathfrak{A}$	*Ereignis*
$\mathfrak{A}$	*Menge der (möglichen) Ereignisse*
Ω	*sicheres Ereignis*
$\emptyset$	*unmögliches Ereignis*
$\omega \in A$	*Ereignis A tritt ein*
$\omega \in A^C$	*Ereignis A tritt* nicht *ein*
$\omega \in A \cup B$	*Ereignis A* oder *B tritt ein*
$\omega \in A \cap B$	*Ereignis A* und *B tritt ein*
$A \subset B$	*Eintreten von Ereignis A* impliziert *das Eintreten von Ereignis B*
$A \cap B = \emptyset$	*Ereignisse A und B schließen einander aus*
$\omega \in \bigcup\limits_{i \in \mathbb{I}} A_i$	mindestens ein *Ereignis A_i , $i \in \mathbb{I}$ tritt ein*
$\omega \in \bigcap\limits_{i \in \mathbb{I}} A_i$	alle *Ereignisse A_i , $i \in \mathbb{I}$ treten ein*
$\omega \in \limsup i \in \mathbb{I} A_i$	unendlich viele *Ereignisse A_i , $i \in \mathbb{I}$ treten ein*
$\omega \in \liminf i \in \mathbb{I} A_i$	alle bis auf endlich viele *Ereignisse A_i , $i \in \mathbb{I}$ treten ein*
$P(A)$	*Wahrscheinlichkeit, dass das Ereignis $A \in \mathfrak{A}$ eintritt (Wahrscheinlichkeit für A)*
$P(A) = 1$	*A tritt sicher ein*
$P(A) = 0$	*A tritt sicher* nicht *ein*
$P(A) > P(B)$	*A ist wahrscheinlicher als B*

4 Bedingte Wahrscheinlichkeit

Idee:

Verarbeitung von Vor- bzw. Zusatzinformationen.
Für $A \in \mathfrak{A}$ ist $P(A)$ die Wahrscheinlichkeit des Eintretens
Sei jetzt bekannt/gefordert: $B \in \mathfrak{A}$ tritt ein
$\hookrightarrow$ Einfluss auf $P(A)$?

Beispiel 4.1

1. *Würfel*
   Wie groß ist die Wahrscheinlichkeit für die 2 unter der Bedingung, dass eine gerade Zahl auftritt ?
   $\hookrightarrow P(\{2\}|\ \{2,4,6\}) = \frac{1}{3}$
   $\hookrightarrow$ *der Wahrscheinlichkeitsraum wird eingeschränkt*

2. *Urnen*
   Ziehe 2 Kugeln ohne Zurücklegen aus einer Urne mit 2 weißen und 3 schwarzen Kugeln
   $P($ *2. Kugel ist schwarz* $|$ *1.Kugel ist weiß*$) = \frac{3}{4}$

   Laplace-Experiment: $|\Omega| = 5 \cdot 4$

   weiße Kugeln: *1,2*

   schwarze Kugeln: *3,4,5*

   Ereignis A: *2.Kugel ist schwarz*

   Ereignis B: *1.Kugel ist weiß*

$$
\begin{aligned}
A \cap B &= \{(1,3),(1,4),(1,5),(2,3),(2,4),(2,5)\} \\
|A \cap B| &= 6 \\
P(A \cap B) &= \frac{6}{20} \\
B &= \{(1,2),(1,3),(1,4),(1,5),(2,1),(2,3),(2,4),(2,5)\} \\
|B| &= 8 \\
P(B) &= \frac{8}{20} \\
\Rightarrow P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{\frac{6}{20}}{\frac{8}{20}} \\
&= \frac{3}{4}
\end{aligned}
$$

Beispiel 4.2 *relative Häufigkeiten*

Befragung von 100 Personen zu den Präferenzen zu Produkt A oder B:

	F	M	
A	10	20	30
B	50	20	70
	60	40	100

Tabelle der relativen Häufigkeiten:

	F	M	
A	0,1	0,2	0,3
B	0,5	0,2	0,7
	0,6	0,4	1

Wie groß ist die Häufigkeit für B in der Gruppe der Frauen ?

$$\frac{\textit{relative Häufigkeit für B und Frau}}{\textit{relative Häufigkeit für Frau}} = \frac{0,5}{0,6} = \frac{5}{6}$$

4.1 Bedingte Wahrscheinlichkeitsverteilung

Definition 4.1 *Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum.*
Für jedes $B \in \mathfrak{A}$ mit $P(B) > 0$ wird durch

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad , A \in \mathfrak{A}$$

eine Wahrscheinlichkeitsverteilung $P(\bullet|B)$ auf $\mathfrak{A}$ definiert,
die sogenannte BEDINGTE VERTEILUNG *unter (der Hypothese) B.*
$P(A|B)$ heißt ELEMENTAR BEDINGTE WAHRSCHEINLICHKEIT *von A und B*

Beweis:
$(\Omega, \mathfrak{A}, P(\bullet|B))$ ist ein Wahrscheinlichkeitsraum, wegen $P(A|B) = P(A \cap B|B)$ ist auch
$(B, \underbrace{\{A \cap B|\ A \in \mathfrak{A}\}}_{\text{Spur-}\sigma\text{-Algebra}}, P(\bullet|B))$ ein Wahrscheinlichkeitsraum,
der sogenannte INDUZIERTE oder EINGESCHRÄNKTE Wahrscheinlichkeitsraum.
Zur Wohldefiniertheit: $P(\bullet|B)$ ist Wahrscheinlichkeitsverteilung, denn

1. $P(A|B) \geqslant 0$ klar

2. $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$

3. A_i paarweise disjunkt:

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

Lemma 4.2 *Sei* $A, B, A_1, A_2, \ldots, A_n \in \mathfrak{A}$
$P(A) > 0, \quad P(B) > 0, \quad P\left(\bigcap\limits_{i=1}^{n-1} A_i\right) > 0$ *dann gilt:*

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

2. $P\left(\bigcap\limits_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P\left(A_n | \bigcap\limits_{i=1}^{n-1} A_i\right)$

Beweis

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

2.

$$\begin{aligned}
P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 \cap A_1) \cdot \ldots \cdot P\left(A_n | \bigcap_{i=1}^{n-1} A_i\right) &= \\
= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \ldots \cdot \frac{P\left(\bigcap_{i=1}^{n} A_i\right)}{P\left(\bigcap_{i=1}^{n-1} A_i\right)} \\
= P\left(\bigcap_{i=1}^{n-1} A_i\right)
\end{aligned}$$

Beispiel 4.3 (Skat) *Wie groß ist die Wahrscheinlichkeit, dass die drei Spieler je genau ein Ass haben ?*
$A_i \widehat{=}$ *Spieler i hält genau ein Ass*
$\Rightarrow P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$

Modell:
10 Karten an Spieler 1, danach
10 Karten an Spieler 2, danach

10 Karten an Spieler 3, danach
2 Karten auf den Skat (Kartenstoß)
Verteilung unerheblich, da Karten gut gemischt.

$$\begin{aligned} P(A_1) &= \frac{\binom{4}{1}\cdot\binom{28}{9}}{\binom{32}{10}} \qquad \rightarrow \textit{hypergeometrische Wahrscheinlichkeitsverteilung} \\ P(A_2|A_1) &= \frac{\binom{3}{1}\cdot\binom{19}{9}}{\binom{22}{10}} \\ P(A_3|A_1\cap A_2) &= \frac{\binom{2}{1}\cdot\binom{10}{9}}{\binom{12}{10}} \end{aligned}$$

4.2 Formel der totalen Wahrscheinlichkeit

Lemma 4.3
Gegeben sei $A \in \mathfrak{A}$, $\quad (B_n)_n \subset \mathfrak{A}$, $\quad B_n$ *paarweise disjunkt*
$A \subset \bigcup\limits_{n=1}^{\infty} B_n$ *(In der Literatur:* $\bigcup\limits_{n=1}^{\infty} B_n = \Omega$ *disjunkte Zerlegung von* Ω*), dann gilt:*

$$P(A) = \sum_{n=1}^{\infty} P(A|B_n)\cdot P(B_n) \quad \textit{gewichtete Summe}$$

Ist $P(B_k) = 0$*, so ist* $P(A|B_k)$ *nicht definiert !*
Konvention: Setze dann $P(B_k)\cdot P(A|B_k) = 0$

Beweis:
Wegen $A \subset \bigcup\limits_{n=1}^{\infty} B_n$ ist $A = \bigcup\limits_{n=1}^{\infty} \underbrace{(A\cap B_n)}_{\text{paarweise disjunkt}}$

$$\Rightarrow P(A) \overset{\sigma\text{-Add.}}{=} \sum_{n=1}^{\infty} P(A\cap B_n) = \sum_{n=1}^{\infty} P(A|B_n)\cdot P(B_n)$$

Beispiel 4.4

1. *Ein Serienartikel wird parallel auf drei Fertigungsstrecken produziert mit einem gemeinsamen Transportband. Mengenanteile der 3 Anlagen werden (etwa durch unregelmäßige Ausfälle) als Wahrscheinlichkeit angesetzt:*

$$P(A_1) = 0,3 \quad P(A_2) = 0,2 \quad P(A_3) = 0,5$$

mit $A_i \widehat{=}$ *Artikel wurde in Anlage i hergestellt*
Desweiteren sind die Wahrscheinlichkeiten für fehlerhafte Artikel der Anlagen gegeben:

1	*2*	*3*
0,05	*0,03*	*0,09*

Gesucht ist die Wahrscheinlichkeit, dass ein zufällig geprüftes Stück fehlerhaft ist.
Sei $B \widehat{=}$ *Stück ist fehlerhaft*

$$\begin{aligned} P(B|A_1) &= 0,05 \\ P(B|A_2) &= 0,03 \\ P(B|A_4) &= 0,09 \end{aligned}$$

$$\Rightarrow P(B) = \sum_{i=1}^{3} P(B|A_i) \cdot P(A_i) = 0,066$$

2. *Eine unter einer Million Münzen hat Zahl auf beiden Seiten, alle anderen sind „fair“ → Symbol und Zahl, jede Seite mit Wahrscheinlichkeit $\frac{1}{2}$*
   Eine zufällig ausgewählte Münze wird 20mal geworfen mit dem Ergebnis 20 mal Zahl.
   Frage: Wie groß ist die Wahrscheinlichkeit, dass die Münze fair war ?

$$\begin{aligned}
A &\;\widehat{=}\; \textit{faire Münze wird gezogen} \\
B &\;\widehat{=}\; A^C \\
\Rightarrow P(A) &= \frac{10^6 - 1}{10^6} = 1 - 10^{-6} \quad , P(B) = 10^{-6} \\
\\
Z_{20} &\;\widehat{=}\; \textit{es fällt 20 mal Zahl} \\
P(Z_{20}) &= \underbrace{P(Z_{20}|A)}_{=\left(\frac{1}{2}\right)^{20}} \cdot P(A) + \underbrace{P(Z_{20}|B)}_{=1} \cdot P(B) \\
&= 2^{-20}(1 - 10^{-6}) + 10^{-6} \\
\\
\Rightarrow \underbrace{P(A|Z_{20})}_{\textit{gesucht}} &= P(Z_{20}|A) \cdot \frac{P(A)}{P(Z_{20})} \\
&\approx 0,4881
\end{aligned}$$

4.3 Bayes'sche Formel

Satz 4.4 *Seien $A, (B_n)_n \subset \mathfrak{A}$, B_n paarweise disjunkt, $A \subset \bigcup_{n=1}^{\infty} B_n$ und $P(A) > 0$, dann gilt:*

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n)} \quad , \forall k \in \mathbb{N}$$

Beweis:

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

Nach dem Satz der totalen Wahrscheinlichkeit gilt die Behauptung.
Bei dieser Formel wird von der „Wirkung“ A auf die „Ursache“ B geschlossen.

Beispiel 4.5 *Ein Arzt stellt Symptom A fest, dass von verschiedenen Krankheiten $B_1, \ldots, B_n$ herrühren kann.*

- *Die relativen Häufigkeiten einer jeden Krankheit sind bekannt → $P(B_i)$*
- *Wenn Krankheit B_k vorliegt, dann kenne ich die relative Häufigkeit für Symptom A → $P(A|B_k)$*

Annahme: B_i paarweise disjunkt, d.h. nur eine Krankheit tritt auf einmal auf.
Gesucht ist die Wahrscheinlichkeit für die Krankheit B_k, wenn Symptom A auftritt.

Bemerkung:

1. $P(B_k)$ heißen auch A PRIORI WAHRSCHEINLICHKEIT
2. $P(A|B_k)$ heißen auch A POSTERIORI WAHRSCHEINLICHKEIT

Fortsetzung zu Beispiel 4.4 Wie groß ist die Wahrscheinlichkeit, dass ein geprüftes und einwandfreies Stück in Anlage 3 hergestellt wurde ?

$$P(A_3|B^C) = \frac{P(A_3) \cdot P(B^C|A_3)}{\sum_{n=1}^{3} P(A_n)P(B^C|A_n)} = 0,49$$

mit

$$P(B^C|A_1) = \frac{P(B^C \cap A_1)}{P(A_1)} = \frac{P(A_1) - P(B \cap A_1)}{P(A_1)} = 1 - P(B|A_1)$$

Fortsetzung Beispiel 4.5

- in $0,9$ der Fälle wird ein Kranker als „krank“ erkannt, d.h. das Verfahren liefert richtigerweise einen positiven Befund
- in $0,05$ der Fälle wird ein Gesunder als „krank“ bezeichnet, d.h. das Verfahren liefert ein falsch-positives Ergebnis.

Modell: $0,01$ der Bevölkerung leiden an dieser Krankheit
Wie groß ist die Wahrscheinlichkeit, dass zufällige Testpersonen gesund sind, obwohl die Diagnose positiv ist ?

$$\begin{aligned} G &\mathrel{\widehat{=}} \text{„Person ist gesund“} \\ B &\mathrel{\widehat{=}} \text{Verfahren liefert positiven Befund} \end{aligned}$$

$$\begin{aligned} P(B|G^C) &= 0,9 \\ P(B|G) &= 0,05 \\ P(G^C) &= 0,01 \end{aligned}$$

$$\begin{aligned} \Rightarrow P(B) &= P(B|G^C) \cdot P(G^C) + P(B|G) \cdot P(G) \\ &= 0,9 \cdot 0,01 + 0,05 \cdot 0,99 \\ &\approx 0,0585 \end{aligned}$$

Gesucht:

$$P(G|B) = \frac{P(B|G) \cdot P(G)}{P(B)} = 0,846$$

Alternative Werte:

$$\begin{aligned} P(B|G^C) &= 0,9 \\ P(B|G) &= 0,01 \quad (0,001) \\ \Rightarrow P(G|B) &= 0,52 \quad (0,10) \end{aligned}$$

Beispiel 4.6 (gestörter Nachrichtenkanal)
Gesendet werden 0 und 1 über n Stationen

$$S_0 \to S_1 \to \ldots \to S_n$$

Die Wahrscheinlichkeit für eine korrekte Übertragung von S_i nach $S_{i+1}, i = 1, 2, \ldots, n$ sei $p(0,1)$.

Gesucht ist die Wahrscheinlichkeit p_n, dass S_n die von S_0 gesendete Nachricht erhält.
Modell:

$$\begin{aligned}
\Omega_n &= \{(\omega_0, \ldots, \omega_{n-1})|\ \omega_i \in \{0,1\}, \quad 0 \leqslant i \leqslant n-1\} \\
\omega_i &\mathrel{\widehat{=}} \text{ Nachricht von } S_i \text{ nach } S_{i+1} \\
\\
E_m &:= \{\omega \in \Omega_n|\ \omega_0 = \omega_{m-1}\} \quad ,1 \leqslant m \leqslant n \\
&\mathrel{\widehat{=}} S_m \text{ erhält korrekte Nachricht} \\
\\
\text{mit } p_m &:= P(E_m) \quad ,1 \leqslant m \leqslant n \text{ ist} \\
p_1 &= 1 \\
p_n &= P(E_n) \\
&= P(E_{n-1} \cap E_n) + P(E_{n-1}^C \cap E_n) \\
&= \underbrace{P(E_n|E_{n-1})}_{=:p} \cdot \underbrace{P(E_{n+1})}_{=p_{n-1}} + \underbrace{P(E_n|E_{n-1}^C)}_{=:1-p} \cdot \underbrace{P(E_{n+1}^C)}_{=1-p_{n-1}} \\
&= (2p-1)p_{n-1} + 1 - p \quad ,n \geqslant 2
\end{aligned}$$

Aus Induktion folgt:

$$p_n = \frac{1}{2} + \frac{1}{2}(2p-1)^{n-1}$$

$$\Rightarrow \lim_{n\to\infty} p_n = \frac{1}{2} \quad (|2p-1| < 1)$$

Beispielwerte:

n	*10*	*100*
p=0,8	*0,57*	*0,5*
0,99	*0,92*	*0,57*
0,999	*0,95*	*0,91*

5 Stochastische Unabhängigkeit

Heuristisch: Zwei Ereignisse A, B sind unabhängig, falls $P(A)$ nicht von der Kenntnis des Eintretens oder Nicht-Eintretens von B abhängt, d.h.

$$\begin{aligned} P(A|B) &= P(A) \quad \text{bzw.} \\ P(B|A) &= P(B) \end{aligned}$$

Beispiel 5.1 *Ziehe aus einer Urne mit 2 weißen und 3 schwarzen Kugeln 2 Kugeln mit Zurücklegen.*

$$\begin{aligned} \Omega &= \{(i,j)|\ 1 \leqslant i,j \leqslant 5\} \\ |\Omega| &= 5^2 = 25 \end{aligned}$$

weiße Kugeln: *1,2*

schwarze Kugeln: *3,4,5*

$$\begin{aligned} A &\mathrel{\widehat{=}} \textit{zweite Kugel ist schwarz} \\ A &= \{(i,j)|\ j \in \{3,4,5\}\} \\ |A| &= 5 \cdot 3 = 15 \end{aligned}$$

$$\begin{aligned} B &\mathrel{\widehat{=}} \textit{erste Kugel ist weiß} \\ B &= \{(i,j)|\ i \in \{1,2\}\} \\ |B| &= 5 \cdot 2 = 10 \end{aligned}$$

$$\begin{aligned} \Rightarrow P(A) &= \frac{15}{25} = \frac{3}{5} \\ P(B) &= \frac{10}{25} = \frac{2}{5} \\ P(A \cap B) &= \frac{6}{25} \end{aligned}$$

$$\begin{aligned} \Rightarrow P(A|B) &= \frac{\frac{6}{25}}{\frac{2}{5}} = \frac{3}{5} \\ &= P(A) \end{aligned}$$

Klar, denn durch das Zurücklegen beeinflussen sich die Ziehungen nicht

Weiterhin:

$$\begin{aligned} P(A) &= P(A|B) = \frac{P(A \cap B)}{P(B)} & (1) \\ \Rightarrow P(A \cap B) &= P(A) \cdot P(B) & (2) \end{aligned}$$

Gleichung 1 erfordert noch, dass $P(B) > 0$, daher:

Definition 5.1 *Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum, $A, B \in \mathfrak{A}$ heißen* STOCHASTISCH UNABHÄNGIG, *falls*

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

Lemma 5.2

1. *Mit* A, B *sind auch* A, B^C *und* A^C, B^C *stochastisch unabhängig*
2. *Ist* $P(B) > 0$, *so gilt:*
$$A, B \text{ stochastisch unabhängig} \Leftrightarrow P(A|B) = P(A)$$
3. *Ist* A *sog. Nullmenge, d.h.* $P(A) = 0$ *so ist* A, B *stochastisch unabhängig* $\forall B \in \mathfrak{A}$

Beweis:

1. A, B stochastisch unabhängig
$$\begin{aligned} P(A \cap B^C) &= P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot (1 - P(B)) \\ &= P(A) \cdot P(B^C) \end{aligned}$$
$$\begin{aligned} P(A^C \cap B^C) &\overset{\text{de Morgan}}{=} P\left((A \cup B)^C\right) = 1 - P(A \cup B) \\ &= 1 - (P(A) + P(B) - \underbrace{P(A \cap B)}_{P(A) \cdot P(B)}) \\ &= (1 - P(A)) \cdot (1 - P(B)) \\ &= P(A^C) \cdot P(B^C) \end{aligned}$$
2. s.o.
3. $P(A) = 0$
$$\begin{aligned} \Rightarrow P(A \cap B) &\leqslant P(A) = 0 \\ \Rightarrow P(A \cap B) &= 0 \end{aligned}$$

Bemerkung: Unabhängigkeit ist abhängig von der Wahrscheinlichkeitsverteilung **!**

Sei $\Omega = \{1, 2, 3\}$, $A = \{1\}$, $B = \{1, 2\}$
$P = \varepsilon_1$, $\quad Q$: Laplaceverteilung

$$\Rightarrow \varepsilon_1(A \cap B) = \varepsilon_1(1) = 1 = \varepsilon_1(A) \cdot \varepsilon_1(B)$$

aber
$$Q(A \cap B) = \frac{1}{3} \neq \frac{2}{9} = \frac{1}{3} \cdot \frac{2}{3} = Q(A) \cdot Q(B)$$

Benötigt wird eine Definition für 2 oder mehr Ereignisse:

Definition 5.3 *Eine Familie von Ereignissen* $(A_i)_{i \in \mathbb{I}} \subset \mathfrak{A}$, $\mathbb{I}$ *beliebige Indexmenge, heißt* PAARWEISE STOCHASTISCH UNABHÄNGIG , *falls*

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \quad , \forall i, j \in \mathbb{I}, i \neq j$$

Beispiel 5.2 *Werfen von 2 unverfälschten Würfeln, d.h. Laplaceverteilung über* $\Omega = \{ (i,j) | \, i,j \in \{1,\ldots,6\} \}$

$$\begin{aligned} A_i &\;\widehat{=}\; \text{Würfel } i \text{ zeigt gerade Zahl}, \quad i = 1,2 \\ \Rightarrow P(A_1 \cap A_2) &= P(A_1) \cdot P(A_2) = \frac{1}{4} \quad \text{, d.h. stochastisch unabhängig} \end{aligned}$$

Weiterhin:

$$\begin{aligned} A_3 &\;\widehat{=}\; \text{Summe der Augen gerade} \\ \Rightarrow P(A_3|A_1) &= P(A_3) \\ \text{und } P(A_3|A_2) &= P(A_3) \end{aligned}$$

d.h. $\{A_1, A_2, A_3\}$ *paarweise stochastisch unabhängig*

$$\begin{aligned} \text{Aber} \quad P(A_1 \cap A_2 \cap A_3) &\neq P(A_1) \cdot P(A_2) \cdot P(A_3) \\ \text{denn} \quad P(A_1 \cap A_2 \cap A_3) &= P(A_1 \cap A_2) = \frac{1}{4} \\ P(A_1) \cdot P(A_2) \cdot P(A_3) &= \frac{1}{8} \end{aligned}$$

5.1 Vollständige stochastische Unabhängigkeit

Benötigt wird also ein stärkerer Begriff der stochastischen Unabhängigkeit:

Definition 5.4 *Eine Familie von Ereignissen* $(A_i)_{i \in \mathbb{I}}$ *heißt* (VOLLSTÄNDIG) STOCHASTISCH UNABHÄNGIG, *falls für jede endliche Auswahl gilt:*

$$P\left(\bigcap_{j \in \mathbb{J}} A_j\right) = \prod_{j \in \mathbb{J}} P(A_j) \qquad , \forall\ \mathbb{I} \neq \emptyset, \mathbb{J} \subset \mathbb{I}, |\mathbb{J}| < \infty$$

Bemerkung:

1. $(A_i)_i$ stochastisch unabhängig $\Rightarrow$ $(A_i)_i$ paarweise stochastisch unabhängig
   Umkehrung gilt nicht ! (siehe Beispiel)

2. Jede Teilfamilie einer stochastisch unabhängigen Familie ist selbst wieder stochastisch unabhängig

3. Beachte, dass die Definition ein System von Gleichungen liefert
   Beispiel: A, B, C stochastisch unabhängig falls

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(A \cap C) &= P(A) \cdot P(C) \\ P(C \cap B) &= P(C) \cdot P(B) \\ &\text{und} \\ P(A \cap B \cap C) &= P(A) \cdot P(B) \cdot P(C) \end{aligned}$$

Satz 5.5

1. *Seien $(A_i)_{i\in\mathbb{I}}$ Familie von stochastisch unabhängigen Ereignissen, $k \notin \mathbb{I}$ und $P(A_k)$ aus der Menge $\{0,1\}$*

$$\Rightarrow (A_i)_{i\in\mathbb{I}\cup\{k\}} \quad \textit{stochastisch unabhängig}$$

2. *$(A_i)_{i\in\mathbb{I}}$ stochastisch unabhängig und $B_i \in \{A_i, A_i^C, \emptyset, \Omega\} \quad \forall i \in \mathbb{I}$*

$$\Rightarrow (B_n)_{n\in\mathbb{I}} \quad \textit{stochastisch unabhängig}$$

3. *Sei $\mathbb{I} = \{1,\ldots,n\}$, $n \in \mathbb{N}$*

$$(A_i)_{i\in\mathbb{I}} \textit{ stochastisch unabhängig} \Leftrightarrow P\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} P(B_i) \quad \forall B_i \in \{A_i, A_i^C\},\ \forall i \in \mathbb{N}$$

Beispiel 5.3
Ein Experiment liefert mit Wahrscheinlichkeit p ein Ereignis A und Ereignis A^C mit Wahrscheinlichkeit $q := 1-p$
Das Experiment wird n-mal „unabhängig" wiederholt
$\hookrightarrow$ *Modell:* $\Omega = \{\ (\omega_1,\ldots,\omega_n) \mid \omega_i \in \{\underbrace{0}_{A^C}, \underbrace{1}_{A}\}, \quad 1 \leqslant i \leqslant n\}$
Interpretiert man experimentelle Unabhängigkeit als stochastische Unabhängigkeit, dann folgt, dass jedes ω mit k Komponenten gleich 1 folgende Wahrscheinlichkeit hat:

$$\begin{aligned} p(\omega) &= p^k(1-p)^{n-k} \\ \sum_{\omega\in\Omega} p(\omega) &= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} \\ &= 1 \end{aligned}$$

Dies bedeutet, dass (Ω, P) einen endlichen Wahrscheinlichkeitsraum beschreibt:
BERNOULLI-MODELL

Bemerkung: Sei die Familie $(A_i)_{i\in\{1,\ldots,n\}}$ stochastisch unabhängig oder $A_1,\ldots,A_n$ stochastisch unabhängig, dann gilt:

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= 1 - P\left(\bigcap_{i=1}^{n} A_i^C\right) \\ &= 1 - \prod_{i=1}^{n} P(A_i^C) \\ &= 1 - \prod_{i=1}^{n}(1 - P(A_i)) \end{aligned}$$

Beispiel 5.4 (Ziegenproblem)
Situation:

- *3 Türen, dahinter 1 Auto und 2 Ziegen*
- *Kandidat wählt eine Türe*
- *Eingriff des Moderators:*
  Dieser öffnet eine der nicht-gewählten Türen, hinter der kein Auto steht
- *Kandidat darf Wahl ändern*

Frage: *Ist die Änderung der Entscheidung sinnvoll ?*
Konkret: *o.B.d.A wegen Symmetrie wählt Kandidat Tür 1, Quizmaster öffnet Tür 3. Soll Kandidat bei Tür 1 bleiben oder zu 2 wechseln ?*
Sei, ohne spezifizierten Wahrscheinlichkeitsraum:

$$\begin{aligned}
A_i &\mathrel{\widehat{=}} \textit{Auto hinter Türe } i \\
P(A_i) &= \frac{1}{3} \quad , i \in \{1,2,3\} \\
K_i &\mathrel{\widehat{=}} \textit{Kandidat wählt Türe } i \\
P(K_i) &= \frac{1}{3} \quad , i \in \{1,2,3\} \\
A_i, K_j & \quad \textit{stochastisch unabhängig} \quad , 1 \leqslant i,j \leqslant 3 \\
\\
Q_i &\mathrel{\widehat{=}} \textit{Quizmaster öffnet Türe } i \textit{ (nicht unabhängig von } A_i, K_j\textit{)} \\
\Rightarrow P\underbrace{(A_1|\ K_1 \cap Q_3)}_{\textit{Bleibestrategie}} &= \frac{P(A_1 \cap K_1 \cap Q_3)}{P(K_1 \cap Q_3)} \\
&= \frac{P(Q_3|A_1 \cap K_1)}{P(K_1 \cap Q_3)} \cdot P(A_1 \cap K_1)
\end{aligned}$$

Es ist:

$$\begin{aligned}
P(K_1 \cap Q_3) &= P(K_1 \cap Q_3 \cap A_1) + P(K_1 \cap Q_3 \cap A_2) + \underbrace{P(K_1 \cap Q_3 \cap A_3)}_{=0} \\
&= \underbrace{P(Q_3|A_1 \cap K_1)}_{=\frac{1}{2}} \cdot \underbrace{P(A_1 \cap K_1)}_{=\frac{1}{9}} \\
&\quad + \underbrace{P(Q_3|A_2 \cap K_1)}_{=1} \cdot \underbrace{P(A_2 \cap K_1)}_{=\frac{1}{9}} \\
&= \frac{1}{2} \cdot \frac{1}{9} + 1 \cdot \frac{1}{9} \\
&= \frac{1}{6}
\end{aligned}$$

Eingesetzt in die Bleibestrategie:

$$\frac{P(Q_3|A_1 \cap K_1)}{P(K_1 \cap Q_3)} \cdot P(A_1 \cap K_1) = \frac{\frac{1}{2}}{\frac{1}{6}} \cdot \frac{1}{9} = \frac{1}{3}$$

Aber

$$P(A_2|K_1 \cap Q_3) \overset{analog}{=} \frac{2}{3}$$

d.h. die Änderung der Entscheidung verdoppelt die Gewinnwahrscheinlichkeit !

Bemerkung: Die Relation „stochastisch unabhängig“ ist **nicht transitiv**, d.h.
Aus A_1, A_2 stochastisch unabhängig und
A_2, A_3 stochastisch unabhängig,
folgt nicht, dass A_1, A_3 notwendigerweise stochastisch unabhängig sind.

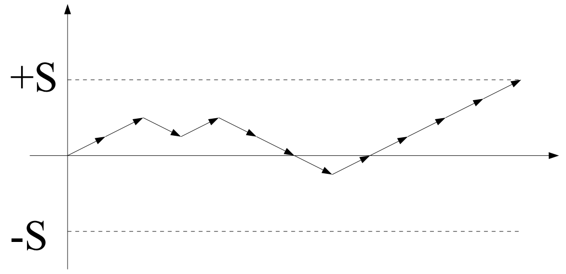
Beispiel 5.5 *Sei* $(\Omega, \mathfrak{A}, P)$ *ein Wahrscheinlichkeitsraum mit* $\Omega = \{(0,0),(1,0),(0,1),(1,1)\}$ *und* P *eine Laplace-Wahrscheinlichkeitsverteilung. Desweiteren:*

$$\begin{aligned} A_1 &= \{(0,0),(0,1)\} \\ A_2 &= \{(0,1),(1,0)\} \\ A_3 &= \{(1,0),(1,1)\} \end{aligned}$$

$$\begin{aligned} P(A_i) &= \frac{1}{2} \quad , i \in \{1,2,3\} \\ &\textit{und} \\ P(A_1 \cap A_2) = P(\{(0,1)\}) &= \frac{1}{4} = P(A_1) \cdot P(A_2) \\ P(A_2 \cap A_3) = P(\{(1,0)\}) &= \frac{1}{4} = P(A_2) \cdot P(A_3) \\ &\textit{aber} \\ P(A_1 \cap A_3) = P(\emptyset) &= 0 \neq P(A_1) \cdot P(A_3) \end{aligned}$$

Folgen von unabhängigen Ereignissen spielen in der Stochastik eine zentrale Rolle:

- Simulation von zufallsbeeinflussten Prozessen
- mehrfacher Münzwurf: Wann fällt das erste Mal „Zahl“ ?
- diskrete Versuche, bei denen interessant ist, wann die Summe der Ergebnisse zum ersten mal größer als ein bestimmter Wert wird
- Irrfahrten (Labyrinth)
  Wenn in einem Irrgarten die Richtungsänderungen zufällig erfolgt ist interessant, nach wievielen Zügen dieser verlassen wird.
- Summenpfad
  Wann überschreitet ein Summenpfad zum ersten Mal eine Grenze S ?



Definition 5.6 *Sei* $(A_i)_{i\in\mathbb{N}} \subset \mathfrak{A}$
$(A_n)_n$ *heißt* KONVERGENT $\Leftrightarrow \limsup\limits_{n\to\infty} A_n = \liminf\limits_{n\to\infty} A_n$

Bemerkung: Es gilt stets $\liminf\limits_{n\to\infty} A_n \subset \limsup\limits_{n\to\infty} A_n$, denn:

$$\text{Sei } \omega \in \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k \Rightarrow \exists n_0, \quad \omega \in \bigcap_{k=n_0}^{\infty} A_k$$

$$\begin{aligned} \Rightarrow \omega &\in A_k \quad \forall k \geqslant n_0 \\ \Rightarrow \omega &\in \bigcup_{k=n}^{\infty} A_k \quad \forall n \geqslant 1 \\ \Rightarrow \omega &\in \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k \end{aligned}$$

Desweiteren gilt:

$$\begin{aligned} \left(\limsup_{n\to\infty} A_n\right)^C &= \left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right)^C \\ &= \bigcup_{k=n}^{\infty}\bigcap_{n=1}^{\infty} A_k^C \\ &= \liminf_{n\to\infty} A_n^C \end{aligned}$$

Lemma 5.7 (von Borel Cantelli) *:*
Sei $(\Omega, \mathfrak{A}, P)$ *ein Wahrscheinlichkeitsraum und* $(A_i)_{i\in\mathbb{N}} \subset \mathfrak{A}$*, dann gilt:*

1. $\sum\limits_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P\left(\limsup\limits_{n\to\infty} A_n\right) = 0$

2. *Ist zusätzlich* $(A_n)_n$ *stochastisch unabhängig,*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P\left(\limsup_{n\to\infty} A_n\right) = 1$$

BEWEIS:

1. Wegen $\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty} A_k \subset \bigcup\limits_{k=n}^{\infty} A_k \quad \forall n \in \mathbb{N}$ gilt:

$$P\left(\limsup_{n\to\infty} A_n\right) \leqslant P\left(\bigcup_{k=n}^{\infty} A_k\right) \overset{\text{sub-}\sigma\text{-Add.}}{\leqslant} \sum_{k=n}^{\infty} P(A_k) \overset{n\to\infty}{\longrightarrow} 0$$

2. Es ist

$$P\left(\limsup_{n\to\infty} A_n\right) \overset{\text{s.o.}}{=} 1 - P\left(\liminf_{n\to\infty} A_n^C\right)$$

$$\begin{aligned}
&\overset{3.3}{=} \quad 1 - \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k^C\right) \\
&\overset{\text{unabh.}}{=} \quad 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} (1 - P(A_k))
\end{aligned}$$

Mit der Abschätzung $1 - x \leqslant e^{-x}, \quad x \in \mathbb{R}$

$$\begin{aligned}
&\geqslant \quad 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} \exp(-P(A_k)) \\
&\geqslant \quad 1 - \lim_{n\to\infty} \exp \underbrace{\left(-\sum_{k=n}^{\infty} P(A_k)\right)}_{\to -\infty \quad \text{nach Vor.}} = 1
\end{aligned}$$

Bemerkung:

1. Analog gilt noch mit de Morgan
   - $\sum\limits_{n=1}^{\infty} P(A_n^C) < \infty \Rightarrow P\left(\liminf\limits_{n\to\infty} A_n\right) = 1$
   - $(A_n)_n$ stochastisch unabhängig :

   $$\sum_{n=1}^{\infty} P(A_n^C) = \infty \Rightarrow P\left(\liminf_{n\to\infty} A_n\right) = 0$$

2. Für stochastisch unabhängige Ereignisse $(A_i)_i$ gilt stets

   $$P\left(\limsup_{n\to\infty} A_n\right), P\left(\liminf_{n\to\infty} A_n\right) \in \{0, 1\}$$

3. Sei $(A_n)_n \subset \mathfrak{A}$, zu der eine stochastisch unabhängige Teilfolge $(A_{n_k})_k$ von $(A_n)_n$ existiert mit $\sum\limits_{k=1}^{\infty} P(A_{n_k}) = \infty$, dann folgt $P\left(\limsup\limits_{n\to\infty} A_n\right) = 1$

Beispiel 5.6

1. *Betrachte eine unendliche Menge von Urnen aus denen je eine Kugel gezogen wird. Die Urne n enthält eine weiße und $n-1$ schwarze Kugeln. Sei nun*

   $$\begin{aligned}
   A_n \quad &\widehat{=} \quad \textit{die gezogene Kugel aus der } n\textit{-ten Urne ist weiß} \\
   &\Rightarrow \quad P(A_n) = \frac{1}{n} \quad \textit{,Ziehungen unabhängig} \\
   &\Rightarrow \quad \sum_{n=1}^{\infty} P(A_n) = \infty \\
   &\Rightarrow \quad P\left(\limsup_{n\to\infty} A_n\right) = 1
   \end{aligned}$$

   dies bedeutet, dass mit Wahrscheinlichkeit 1 unendlich viele weiße Kugeln gezogen werden, obwohl die Wahrscheinlichkeit immer weiter abnimmt.

2. *Wie oben, aber die n-te Urne enthält $n^2 - 1$ schwarze Kugeln*

$$\begin{aligned} &\Rightarrow \quad \sum_{n=1}^{\infty} P(A_n) < \infty \\ &\Rightarrow \quad P\left(\limsup_{n\to\infty} A_n\right) = 0 \end{aligned}$$

d.h. mit Wahrscheinlichkeit 0 werden nur endlich viele weiße Kugeln gezogen.

Beachte:

Die Anzahl der Kugeln kann **nicht so** *gewählt werden, dass*

$$P\left(\limsup_{n\to\infty} A_n\right) \in (0,1)$$

Beispiel 5.7

1. *Unabhängiges Werfen eines Würfels.*
   Gesucht ist die Wahrscheinlichkeit unendlich oft die 6 zu würfeln.

   $$\begin{aligned} A_n &\mathrel{\widehat{=}} 6 \text{ im } n\text{-ten Wurf} \\ P(A_n) &= \frac{1}{6} \quad \forall n \in \mathbb{N} \end{aligned}$$

   $(A_n)_n$ *stochastisch unabhängig, d.h.* $\sum P(A_n) = \infty$

   $$\Rightarrow P\left(\limsup_{n\to\infty} A_n\right) = 1$$

   Beachte: *Das Grundmodell ist überabzählbar !*
   z.B. Münzwurf: $\Omega = \{(\omega_n)_{n\in\mathbb{N}} | \ \omega_n \in \{0,1\}, n \in \mathbb{N}\}$
   kann betrachtet werden als die Dualdarstellung der reellen Zahlen im Intervall $(0,1)$.

2. *Gesucht ist nun die Wahrscheinlichkeit unendlich oft zwei Sechsen hintereinander zu würfeln.*

   $$\begin{aligned} B_{n,n+1} &\mathrel{\widehat{=}} 6 \text{ im } n\text{-ten Wurf und im } n+1\text{-ten} \\ \text{Nicht unabhängig mit } n & \\ P(B_{n,n+1}) &= \frac{1}{36} \end{aligned}$$

   Betrachte also die unabhängig Familie $(B_{2n-1,2n})_{n\in\mathbb{N}}$.
   Beachte, dass diese Familie nicht alle „Treffer“ enthält. Etwa bei Sechsen in den Würfen 3,4 und 5 würde dies nur als einfacher Treffer gezählt. Dennoch ist

   $$\begin{aligned} P\left(\limsup_{n\to\infty} B_{2n-1,2n}\right) &= 1 \\ \Rightarrow P\left(\limsup_{n\to\infty} B_{n,n+1}\right) &= 1 \end{aligned}$$

5.2 Produktexperimente

Idee:

Modelle $(\Omega_i, \mathfrak{A}_i, P_i), 1 \leqslant i \leqslant n$ bekannt, z.B. Ziehen mit Zurücklegen aus Urnen, Würfelexperimente, ...

Ziel:

Ein Modell für ein Experiment, dass aus der unabhängigen Hintereinanderausführung der Teilexperimente besteht.
z.B. n-maliges Ziehen, dann n-maliger Würfelwurf, ...

$$\begin{aligned} \Rightarrow \Omega &= \{(\omega_1, \dots, \omega_n) |\ \omega_i \in \Omega_i, 1 \leqslant i \leqslant n\} \\ &= \prod_{i=1}^{n} \Omega_i \qquad (\texttt{Schreibweise}) \\ &= \mathop{\times}_{i=1}^{n} \Omega_i \qquad (n\text{-faches Kreuzprodukt}) \end{aligned}$$

Ω_i müssen nicht identisch sein.

Definition 5.8 *Für diskrete Wahrscheinlichkeitsräume $(\Omega_i, \mathfrak{A}_i, P_i), 1 \leqslant i \leqslant n$ heißt $(\Omega, \mathfrak{A}, P)$ mit*

$$\Omega = \prod_{i=1}^{n} \Omega_i = \{(\omega_1, \dots, \omega_n) |\ \omega_i \in \Omega_i, 1 \leqslant i \leqslant n\}$$

P definiert durch

$$P = \prod_{i=1}^{n} P_i(\omega_i) =: \mathop{\times}_{i=1}^{n} P_i$$

und $\mathfrak{A}$ ist Potenzmenge von Ω
Produkt der Wahrscheinlichkeitsräume $(\Omega_i, \mathfrak{A}_i, P_i), 1 \leqslant i \leqslant n$
`Schreibweise:`

$$(\Omega, \mathfrak{A}, P) = \bigotimes_{i=1}^{n} (\Omega_i, \mathfrak{A}_i, P_i)$$

Beispiel 5.8 *Binomialverteilung bei n-facher Wiederholung eines Experiments mit Ausgang jeweils 0 oder 1*

$$\begin{aligned} \Omega &= \{0,1\}^n \\ P(\omega) &= \underbrace{p^k}_{\textit{Treffer}} \underbrace{(1-p)^{n-k}}_{\textit{Nieten}} \end{aligned}$$

So ist k die Anzahl der Einsen in einem Tupel der Länge n und $p = P_i(1), 1 \leqslant i \leqslant n$
Bezeichnungen:

$$\begin{aligned} p &: \textit{Erfolgswahrscheinlichkeit} \\ \omega_i = 1 &: \textit{Erfolg im i-ten Teilexperiment} \end{aligned}$$

in n sog. Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p.

Die durch

$$\begin{aligned} E_k &:= \{\omega \in \Omega_i | \sum_{i=1}^{n} \omega_i = k\} \\ &\widehat{=} \quad \textit{Genau } k \textit{ Erfolge mit } k \leqslant n \\ P(E_k) &\overset{s.o.}{=} \binom{n}{k} p^k (1-p)^{n-k} \qquad E_i\textit{'s sind disjunkt} \end{aligned}$$

definierte Wahrscheinlichkeitsverteilung auf $\{0, 1, \ldots, n\}$ *heißt* BINOMIALVERTEILUNG

6 Zufallsvariablen

Zufallsvorgänge werden beschrieben durch einen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$.
Dabei ist häufig $\omega \in \Omega$ nicht von Interesse, sondern eine Funktion X von ω.
Zum Beispiel:

- $\omega :=$ n-facher Münzwurf
  $\rightarrow X(\omega)$: Anzahl von „Zahl“
- $\omega :=$ Telefongespräch
  $\rightarrow X(\omega)$: Gesamtdauer
- $\omega :=$ Aktienmarkt
  $\rightarrow X(\omega)$: Kurs einer Aktie

Ist dieser Grundraum vollständig beschreibbar ?
Sind alle Einflussgrößen beschreibbar ?

Gesucht ist also eine neue, zielorientierte Form der Modellierung durch ZUFALLSVARIABLEN.

Beispiel 6.1

1. *Fortsetzung des n-fachen unabhängigen Münzwurfes*
   $\Omega = \{0, 1\}^n, P(\omega) = p^k(1-p)^{n-k}$, *falls k Einsen in ω*
   Betrachte hierbei die Abbildung:

$$X : \begin{cases} \Omega & \rightarrow \quad \mathbb{N}_0 \\ \omega & \mapsto \quad \sum\limits_{i=1}^{n} \omega_i \end{cases}$$

$$\begin{aligned} \Rightarrow P\left(\{\omega \in \Omega | \ X(\omega) = k\}\right) &= \binom{n}{k} p^k (1-p)^{n-k} \quad , 0 \leqslant k \leqslant n \\ \textit{Nach Def.:} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} &= 1 \end{aligned}$$

   Somit beschreibt $P'(k) := \binom{n}{k} p^k (1-p)^{n-k}$ eine Wahrscheinlichkeitsverteilung auf $\Omega' = \{0, 1, \ldots, n\}$,
   die Binomialverteilung $b(n, p)$.

2. *Fortsetzung des Problems des Chevalier de Néré*

$$\Omega = \{(\omega_1, \omega_2, \omega_3)|\ \omega_i \in \{1, \dots, 6\}, i = 1, 2, 3\}$$

bzw. $\Omega = \{1, \dots, 6\}^3$ *und Laplaceverteilung. Es wurde bereits gezeigt, dass*

$$P(\sum = 11) = \frac{27}{216} > \frac{25}{216} = P(\sum = 12)$$

Allgemeiner:
Gesucht ist ein

$$q_r := P\left(\{\omega \in \Omega|\ \omega_1 + \omega_2 + \omega_3 = r\}\right), \quad 3 \leqslant r \leqslant 18$$

Klar ist, dass $\sum_{r=3}^{18} q_r = 1$

$$\begin{aligned} q_{11} &= \frac{27}{216} \\ q_{12} &= \frac{25}{216} \\ q_{10} &= q_{11} \\ q_9 &= q_{12} \\ q_8 &= q_{13} \\ &\vdots \end{aligned}$$

Betrachte also:

$$X : \begin{cases} \Omega & \to \quad \mathbb{R} \\ (\omega_1, \omega_2, \omega_3) & \mapsto \quad \omega_1 + \omega_2 + \omega_3 \end{cases}$$

Dann wird durch X *eine neue Wahrscheinlichkeitsverteilung* P^X *auf* $\Omega' = \{3, \dots, 18\}$ *erzeugt, wobei* $P^X(r) = q_r$ *keine Laplaceverteilung mehr ist !*
Beachte:

- *Bei der Abbildung* $X : \Omega \to \mathbb{R}$ *ist* $\mathbb{R}$ *nicht abzählbar, aber* Ω *ist abzählbar. Insbesonders ist dann das Bild von* Ω *unter* X *abzählbar.*
- *Es ist möglich, dass* $\omega_1 \neq \omega_2$, *aber* $X(\omega_1) = X(\omega_2) = x$.
  Dann ist $P^X(x) = P\left(\{\omega \in \Omega|\ X(\omega) = x\}\right)$ *!*

Zur Struktur der Beispiele:

Gegeben ist ein diskreter Wahrscheinlichkeitsraum und
eine Abbildung $X : \Omega \to \mathbb{R}$.
Beschreibt nun

$$\begin{aligned} P^X(x) &= P\left(\{\omega \in \Omega|\ X(\omega) = x\}\right) \quad , x \in X(\Omega) \\ \text{bzw. allgemeiner } P^X(B) &= P\left(\{\omega \in \Omega|\ X(\omega) \in B\}\right) \quad , B \subset \mathfrak{P}(\Omega) \\ &= P(X^{-1}(B)) \end{aligned}$$

eine diskrete Wahrscheinlichkeitsverteilung über $X(\Omega)$ und ist diese abzählbar ?

6.1 Die Urbildfunktion

Definition 6.1 *Sei $T : \Omega \to \Omega'$ eine Abbildung, so heißt*

$$T^{-1} : \begin{cases} \mathfrak{P}(\Omega') & \to \quad \mathfrak{P}(\Omega) \\ A' \subset \Omega' & \to \quad T^{-1}(A') := \{\omega \in \Omega |\ T(\omega) \in A'\} = A \end{cases}$$

die zu T-gehörige Urbildfunktion *oder* Umkehrabbildung/Pseudoinverse

Lemma 6.2 (Eigenschaften der Urbildfunktion) *:*
Seien $A', B', A'_i \in \mathfrak{P}(\Omega')$, so gilt:

1. $$\begin{aligned} T^{-1}(\emptyset) &= \emptyset \\ T^{-1}(\Omega') &= \Omega \end{aligned}$$

2. $$\begin{aligned} T^{-1}(A' \backslash B') &= T^{-1}(A') \backslash T^{-1}(B') \\ \textit{Speziell:} \quad T^{-1}(B'^C) &= \left(T^{-1}(B')\right)^C \end{aligned}$$

3. $T^{-1}(\bigcap\limits_{i \in \mathbb{I}} A'_i) = \bigcap\limits_{i \in \mathbb{I}} T^{-1}(A'_i)$

4. $T^{-1}(\bigcup\limits_{i \in \mathbb{I}} A'_i) = \bigcup\limits_{i \in \mathbb{I}} T^{-1}(A'_i)$
   Speziell für disjunkte B'_i:

   $$T^{-1}\left(\sum_{i \in \mathbb{I}} B'_i\right) = \sum_{i \in \mathbb{I}} T^{-1}(B'_i)$$

   mit $\sum$ als Symbol der disjunktiven Vereinigung

5. $A' \subset B' \Rightarrow T^{-1}(A') \subset T^{-1}(B')$

6. *Ist $S : \Omega' \to \Omega''$ eine beliebige Abbildung, dann gilt:*

   $$(S \circ T)^{-1} = T^{-1} \circ S^{-1}$$

Lemma 6.3 *Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$, dann ist*

$$\begin{aligned} P^X \quad &: \quad \mathfrak{P}(\mathbb{R}) \to [0,1] \\ \textit{mit} \quad P^X(A) &= P(X^{-1}(A)) \\ &= P\left(\{\omega \in \Omega |\ X(\omega) \in A\}\right) \end{aligned}$$

eine diskrete Wahrscheinlichkeitsverteilung über (einer abzählbaren Teilmenge) von $\mathbb{R}$ bzw. $X(\Omega)$

Beweis

P^X ist eindeutig über die Urbildfunktion definiert, also gilt:

$$\begin{aligned} P^X(A) &\geqslant 0 \quad \forall A \subset \mathbb{R} \\ P^X(\mathbb{R}) &= P(X^{-1}(\mathbb{R})) \\ &= P\left(\{\omega \in \Omega | \; X(\omega) \in \mathbb{R}\}\right) \\ &= P(\Omega) = 1 \end{aligned}$$

Zur σ-Additivität:
Sei $(A_i)_{i\in\mathbb{N}}$ eine Folge disjunkter Teilmengen von $\mathbb{R}$, dann gilt:

$$\begin{aligned} P^X\left(\sum_{i=1}^{\infty} A_i\right) &= P\left(X^{-1}\left(\sum_{i=1}^{\infty} A_i\right)\right) \\ &= P\left(\sum_{i=1}^{\infty} X^{-1}(A_i)\right) \\ &= \sum_{i=1}^{\infty} P(X^{-1}(A_i)) \\ &= \sum_{i=1}^{\infty} P^X(A_i) \end{aligned}$$

Für überabzählbare Ω ist eine Einschränkung an der Abbildung X erfoderlich um eine Aussage wie oben zu treffen.

6.2 Messbare Funktionen

Definition 6.4 *Eine Abbildung $X : \Omega \to \Omega'$ von einem messbaren Raum $(\Omega,\mathfrak{A})$ in einen weiteren messbaren Raum $(\Omega', \mathfrak{A}')$ heißt messbar, wenn*

$$\forall A' \in \mathfrak{A}' : \qquad X^{-1}(A') \in \mathfrak{A}$$

Bemerkung:

Die Elemente einer σ-Algebra heißen auch messbare Mengen. Dann besagt obiges:
X heißt messbar, falls Urbilder messbarer Mengen wieder messbar sind.

Definition 6.5 *Eine messbare Funktion im obigen Sinne von einem Wahrscheinlichkeitsraum in einen anderen heißt* ZUFALLSVARIABLE *(ZV) bzw.* ZUFALLSVEKTOR*, wenn $\Omega' = \mathbb{R}$*

Der Begriff der Zufallsvariable ist einer der wesentlichen Aspekte in der Wahrscheinlichkeitstheorie zur Modellierung.
Diese Abbildung beinhaltet das jeweils interessante Merkmal oder den jeweiligen Teilaspekt eines übergeordneten evtl. komplexeren Modells.
Falls der Teilaspekt eines Zufallsexperiments durch eine Zufallsvariable beschrieben wird, kann man sich dann darauf beschränken, nur noch die ZV und deren Verteilung zu betrachten, OHNE Rückgriff auf die explizite Gestalt des zugrundeliegenden Wahrscheinlichkeitsraum.

Bemerkung:

1. Ist $\mathfrak{A} = \mathfrak{P}(\Omega)$ wie beim diskreten Wahrscheinlichkeitsraum, so ist jede Abbildung $X : \Omega \to \Omega'$ messbar
2. `Schreibweise`:
$$\{X \in A\} := X^{-1}(A) = \{\omega \in \Omega |\ X(\omega) \in A\}$$
dann kurz: $P(X \in A)$
3. Die Komposition messbarer Funktionen ist messbar

Definition 6.6 *Das Wahrscheinlichkeitsmaß P^X definiert durch*

$$P^X(A) = P(X^{-1}(A)) \quad , A \subset \mathbb{R}$$

heißt VERTEILUNG VON X UNTER P *oder*
X hat Verteilung P^X, X ist verteilt wie P^X, $X \sim P^X$, $X \sim P$

Beispiel 6.2 (Binary Search) *binäre Suche*
$\Omega = \{0, 1, \ldots, 2^n - 1\}, n \in \mathbb{N}$ *mit Laplaceverteilung.*
(geordnetes Feld der Länge $2^n - 1$ plus zusätzliche Möglichkeit 0)

$$\begin{array}{rcl} \omega \geqslant 1, \omega \in \Omega & : & \textit{mögliche Platznummer des gesuchten Schlüsselelements} \\ \omega = 0 & : & \textit{gesuchtes Element ist nicht vorhanden} \end{array}$$

$$\begin{array}{rcl} A_i & \widehat{=} & \textit{Element wird in genau i Schritten gefunden} \\ i = 1 & : & \textit{Element in der Mitte des Feldes, also } \{2^{n-1}\} \\ i = 2 & : & \{2^{n-2}, 2^n - 1 - 2^{n-2}\} \\ & \vdots & \\ i = k & : & \{(2j-1)2^{n-k} |\ 1 \leqslant j \leqslant 2^{k-1}\} \quad , 1 \leqslant k \leqslant n \end{array}$$

Daraus ergbit sich:

$$|A_i| \;=\; 2^{i-1} \quad \textit{und}$$
$$P(A_i) = \frac{2^{i-1}}{2^n} \quad , 1 \leqslant i \leqslant n$$

Beachte: *Voraussetzung der Laplaceverteilung über $0, 1, \ldots, 2^n - 1$ ergibt*

$$P\,(\textit{Element nicht im Feld}) = \frac{1}{2^n}$$

als implizite Bedingung.

Desweiteren sei:

$$\begin{array}{rcl} B_k & \widehat{=} & \textit{Suche nach höchstens } k \leqslant n \textit{ Schritten beendet} \\ & = & \sum_{i=1}^{k} A_i \\ P(B_k) & = & \sum_{i=1}^{k} P(A_i) \end{array}$$

$$
\begin{aligned}
&= \frac{1}{2^n}\sum_{i=1}^{k} 2^{i-1} \\
&= \frac{2^k-1}{2^n}
\end{aligned}
$$

Der ungünstigste Fall wäre damit B_n:

$$P(\textit{weniger } n \textit{ Schritte}) = P(B_{n-1}) = \frac{2^{n-1}-1}{2^n} < \frac{1}{2}$$

Bis hierher die alte Beschreibung

Jetzt:
Sei die Zufallsvariable X definiert durch

$$X(\omega) = \begin{cases} k & ,\omega \in A_k \\ n & ,\omega \in A_o \cup A_n \quad [A_o = \{0\}] \end{cases}$$

X zählt die Schritte bis zum Abbruch des Verfahrens und ordnet jeder Platznummer ω die in genau $k \in \{1, \ldots, n-1\}$ Schritten erreichbar ist, den Wert k zu.

$X = n$, falls das Schlüsselelement nicht existiert oder die maximale Schrittzahl erreicht wurde.
Die Verteilung P^X ist diskret und bestimmt durch

$$P(X = k) = \begin{cases} P(A_k) = 2^{k-1-n} & ,1 \leqslant k < n \\ P(A_0 \cup A_n) = P(A_0) + P(A_n) = \frac{1}{2^n} + \frac{1}{2} & ,k = n \end{cases}$$

Entsprechend für B_k

$$P(X \leqslant k) = \sum_{i=1}^{k} P(X = i) = \frac{2^k - 1}{2^n}$$

Bezeichnung 6.7

1. *Zufallsvariable X ist binomialverteilt, falls*

$$P^X(k) = P(X = k) = \tbinom{n}{k} p^k (1-p)^{n-k}$$

2. *Zufallsvariable X ist poisson-verteilt, falls für $\lambda > 0$ gilt*

$$P^X(k) = P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad ,k \in \mathbb{N}_0$$

3. *Seien $\Omega \neq \emptyset$, P eine diskrete Wahrscheinlichkeitsverteilung über Ω, $A \subset \Omega$.*
   Die Funktion $\mathcal{I}_A : \Omega \to \mathbb{R}$ definiert durch

$$\mathcal{I}_A(\omega) = \begin{cases} 1 & ,\omega \in A \\ 0 & ,\textit{sonst} \end{cases}$$

heißt INDIKATORFUNKTION *von A und ist eine Zufallsvariable.*
Dabei gilt $\mathcal{I}_A \sim b(1,p)$ mit $p = P(A)$, denn

$$\begin{aligned} P(\mathcal{I}_A = 0) &= P\left(\{\omega \in \Omega |\ \mathcal{I}_A(\omega) = 0\}\right) \\ &= P(A^C) \\ &= 1 - p \end{aligned}$$

$$\begin{aligned} P(\mathcal{I}_A = 1) &= P\left(\{\omega \in \Omega |\ \mathcal{I}_A(\omega) = 1\}\right) \\ &= P(A) \\ &= p \end{aligned}$$

zusätzlich gilt noch:

$$\begin{aligned} \mathcal{I}_{A\cup B} &= max(\mathcal{I}_A, \mathcal{I}_B) \\ \mathcal{I}_{A\cup B} &= \mathcal{I}_A + \mathcal{I}_B, \qquad \textit{falls } A \cap B = \emptyset \\ \mathcal{I}_{A\cap B} &= min(\mathcal{I}_A, \mathcal{I}_B) \\ &= \mathcal{I}_A \cdot \mathcal{I}_B \\ \mathcal{I}_{A^C} &= 1 - \mathcal{I}_A \end{aligned}$$

Zum Zusammenhang zwischen Zufallsvariablen $X_1, \ldots, X_m$ und Zufallsvektor $(X_1, \ldots, X_n)$:

- Aufbau eines Vektors durch verschiedene Zufallsvariablen
- Zerlegung eines Vektors in seine Komponenten

Beispiel 6.3 *Verallgemeinertes Bernoulli-Experiment*
Ein Zufallsexperiment liefert eines von $n \geqslant 2$ Ergebnissen A_i,
etwa fällt eine Maschine aufgrund von Defekt i aus, $\quad 1 \leqslant i \leqslant n$
Voraussetzungen*:*

$$A_1, \ldots, A_m \textit{ sind disjunkt, } P(A_i) = p_i \quad , 1 \leqslant i \leqslant m, \sum_{i=1}^{m} P_i = 1$$

Nun betrachte eine n-malige Versuchsreihe $\Omega = \{1, \ldots, m\}^n$
Frage:

Wie groß ist die Wahrscheinlichkeit, dass sich die Defekte in einer bestimmten Weise verhalten:

$$P\left(|A_1| = k_1, |A_2| = k_2, \ldots, |A_n| = k_n, \sum_{i=1}^{n} k_i = n\right)$$

Beachte, dass die Zufallsvariablen sich gegenseitig verdrängen.
Beschreibe nun die Zufallsvariable X_j die Anzahl der Defekte A_j bei n Versuchen,$1 \leqslant j \leqslant m$:

$$\begin{aligned} P(X_1 = k_1, \ldots, X_m = k_m) &= P^{X_1,\ldots,X_m}(\{(k_1, \ldots, k_n)\}) \\ &= P(\{\omega \in \Omega |\ X_1(\omega) = k_1, \ldots, X_m(\omega) = k_m\}) \\ &:= P(Y_1 \cap \ldots \cap Y_n) \qquad \textit{mit } Y_i := (X_i = k_i) \quad 1 \leqslant i \leqslant n \end{aligned}$$

$$
\begin{aligned}
&= P(Y_1) \cdot \frac{P(Y_1 \cap Y_2)}{P(Y_1)} \cdot \frac{P(Y_1 \cap Y_2 \cap Y_3)}{P(Y_1 \cap Y_2)} \cdot \ldots \cdot \frac{P(Y_1 \cap \ldots \cap Y_n)}{P(Y_1 \cap \ldots \cap Y_{n-1})} \\
&= P(Y_1) \cdot P(Y_2 | \, Y_1) \cdot P(Y_3 | \, Y_1 \cup Y_2) \cdot \ldots \cdot P(Y_n | \, Y_1 \cup \ldots \cup Y_{n-1}) \\
&= \tbinom{n}{k_1} \cdot \tbinom{n-k_1}{k_2} \cdot \ldots \cdot \tbinom{n-k_1-\ldots-k_{n-1}}{k_n} \cdot p_1^{k_1} \cdot \ldots \cdot p_m^{k_n} \\
&= \frac{n!}{k_1! \cdot \ldots \cdot k_m!} \cdot \prod_{i=1}^{m} p_i^{k_j} \qquad , k_j \in \mathbb{N}_0, 1 \leqslant j \leqslant m, \sum_{j=1}^{m} k_j = n
\end{aligned}
$$

Dieser Ausdruck definiert eine endliche Wahrscheinlichkeitsverteilung über

$$\{(x_1, \ldots, x_m) \in \mathbb{R}^m | \; x_i \in \mathbb{N}_0, 1 \leqslant i \leqslant m, \sum_{i=1}^{m} x_i = n\}$$

bzw. über $\mathbb{R}^m$ selbst.

Bezeichnung 6.8
$(X_1, \ldots, X_m)$ *genügt der Multinomial/Polynomialverteilung mit Parametern* $n, p_1, \ldots, p_m$

$$
\begin{aligned}
& M(n, p_1, \ldots, p_m) \\
\textit{Spezialfall } m = 2: \quad M(n, p_1, p_2) &= b(n, p_1) \quad \textit{mit } p_1 + p_2 = 1
\end{aligned}
$$

Definition 6.9 *Seien $X_1, \ldots, X_m$ Zufallsvariablen. Die Verteilung $X = (X_1, \ldots, X_m)$ heißt* GEMEINSAME VERTEILUNG *der Zufallsvariablen $X_1, \ldots, X_m$*
`Schreibweise`*:*

$$P^X = P^{X_1, \ldots, X_m}$$

Die Verteilung von $(X_{i_1}, \ldots, X_{i_l}), 1 \leqslant i_1 < i_2 < \ldots < i_l \leqslant m, 1 \leqslant l < m$ heißt
*l-*DIMENSIONALE RANDVERTEILUNG *oder* MARGINALVERTEILUNG *zu $(i_1, \ldots, i_l)$.*
*Die Verteilung von X_i heißt i-*TE RANDVERTEILUNG *bzw. Marginalverteilung $1 \leqslant i < m$*

Bemerkung:
Die gemeinsame Verteilung ist durch die Angabe aller Wahrscheinlichkeiten
$P(X_1 \in A_1, \ldots, X_m \in A_m) A_i \in \mathfrak{A}$ bestimmt.

Lemma 6.10 *Sei P^X eine diskrete Wahrscheinlichkeitsverteilung über $\mathbb{R}^n$*
Die Randverteilung von $(X_{i_1}, \ldots, X_{i_m})$ wird bestimmt durch

$$
\begin{aligned}
P^{(X_{i_1}, \ldots, X_{i_m})}(B) &= P\left(\{(X_{i_1}, \ldots, X_{i_m})^{-1}(B)\}\right) \\
&= P\left((X_{i_1}, \ldots, X_{i_n})^{-1}(B \times \mathbb{R}^{n-m})\right)
\end{aligned}
$$

$$
\begin{aligned}
&\textit{mit } B \in \mathfrak{P}(\mathbb{R}^m) \\
\textit{und } B \times \mathbb{R}^{n-m} &= \{(x_{i_1}, \ldots, x_{i_m}) \in \mathbb{R}^n | \; (x_{i_1}, \ldots, x_{i_n}) \in B\}
\end{aligned}
$$

$$\textit{Also } P^{(X_{i_1}, \ldots, X_{i_n})}(B) = P^X(B \times \mathbb{R}^{n-m})$$

d.h. von den n Elementen des Zufallsvektors werden m fest gewählt und die Wahrscheinlichkeit in Abhängigkeit von den restlichen $n - m$ frei wählbaren bestimmt.

Fortsetzung des defekte-Maschine Beispiels

Bestimmung der 1.ten Randverteilung, sei $B = \{k\}$

$$\begin{aligned}
P^{X_1} &= P^X(B \times \mathbb{R}^{n-1}) \\
&= \sum_{k_2}\sum_{k_3}\cdots\sum_{k_m} \frac{n!}{k_2!k_3!\ldots k_m!} \cdot \frac{p_1^k}{k_1!}\prod_{i=2}^{m} p_i^{k_i} \\
&\quad \text{mit} \sum_{i=2}^{m} k_i = n - k_1 \\
&= \frac{n!}{(n-k_1)!} \cdot \frac{p_1^k}{k_1!}\sum_{k_2}\cdots\sum_{k_m} \frac{(n-k_1)!}{k_2!k_3!\ldots k_m!} \cdot \prod_{i=2}^{m}\left(\frac{p_i}{1-p_1}\right)^{k_i} \cdot (1-p_1)^{k_i} \\
&= \tbinom{n}{k_1} p_1^{k_1} \underbrace{\sum_{k_2}\cdots\sum_{k_m} \frac{(n-k_1)!}{k_2!k_3!\ldots k_m!} \cdot \prod_{i=2}^{m}\left(\tfrac{p_i}{1-p_1}\right)^{k_i}}_{=1} \underbrace{\prod_{j=2}^{m}(1-p)^{k_j}}_{=(1-p)^{n-k_1}} \\
&\quad \left(\text{Multinomialverteilung mit } n-k_1, \frac{p_2}{1-p_1}, \ldots, \frac{p_n}{1-p_1}\right) \\
&= \tbinom{n}{k_1} p_1^{k_1}(1-p_1)^{n-k_1} \\
&= b(n, p_1)
\end{aligned}$$

Bemerkung:
Die eindeutigen Randverteilungen legen die gemeinsame Verteilung **nicht** eindeutig fest! Seien X, Y Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit

$$\begin{aligned}
X(\Omega) &= Y(\Omega) \\
&= \{0,1\}^2 \\
&= \{(0,0),(1,0),(0,1),(1,1)\}
\end{aligned}$$

$$\begin{aligned}
Y, X &: \Omega \to \mathbb{R} \\
X &= (X_1, X_2) \\
Y &= (Y_1, Y_2) \\
&\text{mit} \\
P^X((0,1)) = P^X((1,0)) &= \frac{1}{2} \\
P^X((1,1)) = P^X((0,0)) &= 0 \\
P^Y((0,1)) = P^Y((1,0)) &= 0 \\
P^Y((1,1)) = P^Y((0,0)) &= \frac{1}{2} \\
\Rightarrow P^X &\neq P^Y
\end{aligned}$$

Aber für die Randverteilungen gilt:

$$P^{X_1} = P^{Y_1} = P^{X_2} = P^{Y_2}$$

$$\text{denn } P^{X_1}(j) = P^X(j \times \Omega)$$

$$
\begin{aligned}
&= P^X(j,0) + P^X(j,1) \\
&= \begin{cases} 0 + \frac{1}{2} & ,j = 0 \\ \frac{1}{2} + 0 & ,j = 1 \end{cases} \\
P^{Y_1}(j) &= P^Y(j \times \Omega) \\
&= P^Y(j,0) + P^Y(j,1) \\
&= \begin{cases} \frac{1}{2} + 0 & ,j = 0 \\ 0 + \frac{1}{2} & ,j = 1 \end{cases}
\end{aligned}
$$

Für P^{X_2} und P^{Y_2} analog.

6.3 Unabhängigkeit von Zufallsvariablen

Definition 6.11 *Eine Familie von Zufallsvariablen $X_i : (\Omega, \mathfrak{A}, P) \mapsto (\Omega_i, \mathfrak{A}_i, P^{X_i}), i \in \mathbb{I}$ heißt* STOCHASTISCH UNABHÄNGIG*, oder die Zufallsvariablen $X_i, i \in \mathbb{I}$ heißen* STOCHASTISCH UNABHÄNGIG*, falls die Mengensysteme $X_i^{-1}(\mathfrak{A}_i)$ stochastisch unabhängig sind.*
D.h. jedes Repräsentantensystem $B_i \in X_i(\mathfrak{A}_i), i \in \mathbb{I}$ bildet eine unabhängige Familie von Ereignissen

Bemerkung:
Die Zufallsvariablen $X_i, i \in \mathbb{I}$ sind stochastisch unabhängig genau dann, wenn

$$
\begin{aligned}
P\left(\bigcap_{i \in \mathbb{J}} \{X_i \in A_i\}\right) &= \prod_{i \in \mathbb{J}} P(X_i \in A_i), \\
&\forall\ \mathbb{J} \subseteq \mathbb{I}, \quad |\mathbb{J}| < \infty \text{ und} \\
&\forall\ A_i \subset \mathfrak{A}_i, i \in \mathbb{J}
\end{aligned}
$$

Satz 6.12 *Sind die Zufallsvariablen $X_i : \Omega \mapsto \Omega_i \quad ,i \in \mathbb{I}$ stochastisch unabhängig und sind die Abbildungen $f_i : \Omega_i \mapsto \Omega_i'$ messbar, dann sind die Zufallsvariablen $f_i \circ X_i \quad ,i \in \mathbb{I}$ stochastisch unabhängig.*

Weiterhin:
Seien $\mathbb{I}_j \subset \mathbb{I} \quad ,j \in \mathbb{J}$ disjunkte Teilmengen und $g_j : \bigtimes_{i \in \mathbb{I}_j} \Omega_i \mapsto \Omega_j \quad ,j \in \mathbb{J}$ messbare Abbildungen, dann gilt:
$g_j \circ (X_i, i \in \mathbb{I}_j) \quad ,j \in \mathbb{J}$ sind stochastisch unabhängige, messbare Funktionen von Zufallsvariablen mit disjunkten Indexmengen.

Beispiel 6.4

$$
\begin{aligned}
X_1, X_2, X_3 \textit{ stochastisch unabhängig} &\rightarrow X_2, (X_1, X_3) \textit{ stochastisch unabhängig} \\
&\rightarrow X_2^2, |X_1 - X_3| \textit{ stochastisch unabhängig}
\end{aligned}
$$

Lemma 6.13 *Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum, dann gilt:*
$X_i \quad ,i \in \mathbb{I}$ sind stochastisch unabhängig genau dann, wenn

$$
P(X_j = x_j, j \in \mathbb{J}) = \prod_{i \in \mathbb{J}} P(X_j = x_j) \quad ,\forall x_j \in X_j(\Omega), \forall j \in \mathbb{J} \quad \forall \mathbb{J} \subseteq \mathbb{I}, |\mathbb{J}| < \infty
$$

Bemerkung:

Es gilt mit obiger Bedingung:

$$P^X = P^{(X_1,\dots,X_n)} = \bigtimes_{i=1}^{n} P^{X_i} \text{ also}$$

$X_1, \dots, X_n$ sind stochastisch unabhängig genau dann, wenn

$$\begin{aligned}
P^X(\bigtimes_{i=1}^{n} A_i) &= P^{(X_1,\dots,X_n)}(\bigtimes_{i=1}^{n} A_i) \\
&= P^X(\{\omega \in (\omega_1, \dots, \omega_n) |\ \omega_i \in A_i\}) \\
&= \prod_{i=1}^{n} \sum_{\omega_i \in A_i} P^{X_i}(\{\omega_i\}) \\
&= \prod_{i=1}^{n} P(X_i \in A_i) \\
&= \prod_{i=1}^{n} P^{X_i}(A_i) \qquad \forall A_i \in \mathfrak{A}_i
\end{aligned}$$

Bemerkung:

Es gilt: Die Zufallsvariablen X und Y sind stochastisch unabhängig unter P genau dann, wenn die Zufallsvariablen $f(X)$ und $g(Y)$ stochastisch unabhängig sind $\forall f, g : \mathbb{R}^1 \to \mathbb{R}^1 \quad , f, g$ messbar

6.4 Verteilung der Summe zweier unabhängiger Zufallsvariablen

Satz 6.14
Seien X, Y stochastisch unabhängige Zufallsvariablen auf $\mathbb{Z}$ mit den Zähldichten f und g (d.h. $P(X = n) = f(n), P(Y = m) = g(m)$)
Dann hat $X + Y$ die Zähldichte h, gegeben durch:

$$\begin{aligned}
h(k) &= \sum_{j\in\mathbb{Z}} f(j) \cdot g(k-j) \\
&= \sum_{j\in\mathbb{Z}} f(k-j) \cdot g(j) \qquad , k \in \mathbb{Z} \\
&= P(X + Y) = k
\end{aligned}$$

Bezeichnung 6.15 *h ist die sogenannte* FALTUNG *der Dichten f und g: $h = f * g$*

Beweis:

$$\begin{aligned}
\text{Es ist } \sum_{j\in\mathbb{Z}} P(X = j) &= \sum_{j\in\mathbb{Z}} P(Y = j) = 1 \\
\Rightarrow P(X + Y = k) &\overset{\text{totale W'}}{=} \sum_{j\in\mathbb{Z}} P(X + Y = k |\ Y = j) \\
&= \sum_{j\in\mathbb{Z}} P(X + j = k |\ Y = j)
\end{aligned}$$

$$\overset{\text{unabh.}}{=} \sum_{j\in\mathbb{Z}} \underbrace{P(X=k-j)}_{f(k-j)} \underbrace{P(Y=j)}_{g(j)}$$

Beispiel 6.5

1. *Seien X, Y stochastisch unabhängige, nach $b(1,p)$ verteilte Zufallsvariablen (Münzwurf) (d.h. $P(X=0)=1-p, P(X=1)=p$)*

$$\begin{aligned}
\Rightarrow P(X+Y=k) &= \begin{cases} P(X=0)\cdot P(Y=0) & ,k=0 \\ P(X=1)\cdot P(Y=0)+P(X=0)\cdot P(Y=1) & ,k=1 \\ P(X=1)\cdot P(Y=1) & ,k=2 \end{cases} \\
&= \begin{cases} (1-p)^2 & ,k=0 \\ 2p(1-p) & ,k=1 \\ p^2 & ,k=2 \end{cases} \\
&= \tbinom{2}{k} p^k (1-p)^{2-k} \qquad ,k\in\{0,1,2\} \\
&= b(2,p)
\end{aligned}$$

   per Induktion folgt:
   Seien $X_1,\ldots,X_n$ stochastisch unabhängige, nach $b(1,p)$ verteilte Zufallsvariablen.
   $\Rightarrow \sum_{i=1}^n X_i$ besitzt eine $b(n,p)$ Verteilung.

2. *Seien X, Y stochastisch unabhängige Zufallsvariablen, mit $X\sim po(\lambda), Y\sim po(\mu)$ $\quad ,\lambda,\mu>0$*

$$\begin{aligned}
P(X+Y=k) &= \sum_{j\in\mathbb{N}_0} P(X=k-j)\cdot P(Y=j) \\
&= \sum_{j=0}^{k} \frac{\lambda^{k-j}}{(k-j)!} e^{-\lambda}\cdot \frac{\mu^j}{j!} e^{-\mu} \\
&= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^{k} \tbinom{k}{j} \lambda^{k-j}\mu^j \\
&= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda+\mu)^k \qquad \textit{(Binomialsumme)}
\end{aligned}$$

   d.h. $X+Y\sim po(\lambda+\mu)$ wieder Poissonverteilt

Bisher zur Beschreibung des Zufalls:

$$\begin{aligned}
\text{Zähldichte} \qquad & p: && \sum_{\omega\in\Omega} p(\omega)=1 \\
\text{Wahrscheinlichkeitsverteilung} \qquad & P: && \mathfrak{A}\to[0,1] \\
& P(A)= && \sum_{\omega\in A} p(\omega) \qquad ,A\in\mathfrak{A}
\end{aligned}$$

6.5 Verteilungsfunktionen

Definition 6.16 *Seien $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable mit Wahrscheinlichkeitsverteilung P^X.*
Die Funktion

$$F^X = \begin{cases} \mathbb{R}^1 & \to & \mathbb{R}^1 \\ x & \mapsto & P^X((-\infty, x]) \end{cases}$$

heißt die zu P^X gehörige VERTEILUNGSFUNKTION

`Schreibweisen`*:*

> *F Verteilungsfunktion von X*
> *X verteilt nach F*
> *$X \sim F$*

Bemerkung:

1. Sei p^X die Zähldichte von P^X, dann ist

$$\begin{aligned} F^X(x) &= P^X((-\infty, x]) \\ &= \sum_{\omega \in (-\infty, x]} p^X(\omega) \\ &= \sum_{\substack{\omega \leqslant x \\ \omega \in supp(P^X)}} p^X(\omega) \qquad , x \in \mathbb{R} \end{aligned}$$

2. Für $P^X((-\infty, x])$ schreibt man

$$\begin{aligned} P^X((-\infty, x]) &= P\left(X^{-1}((-\infty, x])\right) \\ &= P(\{\omega \in \Omega |\ X(\omega) \in (-\infty, x]\}) \\ &\overset{\text{Bez.}}{=} P(X \leqslant x) \end{aligned}$$

   d.h. $F^X(x)$ ist die Wahrscheinlichkeit dafür, dass die Zufallsvariable X Werte kleiner oder gleich x annimmt

Lemma 6.17 *Sei F^X die zu P^X gehörige Verteilungsfunktion. Dann gilt:*

1. *(a) F^X ist monoton wachsend*
   (b) F^X ist rechtsseitig stetig
   (c) $\lim_{x \to +\infty} F^X(x) = 1$
   $\lim_{x \to -\infty} F^X(x) = 0$
2. *P^X ist durch F^X eindeutig bestimmt.*

Beweis:

1. (a) Sei $x_1 \leqslant x_2$ beliebig. Wegen $(-\infty, x_1] \subset (-\infty, x_2]$ ist

$$F^X(x_1) = P^X((-\infty, x_1]) \leqslant P^X((-\infty, x_2]) = F^X(x_2)$$

   (b) Sei $(x_n)_{n\in\mathbb{N}}$ eine monoton fallend Folge mit $\lim\limits_{n\to\infty} x_n = x$

   Mit $A := (-\infty, x_n] \quad , k \in \mathbb{N}$, $A_n \supset A_{n+1} \quad , n \in \mathbb{N}$ und $\bigcup\limits_{n=1}^{\infty} A_n = (-\infty, x] =: A$ gilt dann:

$$\overset{\text{Stetigkeit von oben}}{\Rightarrow} \quad \begin{aligned} \lim_{n\to\infty} F^X(x_n) &= \lim_{n\to\infty} P^X(A_n) \\ &= P^X(A) \\ &= F^X(A) \end{aligned}$$

   (c) Sei $(x_n)_{n\in\mathbb{N}}$ eine monoton wachsende Folge mit $\lim\limits_{n\to\infty} x_n = \infty$, dann ist $A_n \uparrow \mathbb{R}$ für $A := (-\infty, x_n]$

$$\Rightarrow \begin{aligned} \lim_{n\to\infty} F^X(x_n) &= \lim_{n\to\infty} P^X(A_n) \\ &= P^X(\mathbb{R}) \\ &= 1 \end{aligned}$$

   Zweiter Grenzwert analog.

Bemerkung:

1. F^X ist durch P^X eindeutig bestimmt, dann sei F^X die Verteilungsfunktion zu P' mit der Zähldichte p'

$$\Rightarrow p'(x) = F^X(x) - F^X(x-) = p(x) \qquad , \forall x \in \mathbb{R}$$

   mit $x- := \lim\limits_{h\to 0} x - h$

2. Es gilt $p^X(x) = F^X(x) - F^X(x-)$ (s.o.)
   d.h. F^X ist stetig in x genau dann, wenn die Zähldichte $p^X(x) = 0$ ist (sonst Sprung)

$$\begin{aligned} p^X(x) = 0 \quad &\Rightarrow \quad F^X \text{ ist linksseitig stetig} \\ \text{obiges Lemma} \quad &\Rightarrow \quad F^X \text{ ist rechtsseitig stetig} \\ &\Rightarrow \quad F^X \text{ ist stetig} \end{aligned}$$

   Weiterhin gilt, dass es höchstens abzählbar viele Punkte $x \in \mathbb{R}$ mit $p^X(x) > 0$ gibt, sonst wäre $\sum p(\omega) = \infty$
   $\Rightarrow$ Es existiert höchstens abzählbare viele Unstetigkeitsstellen der Verteilungsfunktion.

3. Zum vorigen Lemma:
   Die Wahrscheinlichkeitsverteilung über $\mathbb{R}$ ist eindeutig bestimmt durch

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

4. Sei $\{x_1, x_2, \ldots\}$ $, x_i < x_{i+1}, i \in \mathbb{N}$ eine Abzählung im üblichen Sinne des Trägers von p^X, dann gilt:

$$P^X(\underbrace{(-\infty, x_{i+1})}_{\text{offenes Intervall}}) = \underbrace{P^X((-\infty, x_i])}_{=F^X(x_i)} + \underbrace{P^X((x_i, x_{i+1}))}_{=0\text{ ,da nicht Träger}}$$

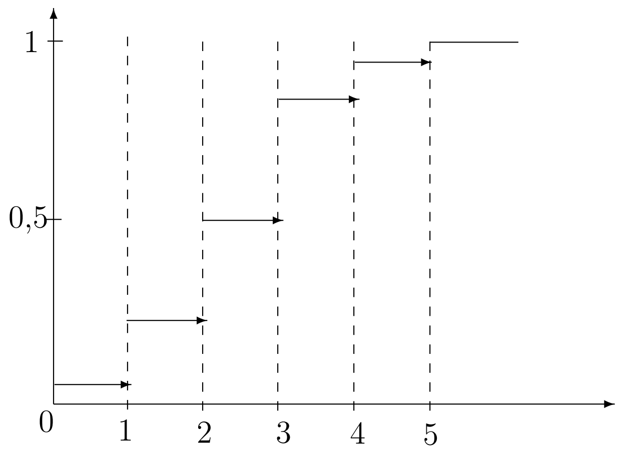
d.h. F^X ist eine sog. Treppenfunktion:
Sprünge an den Trägerpunkten und sonst konstant.

Beispiel 6.6 *Verteilungsfunktion von* $b(5, \frac{1}{2})$

$$p^X(x) = \begin{cases} \underbrace{\binom{5}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{5-x}}_{\left(\frac{1}{2}\right)^5} & , x \in \{0, \ldots, 5\} \\ 0 & , sonst \end{cases}$$

$$supp(p^X) = \{0, \ldots, 5\}$$

$$F^X(x) = \begin{cases} 0 & , x < 0 \\ \frac{1}{2^5} \sum_{i=0}^{\lfloor x \rfloor} \binom{5}{i} & , 0 \leqslant x \leqslant 5 \\ 1 & , x > 5 \end{cases}$$



6.6 Mehrdimensionale Zufallsvariablen

Definition 6.18 *Sei* $X = (X_1, \ldots, X_n)$ *ein Zufallsvektor mit der Wahrscheinlichkeitsverteilung* P^X*. Die durch*

$$F^X := P^X((-\infty, x_1] \times \ldots \times (-\infty, x_n]) \qquad , (x_1, \ldots, x_n) \in \mathbb{R}^n$$

definierte Funktion heißt MULTIVARIANTE VERTEILUNGSFUNKTION.

Bemerkung:

Sind X_1, X_2 stochastisch unabhängig, so gilt:

$$F^{(X_1,X_2)}(x_1,x_2) = F^{X_1}(x_1) \cdot F^{X_2}(x_2) \qquad ,(x_1,x_2) \in \mathbb{R}^2$$

Beweis direkt aus Definition mit $A_i = (-\infty, x_i]$

Insbesonders ist bei stochastisch unabhängigen Zufallsvariablen die gemeinsame Verteilungsfunktion eindeutig durch die Verteilungsfunktionen der eindimensionalen Randwerte bestimmt.

7 Erwartungswerte

Beispiel 7.1 (Würfelspiel)
Ergebnis: $i = \{1, \dots, 6\} \Rightarrow$ *Auszahlung* i *Euro.*
Gesucht: Durschnittlich zu erwartende Auszahlung

$$\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \ldots + \frac{1}{6} \cdot 6 = 3,5$$

3,5 ist der ERWARTUNGSWERT *der Zufallsvariable*

$$X : \begin{cases} \{1, \dots, 6\} & \rightarrow & \mathbb{R} \\ i & \mapsto & i \end{cases}$$

wobei P^X *laplaceverteilt über* $\{1, \dots, 6\}$ *ist.*

Definition 7.1 *Sei* (Ω, P) *ein diskreter Wahrscheinlichkeitsraum.*

1. *Sei* $X(\Omega) \subset \mathbb{R}^+$ *oder* $X(\Omega) \subset \mathbb{R}^-$

$$EX := E(X) := \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\})$$

heißt ERWARTUNGSWERT *von* X *unter* P

2. *Sei* X *eine allgemeine Zufallsvariable mit*

$$E(max(X,0)) < \infty \text{ oder } E(min(X,0)) > -\infty$$

dann heißt

$$EX := E(X) := \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\})$$

der ERWARTUNGSWERT *von* X *unter* P.
(Definition verhindert „$-\infty + \infty$“ in der Summe)

Bemerkung:

1. Für nicht-negative Zufallsvariablen ist E immer wohldefiniert. $E(X) = \infty$ ist erlaubt

2. Für Zufallsvariablen mit positiven und negativen Werten muß die Wohldefiniertheit der Reihe
$$\sum_{\omega\in\Omega} X(\omega)\cdot P(\{\omega\})$$
gewährleistet werden.
Werden nur endliche Erwartungswerte betrachtet, kann absolute Konvergenz
$\sum\limits_{\omega\in\Omega} |X(\omega)|P(\{\omega\}) < \infty$
gefordert werden, d.h. die Änderung der Summationsreihenfolge ist erlaubt.

3. Im Folgenden wird stets die Wohldefiniertheit der auftretenden Erwartungswerte vorausgesetzt.

4. $E(X)$ hängt von der Verteilung von X ab:
Sei $x_1, x_2, \ldots$ eine Abzählung von $X(\Omega)$, dann gilt:
$$\begin{aligned} E(X) &= \sum_{i=1}^{\infty} \sum_{\{\omega|\ X(\omega)=x_i\}} X(\omega)\cdot P(\{\omega\}) \\ &= \sum_{i=1}^{\infty} x_i\cdot P(X=x_i) \\ &= \sum_{i=1}^{\infty} x_i\cdot P^X(x_i) \end{aligned}$$
d.h. $E(X)$ kann ebenso über die Summe $\sum x_i\cdot P^X(x_i)$ erklärt werden:
$$E(X) = \sum_{x\in\mathbb{R}} x\cdot P^X(x) \qquad , P^X(x) = 0 \text{ falls } x\notin supp(P)$$

5. Der Erwartungswert (als mögliche Kenngröße) dient dem Vergleich von Verteilungen.

Beispiel 7.2 *1. Sei $X \sim b(n,p)$:*
$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k\cdot P^X(k) \\ &= \sum_{k=1}^{n} k\cdot \tbinom{n}{k} p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n} n\tbinom{n-1}{k-1} p^k(1-p)^{n-k} \\ &= n\cdot p\sum_{k=1}^{n} \tbinom{n-1}{k-1} p^{k-1}(1-p)^{n-k} \\ &= n\cdot p\underbrace{\sum_{k=0}^{n-1} \tbinom{n-1}{k} p^k(1-p)^{((n-1)-k)}}_{=1,\ b(n-1,p)-Verteilung} \end{aligned}$$
Nach binomischen Lehrsatz:

$$\begin{aligned}&(x+y)^n = \sum_{k=0}^{n} \tbinom{n}{k} x^k y^{n-k}\\ =\ & n \cdot p \cdot (p + (1-p))^{n-1}\\ =\ & n \cdot p\end{aligned}$$

2. *Sei $X \sim po(\lambda)$*

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

3. *Fortsetzung der „binären Suche“*
   Zufallsvariable X zählt die Schritte bist zum Abbruch des Verfahrens.

$$X(\omega) \in \{1, \ldots, n\}$$

$$P(X = k) = \begin{cases} 2^{k-1-n} & , 1 \leqslant k \leqslant n-1 \\ \frac{1}{2^n} + \frac{1}{2} & , k = n \end{cases}$$

$$\begin{aligned}\Rightarrow E(X) &= \sum_{i=1}^{n} i \cdot P(X = i)\\ &= \sum_{i=1}^{n-1} i \cdot 2^{i-1-n} + n\left(\frac{1}{2^n} + \frac{1}{2}\right)\\ &= \sum_{i=1}^{n} i \cdot 2^{i-1-n} + \frac{n}{2^n}\\ &= \ldots\\ &= n - 1 + \frac{n+1}{2^n}\end{aligned}$$

d.h. die erwartete Schrittzahl bis zum Abbruch des Algorithmus ist (für große n) praktisch nur um 1 besser als im „worst case“
($\hookrightarrow$ Average Case Analysis)

4. *Zufallsvariable X mit Verteilung P^X und Träger $\mathbb{N}$*

$$\begin{aligned}P^X(n) &= const \cdot \frac{1}{n^2}\\ & \mathit{mit} \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \ \mathit{folgt}\ const = \frac{6}{\pi^2}\end{aligned}$$

$$\begin{aligned} \Rightarrow E(X) &= \sum_{n=1}^{\infty} n \cdot const \cdot \frac{1}{n^2} \\ &= \infty \end{aligned}$$

5. *Zufallsvariable X mit Verteilung P^X*
   Träger: $\{x_i, i \in \mathbb{N} |\ x_i = (-1)^i \cdot \frac{2^i}{i}\} \subset \mathbb{R}$
   und $P^X(x_i) = \frac{1}{2^i}$ $\quad \sum\limits_{i=1}^{\infty} \frac{1}{2^i} = 1$ Dann:

$$\begin{aligned} E(max(X,0)) &= \sum_{i=1}^{\infty} \frac{2^{2i}}{2i} \cdot \frac{1}{2^{2i}} && \textit{alle geraden Zahlen} \\ &= \sum_{i=1}^{\infty} \frac{1}{2i} \\ &= \infty \\ E(min(X,0)) &= -\sum_{i=1}^{\infty} \frac{2^{2i-1}}{2i-1} \cdot \frac{1}{2^{2i-1}} && \textit{alle ungeraden Zahlen} \\ &= -\sum_{i=1}^{\infty} \frac{1}{2i-1} \\ &= -\infty \end{aligned}$$

   d.h. der Erwartungswert von X unter P existiert nicht, aber

$$\sum_{i=1}^{\infty} x_i P^X(x_i) = \sum_{i=1}^{\infty} (-1)^i \cdot \frac{1}{i} \overset{Leibniz}{=} -\ln(2)$$

   $\Rightarrow$ das Ergebnis hängt von der Summationsreihenfolge ab und ist nicht absolut konvergent

6. *Sei $X = Ind(A) = \mathcal{I}_A$, dann*

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} \mathcal{I}_A(\omega) \cdot p(\omega) \\ &= \sum_{\omega \in A} p(\omega) \\ &= P(A) \end{aligned}$$

Für Zufallsvariablen mit Werten in $\mathbb{N}_0$ gibt es andere Berechnungsmöglichkeiten für den Erwartungswert.
Dazu zunächst:

Lemma 7.2 *Sei $(a_n)_{n \in \mathbb{N}}$ eine Folge aus $\mathbb{R}^+$,*
$b_j := \sum\limits_{n=j}^{\infty} a_n$, dann gilt:

$$\sum_{j=1}^{\infty} b_j = \sum_{n=1}^{\infty} n \cdot a_n$$

Beweis:

1.Fall: $\sum\limits_{n=1}^{\infty} n \cdot a_n < \infty$

$$\begin{aligned}
\sum_{n=1}^{\infty} n \cdot a_n &= \sum_{n=1}^{\infty} \sum_{j=1}^{n} a_n \\
&= a_1 + a_2 + a_2 + a_3 + a_3 + a_3 + \ldots \qquad \text{umsortieren, da absolut konvergent} \\
&= \begin{array}{lllll} a_1 & +a_2 & +a_3 & +a_4 & +\ldots \\ & +a_2 & +a_3 & +a_4 & +\ldots \\ & & +a_3 & +a_4 & +\ldots \\ & & & \ldots & \end{array} \\
&= \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} a_n \\
&= \sum_{j=1}^{\infty} b_j
\end{aligned}$$

2.Fall: $\sum\limits_{n=1}^{\infty} n \cdot a_n = \infty$, also

$$\forall N \geqslant 0 \quad \exists k_0 \in \mathbb{N} \text{ mit } \sum_{n=1}^{k} n \cdot a_n \geqslant N \quad \forall k \geqslant k_0$$

Für solche k gilt:

$$\begin{aligned}
N &\leqslant \sum_{n=1}^{k} n \cdot a_n \\
&= \sum_{n=1}^{k} a_n \cdot \sum_{j=1}^{n} 1 \\
&= \sum_{j=1}^{k} \sum_{n=j}^{k} a_n \\
&\leqslant \sum_{j=1}^{k} \sum_{n=j}^{\infty} a_n \\
&= \sum_{j=1}^{k} b_j \leqslant \sum_{j=1}^{\infty} b_j
\end{aligned}$$

das heißt:

$$\sum_{j=1}^{\infty} b_j \geqslant N \quad \forall N \geqslant 0 \quad \Rightarrow \sum_{j=1}^{\infty} b_j = \infty$$

Korrolar 7.3 *Sei $(\Omega, \mathfrak{A}, P)$ ein diskreter Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{N}_0$, dann gilt:*

$$E(X) = \sum_{n=1}^{\infty} n \cdot P^X(n)$$

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} P^X([n,\infty)) \\
&= \sum_{n=1}^{\infty} P(X \geqslant n)
\end{aligned}
$$

Der Beweis folgt direkt aus dem vorigen Lemma mit
$a_n := P^X(n), n \in \mathbb{N}$ und $b_n = P^X([n,\infty))$

Anmerkung:

$$
\begin{aligned}
P(X \geqslant n) &= P(X > n-1) \\
&= 1 - P(X \leqslant n-1) \\
&= 1 - F^X(n-1)
\end{aligned}
$$

d.h., dass die Verteilungsfunktion als Ausgangspunkt zur Erwartungswertbildung benutzt werden kann.

7.1 Geometrische Wahrscheinlichkeitsverteilung

Beispiel 7.3 *Münzwurf $b(n,p)$ mit $p \in (0,1)$ für „Kopf“*
$\hookrightarrow$ *n-facher Münzwurf*
$\hookrightarrow$ *Zufallsvariable X beschreibe die Anzahl der Würfe bis zum ersten mal „Kopf“*

Damit ist die Wahrscheinlichkeit, dass Kopf zum ersten Mal im n-ten Versuch auftritt:

$$P^X(n) = (1-p)^{n-1} \cdot p \qquad , n \in \mathbb{N}$$

$$
\begin{aligned}
E(X) &= \sum_{n=1}^{\infty} P(X \geqslant n) \\
&= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P^X(k) \\
&= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} p \cdot (1-p)^{k-1} \\
&= \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} p \cdot (1-p)^{((k+n-1)-1)} \\
&= p \cdot \sum_{n=1}^{\infty} \left((1-p)^{n-1} \sum_{k=1}^{\infty} (1-p)^{k-1} \right) \qquad \textit{geometrische Reihen} \\
&= p \cdot \frac{1}{p} \cdot \frac{1}{p} = \frac{1}{p}
\end{aligned}
$$

Bezeichnung 7.4 *Die Wahrscheinlichkeitsverteilung $P^X(n) = (1-p)^{n-1} \cdot p \qquad , n \in \mathbb{N}$ nennt man* GEOMETRISCHE VERTEILUNG

Betrachte nun E' bei Abbildung von Zufallsvariablen.

Satz 7.5 *Sei* $(\Omega, \mathfrak{P}(\Omega)\ , P)$ *ein diskreter Wahrscheinlichkeitsraum,*
$X : \Omega \to \mathbb{R}^k$ *ein Zufallsvektor und*
$f : \mathbb{R}^k \to \mathbb{R}$ *eine messbare Abbildung.*

Ferner existiere der Erwartungswert der Zufallsvariablef $\circ X$*.*
Dann gilt:

$$\begin{aligned} E(f \circ X) &= \sum_{\omega \in \Omega} (f \circ X)(\omega) \cdot P(\omega) && (1) \\ &= \sum_{x \in \mathbb{R}} x \cdot P^{f \circ X}(x) && (2) \\ &= \sum_{t \in \mathbb{R}^k} f(t) \cdot P^X(t) && (3) \\ &= E_{P^X}(f) && (4) \end{aligned}$$

Gleichung 1 setzt Kenntnis des Grundraums voraus, 3 nicht !

Beweis:

$$\begin{aligned} E(f \circ X) &= \sum x \in \mathbb{R} P^{f(x)}(x) \\ &= \sum_{x \in \mathbb{R}} x \cdot P\left(X^{-1}(f^{-1}(x))\right) \\ &= \sum_{x \in \mathbb{R}} x \cdot P^X(\{f = x\}) \\ &= \sum_{x \in \mathbb{R}} x \cdot \sum_{t \in \{f = x\}} P^X(t) \\ &= \sum_{x \in \mathbb{R}} \sum_{t \in \{f = x\}} f(t) P^X(t) \\ &= \sum_{t \in \mathbb{R}^k} f(t) \cdot P^X(t) \end{aligned}$$

7.2 Eigenschaften von Erwartungswerten

Lemma 7.6 *Seien X, Y Zufallsvariablen mit endlichen Erwartungswerten, $a \in \mathbb{R}$, dann gilt:*

1. $E(a) = a$
2. $E(aX) = a \cdot E(X)$ *(Skalarität)*
3. $E(|X + Y|) \leqslant E(|X|) + E(|Y|)$ *(Dreiecksungleichung)*
4. $E(X + Y) = E(X) + E(Y)$ *(Additivität)*

   Zusammen mit 2 ergibt sich die Linearität
5. $X \leqslant Y \Rightarrow \; E(X) \leqslant E(Y)$ *(Ordnungserhaltung)*

   Speziell: $Y \geqslant 0 \Rightarrow E(Y) \geqslant 0$

   $E(X) \leqslant E(|X|)$
6. $E(|X|) = 0 \iff P(X \neq 0) = 0$

Beweis:

1. $E(a) = \sum_{\omega \in \Omega} a \cdot p(\omega) = a \cdot \sum_{\omega \in \Omega} p(\omega) = a \cdot 1$

2. $E(a) = \sum_{\omega \in \Omega} aX(\omega) \cdot P(\omega) = a \cdot \sum_{\omega \in \Omega} X(\omega)P(\omega) = a \cdot E(X)$

3.
$$\begin{aligned} E(|X+Y|) &= \sum_{\omega \in \Omega} |X(\omega) + Y(\omega)| \cdot P(\omega) \\ &\leqslant \sum_{\omega \in \Omega} (|X(\omega)| + |Y(\omega)|) \cdot P(\omega) \\ &= \sum_{\omega \in \Omega} |X(\omega)| \cdot P(\omega) + \sum_{\omega \in \Omega} |Y(\omega)| \cdot P(\omega) \\ &= E(|X|) + E(|Y|) \end{aligned}$$

4. Trivial

5. Trivial

6.
$$\begin{aligned} E(|X|) = 0 &\iff \sum_{\omega \in \Omega} |X(\omega)| \cdot P(\omega) = 0 \\ &\iff |X(\omega)| \cdot P(\omega) = 0 \qquad \forall \omega \in \Omega \\ &\iff X(\omega) \cdot P(\omega) = 0 \qquad \forall \omega \in \Omega \\ &\iff P(X \neq 0) = 0 \end{aligned}$$

Lemma 7.7 *Seien $(X_i)_{i\in\mathbb{I}}$ Zufallsvariablen mit endlichen Erwartungswerten, dann gilt:*

1. $E(\sup_{i\in\mathbb{I}} X_i) \geqslant \sup_{i\in\mathbb{I}} E(X_i)$
2. $E(\inf_{i\in\mathbb{I}} X_i) \leqslant \inf_{i\in\mathbb{I}} E(X_i)$

Beweis:

1. Für ein $i_0 \in \mathbb{I}$ gilt:

$$\begin{aligned} E(\sup_{i\in\mathbb{I}} X_i) &= \sum_{\omega\in\Omega} (\sup_{i\in\mathbb{I}} X_i(\omega)) \cdot P(\omega) \\ &\geqslant \sum_{\omega\in\Omega} X_{i_0}(\omega) \cdot P(\omega) \\ &= E(X_{i_0}) \end{aligned}$$

Da i_0 beliebig gewählt ist, gilt dies auch für das Infimum.

2. analog oder mit $-\sup_{i\in\mathbb{I}} X_i = \inf_{i\in\mathbb{I}}(-X_i)$

Satz 7.8 (Multiplikationssatz)
X, Y *stochastisch unabhängig Zufallsvariablen und* $E(|X|) \leqslant \infty \geqslant E(|Y|)$, *dann gilt:*

$$E(X \cdot Y) \leqslant \infty \text{ und } E(X \cdot Y) = E(X) \cdot E(Y)$$

Beweis:

Sei $\{x_i, i \in \mathbb{N}\} = supp(X)$ und
$\{y_i, i \in \mathbb{N}\} = supp(Y)$

$$\begin{aligned} E(X) \cdot E(Y) &= \left(\sum_{i=1}^{\infty} x_i P^X(x_i)\right) \cdot \left(\sum_{j=1}^{\infty} y_j P^Y(x_j)\right) \\ &\overset{\text{konv.}}{=} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j \cdot P^X(x_i) \cdot P^Y(x_j) \\ &\overset{\text{unabh.}}{=} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j P^{(X,Y)}(x_i, y_j) \\ &= \sum_{z\in\mathbb{R}} z \cdot P^{(X\cdot Y)}(z) \\ &= E(X \cdot Y) \end{aligned}$$

folgt aus Satz 7.5 mit $f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x, y) = x \cdot y$

7.3 Moment und Varianz

Definition 7.9 *Seien X, Y Zufallsvariablen und $c \in \mathbb{R}, k \in \mathbb{N}$:*

1. *$E((X-c)^k)$ heißt k-TES MOMENT von X und c (nichtzentrales Moment, $c=0$: (zentrales) Moment)*

2. *$E((X-EX)^2)$ heißt VARIANZ oder STREUUNG von X kurz: $Var(X)$ oder $VarX$*

3. *$E((X-EX)(Y-EY))$ heißt KOVARIANZ von X und Y kurz: $Cov(X,Y)$*

Satz 7.10 (Jensen'sche Ungleichung für Momente)
Seien X eine Zufallsvariable, $f:\mathbb{R}\to\mathbb{R}$ eine konvexe Funktion, so dass $E(f\circ X)$ und $E(X)$ existieren, dann gilt:

$$E(f\circ X)\geqslant f(EX)$$

Speziell $E(X^2)\geqslant (EX)^2$

Beweis siehe Plachkey Seite 133/134.

Korrolar 7.11 *Seien X,Y reellwertige Zufallsvariablen.*

1. *Aus $0\leqslant |X|\leqslant |Y|,\quad E(|Y|)\leqslant\infty\Rightarrow EX$ existiert und $E(|X|)\leqslant\infty$*

2. *Aus $E(X^k)\leqslant\infty$ für ein $k\in\mathbb{N}\Rightarrow E(X^l)\leqslant\infty\quad\forall\ l\leqslant k$*

3. *Aus $E(X^2)\leqslant\infty\Rightarrow E((X+a)^2)\leqslant\infty\quad\forall a\in\mathbb{R}$ insbesonders $Var(X)\leqslant\infty$*

Korrolar 7.12 (von Ljupunoff)
Sei X eine reellwertige Zufallsvariable, $E(|X|^r)<\infty$ für $r\in(0,\infty)$
Dann existiert auch $E(|X|^s)\quad\forall\ 0<s\leqslant r$ und

$$(E(|X|^r))^{\frac{1}{r}}\geqslant (E(|X|^s))^{\frac{1}{s}}$$

Beweis:

$|X|^s\leqslant |X|^r+1\Rightarrow$ Es existiert $E(|X|^s)$
Nach Satz von Jensen mit $f(X)=|X|^{\frac{r}{s}}$ und $|X|^s$ folgt

$$E(|X|^r)=E\left((|X|^s)^{\frac{r}{s}}\right)\geqslant (E(|X|^s))^{\frac{r}{s}}$$

7.4 Eigenschaften der Varianz

Lemma 7.13 *Sei $X<\infty,\quad a,b\in\mathbb{R}$*

1. *$Var(aX+b)=a^2Var(X)$*

2. *$Var(X)=E(X^2)-(EX)^2$*
   Häufige Methode zur Berechnung von $Var(X)$

3. *$Var(X)=0\Longleftrightarrow P(X\neq EX)=0$*

4. *$Var(X)=\min\limits_{a\in\mathbb{R}}E((X-a)^2)$*

Beweis:

1.

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

2.

$$
\begin{aligned}
Var(X) &= E((X-EX)^2) \qquad (\mu := EX) \\
&= \sum_{i=1}^{\infty} (x_i-\mu)^2 P^X(x_i) \\
&= \sum_{i=1}^{\infty} x_i^2 P^X(x_i) - 2\mu \cdot \sum_{i=1}^{\infty} x_i P^X(x_i) + (\mu^2 \cdot 1) \\
&= E(X^2) - 2(EX)^2 + (EX)^2
\end{aligned}
$$

3.

$$
\begin{aligned}
Var(X) = 0 &\iff \sum_{x_i \in supp(X)} (x_i-\mu)^2 P^X(x_i) = 0 \\
&\iff (x_i - \mu) = 0 \qquad \forall\, x_i \in supp(X)
\end{aligned}
$$

4.

$$
\begin{aligned}
E((X-a)^2) &= E((X-\mu+\mu-a)^2) \\
&= E(X-\mu)^2 + 2(\mu-a) \cdot \underbrace{E(X-\mu)}_{=0} + (\mu-a)^2 \\
&= Var(X) + (\mu-a)^2 \geqslant Var(X)
\end{aligned}
$$

d.h. $a = EX$ minimiert die mittlere quadratische Abweichung von X

Bezeichnung 7.14
Eine Zufallsvariable X mit $EX = 0$ und $Var(X) = 1$ heißt STANDARDISIERT
Weiterhin:
Sei Y Zufallsvariable mit $EY = \mu < \infty$ und $0 < Var(Y) =: \sigma^2$, dann gilt:

$$
X := \frac{Y - EY}{\sqrt{Var(Y)}} = \frac{Y-\mu}{\sigma} \text{ erfüllt gerade } EX = 0 \text{ und } Var(X) = 1
$$

Dieser Vorgang heißt STANDARDISIERUNG
Dies hat den Vorteil der Tabellierung und beim Vergleichen.

Satz 7.15 *Seien X, Y Zufallsvariablen mit $Var(X), Var(Y) < \infty$, dann gilt*

$$
Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y)
$$

wobei

$$
\begin{aligned}
Cov(X,Y) &= E((X-EX)\cdot(Y-EY)) \\
&= E(XY) - EX \cdot EY
\end{aligned}
$$

Beweis:

Mit $|X \cdot Y| \leqslant \frac{X^2+Y^2}{2}$ folgt $E(X \cdot Y) < \infty$ und $E(X+Y)^2$ existiert.

$$\begin{aligned} Var(X+Y) &= E((X+Y) - EX - EY)^2) \\ &= E((X-EX)+(Y-EY))^2 \\ &= E((X-EX)^2 + (Y-EY)^2 + 2(X-EX)(Y-EY)) \\ &= Var(X) + Var(Y) + 2 \cdot Cov(X,Y) \end{aligned}$$

Bemerkung:
Der Varianzoperator ist nicht linear sondern benötigt den Korrekturterm $Cov(X,Y)$
Dieser ist ein Maß für den linearen Zusammenhang zwischen X, Y

Korrolar 7.16 *Seien $X_1, \ldots X_n$ reellwertige Zufallsvariablen mit $EX_i^2 < \infty \quad, i = 1, \ldots, n$, so gilt nach vollständiger Induktion:*

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2 \cdot \sum_{i=1}^{n} \sum_{j=i+1}^{n} Cov(X_i, X_j)$$

Satz 7.17 (Cauchy-Schwarz-Ungleichung)
Seien X, Y Zufallsvariablen mit $EX^2, EY^2 < \infty$, dann gilt:

$$(E(X \cdot Y))^2 \leqslant EX^2 \cdot EY^2$$

Wobei Gleicheit genau dann gilt, wenn $\exists\ a \in \mathbb{R} \quad P(aX = Y) = 1$ gilt, d.h. X ein Vielfaches von Y ist.

Beweis:

$$0 \leqslant E(X+aY)^2 = EX^2 + 2a \cdot E(X \cdot Y) + a^2 \cdot EY^2 =: h(a)$$

Die Funktion h hat ein Minimum bei $a^* = -\frac{E(XY)}{EY^2}$, falls $EY^2 > 0$

$$\Rightarrow h(a^*) = EX^2 - 2\frac{E^2(XY)}{EY^2} + \frac{E^2(XY)}{E^2(Y^2)} \cdot E(Y^2) \geqslant 0$$

Falls

$$\begin{aligned} E(Y^2) = 0 \quad &\Rightarrow \quad P(Y=0) = 1 \\ &\Rightarrow \quad E(XY) = 0 \end{aligned}$$

Dabei gilt die Gleichheit bei $E(X+aY)^2 = 0$ genau dann, wenn $P(X + aY = 0) = 1$

7.5 Eigenschaften der Kovarianz

Lemma 7.18
1. *$Cov(X,Y) = E(X \cdot Y) - EX \cdot EY$*
2. *$Cov(X,X) = Var(X)$*
3. *$Cov(X,Y) = Cov(Y,X)$*

4. $Cov(aX + b, Y) = a \cdot Cov(X, Y)$

5. $Cov^2(X, Y) \leqslant Var(X) \cdot Var(Y)$

6. X, Y *stochastisch unabhängig* $\Rightarrow Cov(X, Y) = 0$
Umkehrung gilt nicht !

Beweis:

1.-4. Trivial

5. mit Cauchy-Schwarz

$$\begin{aligned} Cov(X,Y) &= E((X - EX) \cdot (Y - EY)) \\ &\leqslant (Var(X) \cdot Var(Y))^{\frac{1}{2}} \end{aligned}$$

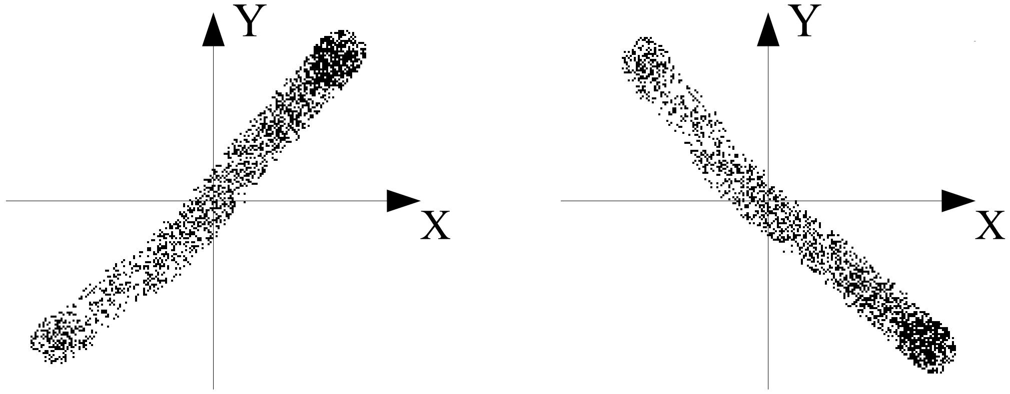
6. $Cov(X, Y) = E(XY) - EX \cdot EY = 0$
Nach dem Multiplikationssatz: $E(XY) \overset{unabh.}{=} EX \cdot EY$

Bemerkung:

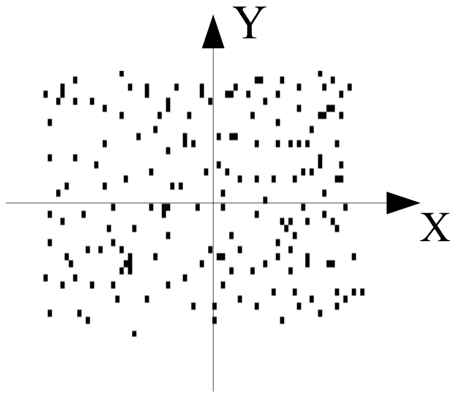
1. Die Kovarianz ist eine symmetrische Bilinearform
2. Aus der Eigenschaft 6. folgt
   X, Y stochastisch unabhängig$\Rightarrow Var(X + Y) = Var(X) + Var(Y)$
3. X, Y heißen UNKORRELIERT, wenn $Cov(X, Y) = 0$
4. Der KORRELATIONSKOEFFIZIENT ist definiert durch

$$Korr(X,Y) := \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}} \in [-1, 1]$$

Graphisch: $\Omega = \{1, \dots, n\}$ Laplaceverteilt, $X(\omega), Y(\omega)$ als Punkt in der Ebene:



$Korr(X, Y) \approx 1$ $\qquad$ $Korr(X, Y) \approx -1$



$$Korr(X,Y) \approx 0$$

Damit ist der Korrelationskoeffizient ein Maß für den linearen Zusammenhang zweier Zufallsvariablen.

Anwendung zu 2.:
Seien $X_1, \dots, X_n$ stochastisch unabhängig Zufallsvariablen mit gleicher Verteilung

$$\Rightarrow Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n} Var(X_1)$$

Bemerkung: Aus der Unkorreliertheit folgt im Allgemeinen nicht die Unabhängigkeit !

Beispiel 7.4 *Seien* $\Omega = \{1,2,3\}$ *mit* $P(\omega) = \frac{1}{3}$ $\quad, \forall \omega \in \Omega,$

X, Y Zufallsvariablen mit

	1	*2*	*3*
X	*1*	*0*	*-1*
Y	*1*	*0*	*1*

$\Rightarrow$ *Die gemeinsame Verteilung:*

$$\begin{aligned} P^{(X,Y)}(\{(1,1)\}) &= P(\{1\}) &= \frac{1}{3} \\ P^{(X,Y)}(\{(0,0)\}) &= P(\{2\}) &= \frac{1}{3} \\ P^{(X,Y)}(\{(-1,1)\}) &= P(\{3\}) &= \frac{1}{3} \end{aligned}$$

Alle weiteren sind demnach gleich Null.
Hieraus ergeben sich folgende Randverteilungen:

$$\begin{aligned} P(X=-1) &= P(X=0) = P(X=1) \\ &= \frac{1}{3} \\ P(Y=0) &= \frac{1}{3} \\ P(Y=1) &= \frac{2}{3} \end{aligned}$$

und

$$\begin{aligned}
P^{X\cdot Y}(\{-1\}) &= P^{(X,Y)}(\{(-1,1)\}) = \frac{1}{3}\\
P^{X\cdot Y}(\{0\}) &= P(\{2\}) = \frac{1}{3}\\
P^{X\cdot Y}(\{1\}) &= P(\{1\}) = \frac{1}{3}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow EX &= \frac{1}{3}(-1+0+1) = 0\\
EY &= 0\cdot\frac{1}{3} + 1\cdot\frac{2}{3} = \frac{2}{3}\\
E(X\cdot Y) &= \frac{1}{3}(-1+0+1) = 0
\end{aligned}$$

$$\Rightarrow E(X\cdot Y) = EX\cdot EY$$

d.h. X, Y sind unkorreliert aber X, Y sind nicht stochastisch unabhängig, denn

$$P(X=1, Y=1) = \frac{1}{3} \neq \frac{2}{9} = \frac{2}{3}\cdot\frac{1}{3} = P(X=1)\cdot P(Y=1)$$

8 Das schwache Gesetz großer Zahlen

Seien $X_1, \dots, X_n$ stochastisch unabhängig, gleichverteilte Zufallsvariablen
`(iid = independent identical distribution)`

$$\begin{aligned}
E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) &= \frac{1}{n}\sum_{i=1}^{n} EX_i \qquad ,\text{mit } EX_i = \mu,\ Var(X_i) = \sigma^2\\
&= \frac{n\mu}{n} = \mu\\
Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)\\
&\overset{unabh.}{=} \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{Var(X_i)}_{=\sigma^2}\\
&= \frac{n\cdot\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}$$

$\hookrightarrow$ siehe Statistik.

Ziel:
Arithmetisches Mittel von iid Zufallsvariablen mit Erwartungswert μ konvergiert gegen μ

Zum Konvergenzbegriff:
Aus der Analysis bekannt ist die punktweise Konvergenz:

$$f_n(x) := c - \frac{x}{n} \quad ,c, x \in \mathbb{R}, n \in \mathbb{N} \quad \Rightarrow \lim_{n\to\infty} f_n(x) = c \quad \forall\, x \in \mathbb{R}$$

Sei nun $(X_n)_n$ eine Folge von Zufallsvariablen mit

$$P(X_n = c) = 1 - \frac{1}{2^n} \quad , P(X_n = 2c) = \frac{1}{2^n} \qquad ,n \in \mathbb{N}, c \neq 0$$

Hier ist eine punktweise Konvergenz nicht sinnvoll, sondern

$$P(|X_n - c| > \varepsilon) = P(X_n = 2c) = \frac{1}{2^n} \stackrel{n\to\infty}{\longrightarrow} 0 \quad ,\forall\ 0 < \varepsilon < |c|$$

Bezeichnung: X_n KONVERGIERT STOCHASTISCH gegen c

8.1 Der stochastische Konvergenzbegriff

Definition 8.1

1. *Eine Folge $(X_n)_n$ von Zufallsvariablen über $(\Omega, \mathfrak{A}, P)$ heißt* STOCHASTISCH KONVERGENT *gegen 0, falls*

   $$\lim_{n\to\infty} P(|X_n| > \varepsilon) = 0 \qquad ,\forall\ \varepsilon > 0$$

   `Schreibweise:` $P\text{-}\lim\limits_{n\to\infty} X_n = 0$

2. *Eine Folge $(X_n)_n$ heißt* STOCHASTISCH KONVERGENT *gegen $c \in \mathbb{R}$ bzw. gegen Zufallsvariable X, falls*

   $$P\text{-}\lim_{n\to\infty}(X_n - c) = 0 \quad \textit{bzw.} \quad P\text{-}\lim_{n\to\infty}(X_n - X) = 0$$

   `Schreibweise:`
   $P\text{-}\lim\limits_{n\to\infty} X_n = c$ *bzw.* $P\text{-}\lim\limits_{n\to\infty} X_n = X$
   oder $X_n \stackrel{P}{\longrightarrow} c$ *bzw.* $X_n \stackrel{P}{\longrightarrow} X$

Der *Grenzwert* einer P-stochastisch konvergenten Folge ist im folgenden Sinne eindeutig definiert:

Satz 8.2

$$X_n \stackrel{P}{\longrightarrow} X \quad \textit{und} \quad X_n \stackrel{P}{\longrightarrow} Y \Rightarrow P(X = Y) = 1$$

aber nicht zwangsläufig $X = Y$!

Beweis: Sei $\varepsilon > 0$ beliebig

$$\begin{aligned}
\{\ \underbrace{|X - Y|}_{=|X - X_n + X_n - Y|}\ > \varepsilon\} \quad &\subset \quad \left\{|X - X_n| > \frac{\varepsilon}{2}\right\} \cup \left\{|X_n - Y| > \frac{\varepsilon}{2}\right\} \\
|X(\omega) - Y(\omega)| > \varepsilon \quad &\Rightarrow \quad |X - X_n| + |X_n - Y| > \varepsilon \\
&\Rightarrow \quad |X_n - X| > \frac{\varepsilon}{2} \quad \vee \quad |X_n - Y| > \frac{\varepsilon}{2} \\
\Rightarrow P(\{|X - Y| > \varepsilon\}) \quad &\leqslant \quad P\left(\left\{\underbrace{|X_n - X|}_{\stackrel{n\to\infty}{\longrightarrow} 0} > \frac{\varepsilon}{2}\right\}\right) + P\left(\left\{\underbrace{|X_n - Y|}_{\stackrel{n\to\infty}{\longrightarrow} 0} > \frac{\varepsilon}{2}\right\}\right) \\
\Rightarrow P(\{|X - Y| > \varepsilon\}) \quad &= \quad 0 \qquad ,\forall \varepsilon > 0
\end{aligned}$$

Sei $A := \{X = Y\} \quad , A^C = \bigcup\limits_{n=1}^{\infty} \{|X - Y| > \frac{1}{n}\}$

$$P(A^C) = P\left(\bigcup_{n=1}^{\infty}\left\{|X-Y| > \frac{1}{n}\right\}\right) \quad \leqslant \quad \sum_{n=1}^{\infty} P\left(\{|X-Y| > \frac{1}{n}\}\right) \overset{\text{s.o.}}{=} 0$$

$$\Rightarrow P(A^C) = 0 \quad \Rightarrow \quad P(A) = 1$$

Satz 8.3

1. $X_n \xrightarrow{P} X \quad , Y_n \xrightarrow{P} Y \quad , g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *stetig*

$$\Rightarrow g(X_n, Y_n) \xrightarrow{P} g(X, Y)$$

2. $X_n \longrightarrow X$ *(punktweise Konvergenz)*

$$\Rightarrow X(\omega) = \lim_{n\to\infty} X_n(\omega) \quad \forall \omega \in \Omega \Rightarrow X_n \xrightarrow{P} X$$

Desweiteren sind für den Nachweis der stochastischen Konvergenz Ungleichungen für Wahrscheinlichkeiten nützlich:

8.2 Markov'sche und Tschebyschoff Ungleichung

Satz 8.4 (Markov'sche Ungleichung)
Sei X Zufallsvariable, $g : \mathbb{R}^+ \to \mathbb{R}^+$ monoton wachsend, dann gilt:

$$P(|X| > \varepsilon) \quad \leqslant \quad P(|X| \geqslant \varepsilon) \quad \leqslant \quad \frac{1}{g(\varepsilon)} \cdot E(g(|X|)) \qquad , \forall \varepsilon > 0, g(\varepsilon) > 0$$

Beweis:

$$\begin{aligned} E(g(|X|)) &= \sum_{x\in\mathbb{R}} g(|x|) \cdot P^X(x) \\ &= \sum_{|x|\geqslant\varepsilon} g(|x|) \cdot P^X(x) + \sum_{|x|<\varepsilon} g(|x|) \cdot P^X(x) \\ &\geqslant \sum_{|x|\geqslant\varepsilon} g(|x|) \cdot P^X(x) \\ &\geqslant g(\varepsilon) \underbrace{\sum_{|x|\geqslant\varepsilon} P^X(x)}_{P(|X|\geqslant\varepsilon)} \end{aligned}$$

Bemerkung: Spezialfälle der Markov'schen Ungleichung

1. $g(t) := t^k \qquad , k > 0, t > 0$

$$\Rightarrow P(|X| \geqslant \varepsilon) \leqslant \frac{E|X|^k}{\varepsilon^k}$$

   Abschätzung gegen ein k-tes Moment

2. $g(t) := t^2$, Anwendung auf $Y := X - EX$

$$\Rightarrow P(|X - EX| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} Var(X)$$

$\hookrightarrow$ Die **Tschebyschoff-Ungleichung** ist grob, aber sehr wichtig.

3. $g_t(x) = e^{tx} \qquad , t > 0, x > 0$

$$\Rightarrow P(|X| \geqslant \varepsilon) \leqslant e^{-t\varepsilon} \cdot E(e^{t|X|}) \qquad (Moment-erzeugendeFunktion)$$

4. Beweis der Markov'schen Ungleichung ohne Rückgriff auf diskreten Wahrscheinlichkeitsraum

$$\begin{aligned} \text{Sei } Y(\omega) &:= \begin{cases} g(\varepsilon) & \text{, falls } |X(\omega)| \geqslant \varepsilon \\ 0 & \text{, sonst} \end{cases} \\ \Rightarrow Y &\leqslant g(|X|) \\ \Rightarrow E(g(|X|)) &\geqslant EY = g(\varepsilon) \cdot P(|X| > \varepsilon) \end{aligned}$$

8.3 Eine Version vom schwachen Gesetz großer Zahlen

Satz 8.5 *Seien $(X_i)_{i \in \mathbb{N}}$ paarweise unkorrelierte Zufallsvariablen mit $EX_i =: \mu \quad \forall i \in \mathbb{N}$ und $Var(X_i) \leqslant M < \infty$, dann gilt:*

$$P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \varepsilon \right) \leqslant \frac{M}{n \cdot \varepsilon^2} \stackrel{n \to \infty}{\longrightarrow} 0$$

$\Rightarrow P - \lim\limits_{n \to \infty} \frac{1}{n} \sum\limits_{i=1}^{n} X_i = \mu$

d.h. das arithmetische Mittel von „Einzelversuchen" (beschrieben durch $X_1, X_2, \ldots$) konvergiert stochastisch gegen den (unbekannten) Erwartungswert μ.

Beweis mit Tschebyschoff-Ungleichung:

$$\begin{aligned} E\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) &= \frac{1}{n} \sum_{i=1}^{n} \underbrace{EX_i}_{=\mu} \\ &= \frac{1}{n} \cdot n \cdot \mu = \mu \\ Var\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) &= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) \\ P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \varepsilon \right) &\leqslant \frac{\frac{1}{n^2} \sum_{i=1}^{n} Var(X_i)}{\varepsilon^2} \\ &\leqslant \frac{M}{n \cdot \varepsilon^2} \stackrel{n \to \infty}{\longrightarrow} 0 \end{aligned}$$

Beispiel 8.1 *Das arithmetische Mittel als Schätzung für den Erwartungswert*

$\hookrightarrow$ *häufiges Würfeln:* $X_i := \begin{cases} 1 & , 6 \text{ fällt} \\ 0 & , \text{ sonst} \end{cases}$

Aussage:
Die relative Häufigkeit für das Würfeln einer 6 *konvergiert stochastisch gegen* $EX = \frac{1}{6}$

Beispiel 8.2 *Sei* $(\Omega, \mathfrak{P}(\Omega)\ , P)$ *ein diskreter Wahrscheinlichkeitsraum,*
$A \subseteq \Omega, P(A) = p\ , 0 < p < 1$ *mit n-facher Wiederholung eines Zufallsexperiments*
$\hookrightarrow$ *Produktraum*
$A_i \hat{=} A$ *tritt im i-ten Versuch ein*
Seien $X_i = \mathcal{I}_{A_i} \quad , 1 \leqslant i \leqslant n,$ *stochastisch unabhängig*
$\Rightarrow X_i$ *unkorreliert und identisch verteilt.*
$\Rightarrow EX_i = P(A_i) = p, \quad Var(X_i) = p \cdot (1-p)$ *Binomialverteilung*

$$\Rightarrow P - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = p$$

d.h. falls p unbekannt ist, dann ist das arithmetische Mittel $\frac{1}{n} \sum\limits_{i=1}^{n} X_i$ *ein guter „Schätzer“ für diesen Parameter.*
($\frac{1}{n} \sum\limits_{i=1}^{n} X_i$ *ist die relative Häufigkeit von A in n Versuchen)*

Beispiel 8.3 *eventuell gefälschte (nicht-Laplace-) Münze.*
$\hookrightarrow$ *Qualitätsprüfung (Gut/Schlecht-Prüfung)*
Stichprobe vom Umfang n

Wie oft muß die Münze geworfen werde (wie viele Teile müßen untersucht werden), damit $p =$ *„Wahrscheinlichkeit für Zahl“ mit einer Sicherheitswahrscheinlichkeit* $\geqslant 0,95$ *auf* $0,01$ *genau berechnet werden kann ?*

Mit Tschebyschoff-Ungleichung:

$$\begin{aligned}
X_i &= \mathcal{I}_{\textit{Zahl im i-ten Versuch}} \\
EX_i &= p \qquad \textit{(unbekannt)} \\
\Rightarrow (X_i)_{i\in\mathbb{N}} \textit{ iid. Zufallsvariable} &\sim b(1,p) \\
\overset{s.o.}{\Rightarrow} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \geqslant 0,01\right) &\leqslant \frac{p\cdot(1-p)}{n\cdot(0,01)^2} = \frac{Var(b(1,p))}{n\cdot\varepsilon^2} \\
(\textit{Mit der Abschätzung } x\cdot(1-x) &\leqslant \frac{1}{4} \qquad , x \in (0,1)) \\
\leqslant \frac{\frac{1}{4}}{n\cdot(0,01)^2} &\leqslant 0,05 = 1 - 0,95 \\
\Rightarrow n &\geqslant \frac{1}{4\cdot(0,01)^2\cdot 0,05} \\
&= 50000
\end{aligned}$$

(die grobe Abschätzung führt zu großem n; es existieren bessere Abschätzungen)
$\hookrightarrow$ *Versuchsplanung*

9 Borelmengen und Maße

Bisher wurde das diskrete (atomare) Wahrscheinlichkeitsmaß P benutzt mit
$P(A) = \sum\limits_{x \in A} p(\{x\})$

Zur Wiederholung:

Seien $\Omega \neq \emptyset$ mit $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$.
$\mathfrak{A}$ heißt σ-Algebra über Ω, falls

1. $\Omega \in \mathfrak{A}$
2. $A \in \mathfrak{A} \Rightarrow \ A^C \in \mathfrak{A} \qquad , \forall A \in \mathfrak{A}$
3. $(A_n)_n \subset \mathfrak{A} \Rightarrow \ \bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$

$\hookrightarrow$ Erzeugung von σ-Algebren über gegebenen Mengen

Lemma 9.1 *Ist $\mathcal{E} \subset \mathfrak{P}(\Omega)$, so existiert eine kleinste σ-Algebra $\mathfrak{A}(\mathcal{E})$, die $\mathcal{E}$ enthält, d.h.*

1. *$\mathfrak{A}(\mathcal{E})$ ist eine σ-Algebra*
2. *$\mathcal{E} \subset \mathfrak{A}(\mathcal{E})$*
3. *Ist $\mathfrak{A}'$ eine σ-Algebra mit $\mathcal{E} \subset \mathfrak{A}' \Rightarrow \mathfrak{A}(\mathcal{E}) \subseteq \mathfrak{A}'$*

Bemerkung:

1. $\mathfrak{A}(\mathcal{E})$ heißt die von $\mathcal{E}$ erzeugte σ-Algebra.
   $\mathcal{E}$ heißt Erzeuger
2. Ist $\mathcal{E}$ σ-Algebra $\Rightarrow \mathfrak{A}(\mathcal{E}) = \mathcal{E}$
3. $\mathcal{E} = \{A\} \Rightarrow \mathfrak{A}(\mathcal{E}) = \{\emptyset, A, A^C, \Omega\}$
4. Ist $\mathcal{E} = \{A_1, \ldots, A_n\}$ mit $\Omega = \sum\limits_{i=1}^{n} A_i$ (disjunkte Zerlegung)
   $\Rightarrow |\mathfrak{A}(\mathcal{E})| = 2^n$

Bezeichnung 9.2 *Seien $k \in \mathbb{N}$ und $\mathcal{E}^n := \{(a, b] |\ a, b \in \mathbb{R}^n\}$*
(für $n = 1: \ F^X(b) - F^X(a) = P(X \in (a, b])$)

so heißt $\mathfrak{B}^n := \mathfrak{A}(\mathcal{E}^n)$ ($\mathfrak{B}^1 = \mathfrak{B}$) Borel'sche σ-Algebra *oder*
σ-Algebra der Borelmenge *über $\mathbb{R}^n$.*
Jedes $B \in \mathfrak{B}^n$ heißt Borelmenge.

Das Ziel ist die Zuordnung einer Maßzahl (Länge, Fläche, Volumen) für **jede** Menge aus $\mathfrak{B}^n$.

Definition 9.3 *Sei* $(\Omega, \mathfrak{A})$ *ein messbarer Raum.*
Eine Abbildung $\mu : \mathfrak{A} \to [0, \infty]$ *heißt* MASS *über* $(\Omega, \mathfrak{A})$*, falls gilt:*

1. $\mu(\emptyset) = 0$
2. *Für alle Familien* $(A_i)_{i \in \mathbb{I}}$ *von paarweise disjunkten* $A_i \in \mathfrak{A}$ *mit abzählbarer Indexmenge* $\mathbb{I}$ *gilt:*

$$\mu\left(\sum_{i \in \mathbb{I}} A_i\right) = \sum_{i \in \mathbb{I}} \mu(A_i)$$

$(\Omega, \mathfrak{A}, \mu)$ *heißt* MASSRAUM.

Ein Maß μ *mit* $\mu(\Omega) = 1$ *heißt* WAHRSCHEINLICHKEITSMASS P *(vgl. erstes Kapitel) und* $(\Omega, \mathfrak{A}, P)$ *heißt* WAHRSCHEINLICHKEITSRAUM.

Satz 9.4 *Seien* $(\Omega, \mathfrak{A}, P)$ *ein Maßraum und* $A, B, A_1, A_2, \ldots \in \mathfrak{A}$
Das Maß μ *besitzt folgende Eigenschaften:*

1. Nulltreue: $\mu(\emptyset) = 0$

2. Positivität: $\mu(A) \geqslant 0 \quad \forall A \in \mathfrak{A} \qquad$ ($\mu(A) = \infty$ *ist möglich !*)

3. Additivität: *Ist* $A \cap B = \emptyset$*, so gilt:* $\mu(A \cup B) = \mu(A) + \mu(B)$

4. Additivität: *Aus 3 folgt für paarweise disjunkte Mengen* $A_1, \ldots, A_n$*:*
$\mu\left(\sum\limits_{i=1}^{n} A_i\right) = \sum\limits_{i=1}^{n} \mu(A_i)$

5. Isotonie: *Ist* $A \subset B$*, so gilt:* $\mu(A) \leqslant \mu(B)$

6. Subtraktivität: *Sind* $A \subset B$ *und* $\mu(\Omega) < \infty$*, so gilt:* $\mu(B \backslash A) = \mu(B) - \mu(A)$

7. Komplementarität: *Ist* $\mu(\Omega) < \infty$*, so gilt:* $\mu(A^C) = \mu(\Omega) - \mu(A)$

8. Stetigkeit von unten: *Ist die Folge* $(A_n)_{n \in \mathbb{N}}$ *isoton (monoton wachsend), so gilt:*
$\mu\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n \to \infty} \mu(A_n)$

9. Stetigkeit von oben: *Ist die Folge* $(A_n)_{n \in \mathbb{N}}$ *antiton (monoton fallend), so gilt:*
$\mu\left(\bigcap\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n \to \infty} \mu(A_n)$

10. Sub-Additivität: *Für Ereignisse* $A_1, \ldots, A_n$ *gilt:*
$\mu\left(\bigcup\limits_{i=1}^{n} A_i\right) \leqslant \sum\limits_{i=1}^{n} \mu(A_i)$

11. Sub-σ-Additivität: *Für eine Ereignisfolge* $(A_i)_{i \in \mathbb{N}}$ *gilt:*
$\mu\left(\bigcup\limits_{i=1}^{\infty} A_i\right) \leqslant \sum\limits_{i=1}^{\infty} \mu(A_i)$

Bemerkung: $\mathfrak{B}^n \neq \mathfrak{P}(\Omega)$, $\mathfrak{B}^n$ enthält alle „vernünftigen“ Mengen von $\mathbb{R}^n$

Lemma 9.5 $\mathfrak{B}^n$ *enthält alle offenen und abgeschlossenen Teilmengen des* $\mathbb{R}^n$

9.1 Stetige Verteilungsfunktionen

Bemerkung:

- Ein Messraum $(\Omega, \mathfrak{A})$ kann jetzt mit einem Wahrscheinlichkeitsmaß P versehen werden $\hookrightarrow$ 1.7 Kolmogorov-Axiome

  Sei $\mathfrak{A}$ σ-Algebra über $\Omega \neq \emptyset$.
  Eine Abbildung $P : \mathfrak{A} \to [0,1]$ mit

  1. $P(A) \geqslant 0$
  2. $P(\Omega) = 1$
  3. $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ für alle disjunkten Mengen $A_n \in \mathfrak{A}$

  heißt Wahrscheinlichkeitsverteilung oder Wahrscheinlichkeitmass.
  $(\Omega, \mathfrak{A}, P)$ heißt Wahrscheinlichkeitsraum

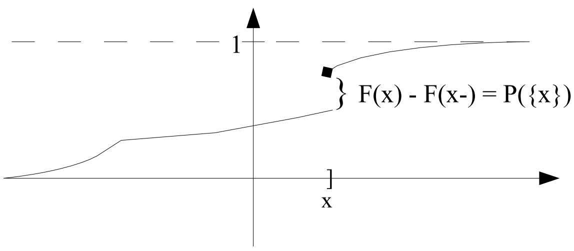
- Alle Eigenschaften von P gelten völlig allgemein ohne Rückgriff auf diskrete Struktur.

- Zuordnung der Verteilungsfunktion im Fall $(\mathbb{R}^1, \mathfrak{B}^1, P)$

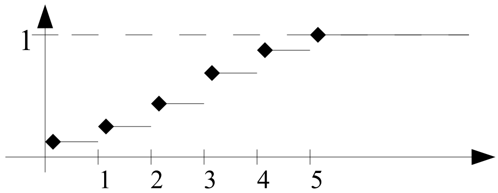
$$F_P : \mathbb{R} \to [0,1] \text{ mit} F_P(x) = P((-\infty, x]) \qquad , x \in \mathbb{R} (\text{siehe def. } 6.16)$$

  Bemerkung:
  $\{(-\infty, x] |\ x \in \mathbb{R}\}$ ist auch ein Erzeuger von $\mathfrak{B}$
  Eigenschaften (siehe Lemma 6.16)

  - F_P ist monoton wachsend und rechtsseitig stetig
  - $F_P(y) \to 0, y \to -\infty$
  - $F_P(y) \to 1, y \to \infty$



  Im Gegensatz dazu die diskrete Verteilung (Poisson):



- Klar ist: $P(\{x\}) = 0 \iff F_P$ stetig in x
  P heißt stetige Wahrscheinlichkeitsverteilung, falls $P(\{x\}) = 0$ $, \forall x \in \mathbb{R}$

Satz 9.6 (über die Existenz und Eindeutigkeit für Maße auf $(\mathbb{R}, \mathfrak{B})$)
Sei $G : \mathbb{R} \to \mathbb{R}$ eine monoton wachsende, rechtsseitig stetige Funktion
Dann existiert genau ein Maß μ auf $(\mathbb{R}, \mathfrak{B})$ mit

$$\mu((a,b]) = G(b) - G(a) \qquad , \forall a, b \in \mathbb{R}, a < b$$

Bezeichnung 9.7

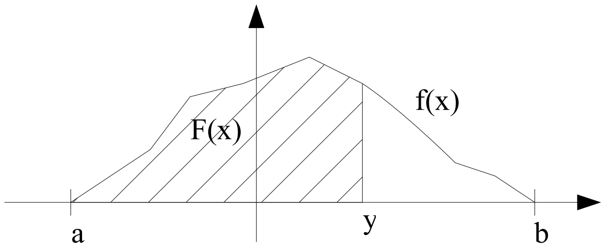
1. *Das zu $G(x) = x$ gehörige Maß heißt* LEBESGUE-MASS *(Le'bek)*
   $\hookrightarrow$ *natürliche Verallgemeinerung des Längenbegriffs*
2. *Anwendung auf die Verteilungsfunktion: $G = F$*
   Ist $F : \mathbb{R} \to [0,1]$ eine Abbildung mit den Eigenschaften aus 6.16(s.o.), dann existiert genau ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B})$, das F als Verteilungsfunktion besitzt.

Jetzt: Die spezielle Gestalt von Wahrscheinlichkeitmaßen mit „einfacher“ Handhabung.
(Teilweise läßt sich mit der Verteilungsfunktion einfacher rechnen)

Lemma 9.8 *Sei $f : \mathbb{R} \to \mathbb{R}^+$ und $I = (a,b) \subset \mathbb{R}, -\infty \leqslant a < b \leqslant \infty$,*
mit $f(x) = 0\ \forall x \in I^C$, f stetig auf I und

$$\int_{-\infty}^{\infty} f(x)dx = \int_a^b f(x)dx = 1 \qquad \textit{(Riemann-Integral)}$$

Definiere $F : \mathbb{R} \to [0,1]$ durch $F(y) = \int_{-\infty}^{y} f(x)dx$



dann gilt: F ist Verteilungsfunktion

- $F(y) \in [0,1] \quad \forall y \in \mathbb{R}$, *$F$ ist monoton steigend und stetig*
- $\lim\limits_{y \to -\infty} F(y) = \lim\limits_{y \to -\infty} \int_{-\infty}^{y} f(x)dx = 0$
- $\lim\limits_{y \to +\infty} F(y) = \lim\limits_{y \to +\infty} \int_{-\infty}^{y} f(x)dx = 1$

Nach dem Existenz- und Eindeutigkeitssatz existiert genau ein Wahrscheinlichkeitsmaß P.
Diese Funktion f heißt (RIEMANN-)DICHTE(-FUNKTION) *von P.*

Eigenschaften von P:

- $P(\{x\}) = 0 \qquad \forall x \in \mathbb{R}$
- $P(A) = 0 \qquad$ für jede abzählbar große Menge A

- $P((a,b]) = P([a,b]) = P([a,b)) = P((a,b)) = \int_a^b f(x)dx \quad \forall a < b$

Hierbei sind die Ereignisse in Intervallform.

Allgemeiner: Was ist $P(B)$ für irgendein $B \in \mathfrak{B}$
$\hookrightarrow$ neuer Integralbegriff notwendig!
$\hookrightarrow$ siehe höhere Stochastik

Bemerkung: Umgekehrte Herleitung

P liegt vor mit stetiger Verteilungsfunktion F,
so dass F auf $\{y|\ 0 < F(y) < 1\} =: (a,b)$ stetig differenzierbar, so wird durch

$$f(x) := \begin{cases} 0 & , x \leqslant a \\ F'(x) & , a < x < b \\ 0 & , x \geqslant b \end{cases}$$

die zu P gehörige Dichte erklärt

Bemerkung: Wahrscheinlichkeitsmaße auf Teilmengen der reellen Zahlen

Seien $\omega \subset \mathbb{R}, \omega \in \mathfrak{B}$
Betrachte die Spur-σ-Algebra

$$\omega \cap \mathfrak{B} := \{\omega \cap b | b \in \mathfrak{B}\} \subseteq \mathfrak{B}$$

Ist Ω abzählbar, so ist $\Omega \cap \mathfrak{B} = \mathfrak{P}(\Omega)$,
denn für jedes $A \in \Omega$ ist $A = \bigcup\limits_{\omega \in A} \Omega \cap \{\omega\} \in \Omega \cap \mathfrak{B}$ eine abzählbare Vereinigung

`Schreibweise`: $\mathfrak{B}|_\Omega$ ($\mathfrak{B}$ eingeschränkt auf Ω)

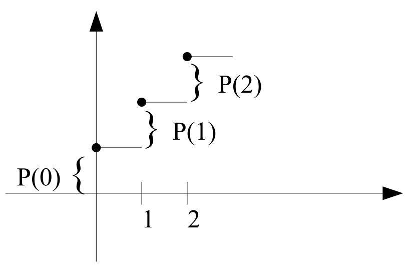
Weiterhin:

Ist P Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{B}|_\Omega)$, so kann P fortgesetzt werden zu

$$\widetilde{P} : \mathfrak{B} \to [0,1] \quad \text{mit} \quad \widetilde{P}(B) = P(B \cap \Omega)$$

D.h. jedes Maß P auf $\Omega \subset \mathbb{R}$ kann aufgefasst werden als ein Maß über den gesamten reellen Zahlen. Ordne dann P die Verteilungsfunktion von $\widetilde{P}$ zu.

z.B. Verteilungsfunktion der Poissonverteilung



$$
\begin{aligned}
F(t) &= \widetilde{P}((-\infty, t]) \\
&= P((-\infty, t] \cap \mathbb{N}_0) \\
&= \sum_{\omega \in (-\infty, t] \cap \mathbb{N}_0} p(\omega) \\
&= \sum_{t \geqslant \omega \in \mathbb{N}_0} e^{-\lambda} \cdot \frac{\lambda^{\omega}}{\omega!}
\end{aligned}
$$

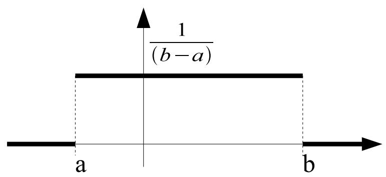
10 Wahrscheinlichkeitsmaße mit Riemann-Dichten über $\mathbb{R}$

Zunächst Beispiele für Riemann-Dichten (nach 9.8)

10.1 Rechteckverteilung (stetige Gleichverteilung)

auf (a, b), $\quad a, b \in \mathbb{R}, a < b$

$$
f(x) = \frac{1}{b-a} \cdot \mathcal{I}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & , x \in [a, b] \\ 0 & , \text{sonst} \end{cases}
$$



Die durch Dichtefunktion definierte Wahrscheinlichkeitsverteilung über (a, b) heißt STETIGE GLEICHVERTEILUNG
Verteilungsfunktion:

$$
F(x) = \begin{cases} 0 & , x < a \\ \frac{x-a}{b-a} & , a \leqslant x \leqslant b \\ 1 & , x > b \end{cases}
$$

Zum Berechnen der Wahrscheinlichkeiten

$$
\begin{aligned}
P(X \in (c, d)) &= \int_c^d f(x) dx \qquad (c, d) \subset (a, b) \\
&= P(X \in (-\infty, d) \setminus (-\infty, c)) \\
&= P(X \in (-\infty, d)) - P(X \in (-\infty, c)) \\
&= P(X \in (-\infty, d]) - P(X \in (-\infty, c]) \\
&= F(d) - F(c)
\end{aligned}
$$

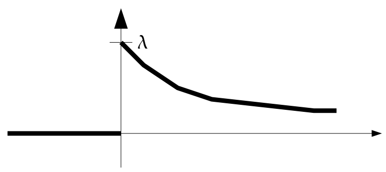
Damit ergibt sich der Zusammenhang: $F'(t) = f(t) \quad \forall\, t \in \mathbb{R}$
und gegenüber dem Lebesgue-Maß: $P(B) = \frac{1}{b-a} \cdot \mu([a,b] \cap B) \quad B \in \mathfrak{B}^1$

Die Wahrscheinlichkeit eines Intervalls häng nur von dessen Länge ab, nicht von der Lage

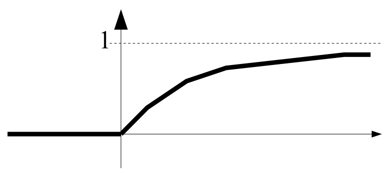
`Schreibweise`: $R(a,b)$ bzw $R[a,b]$
heißt RECHTECKVERTEILUNG mit Parametern a und b

10.2 Exponentialverteilung

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & , x > 0 \\ 0 & , x \leqslant 0 \end{cases} \qquad , x \in \mathbb{R}, \lambda > 0$$



$$F(X) = \begin{cases} 0 & , x \leqslant 0 \\ 1 - e^{-\lambda x} & , x > 0 \end{cases}$$



Vielfache Anwendung in der Praxis beim Modellieren von Wartezeiten, Lebensdauern, etc. aufgrund vieler nützlicher und einfacher Eigenschaften
$\hookrightarrow$ Gedächtnislosigkeit, konstante Abfallrate, etc.

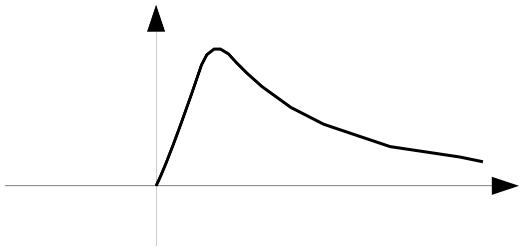
`Schreibweise:` $Exp(\lambda)$: EXPONENTIALVERTEILUNG mit Parameter λ

10.3 Weibull-Verteilung

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} & , x > 0 \\ 0 & , x \leqslant 0 \end{cases} \qquad , x \in \mathbb{R}, \quad \alpha, \beta > 0$$

(für $\alpha = \lambda, \beta = 1$ ergibt sich $Exp(\lambda)$)

Möglicher Graph:



Anwendung z.B. bei Lebensdauern.

$$F(X) = \begin{cases} 0 & , x \leqslant 0 \\ 1 - e^{-\alpha x^{\beta}} & , x > 0 \end{cases}$$

`Schreibweise:` $Wei(\alpha, \beta)$ WEIBULLVERTEILUNG mit Parametern α, β

10.4 Gammaverteilung

$$f(x) = \frac{b^p}{\Gamma(p)} \cdot e^{-bx} \cdot x^{p-1} \cdot \mathcal{I}_{(0,\infty)}(x) \qquad , x \in \mathbb{R}, \quad b, p > 0$$

(Für $b = \lambda, p = 1$ ergibt sich $Exp(\lambda)$)
$\Rightarrow$ Probleme bei der Angabe von F, da dies nur für $p \in \mathbb{N}$ möglich ist.

Anwendung: Zuverlässigkeitstheorie ($\hookrightarrow$ Reliabilitytheorie)

10.5 Gauß'sche Normalverteilung

Die wohl wichtigste Verteilung der mathematischen Statistik

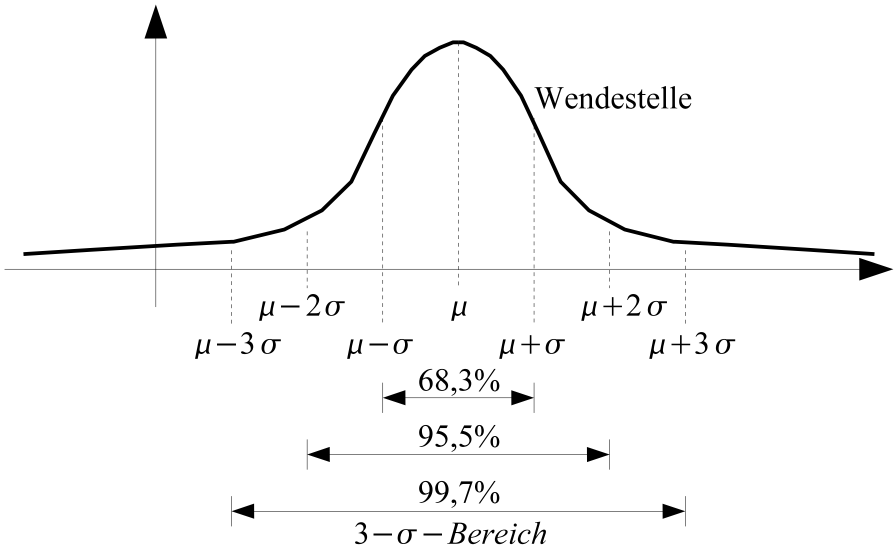
$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma^2} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma}\right) \qquad , \mu, x \in \mathbb{R}, \ \sigma > 0$$

`Schreibweise:` $\mathcal{N}(\mu, \sigma^2)$: NORMALVERTEILUNG mit Parametern μ und σ^2
(In älterer Literatur auch noch $\mathcal{N}(\mu, \sigma)$!)

Standardisierung:
Definiere eine neue Zufallsvariable $Y := \frac{x-\mu}{\sigma}$ mit $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned} \Rightarrow P(Y \leqslant y) &= P(\frac{x-\mu}{\sigma} \leqslant y) \\ &= P(X \leqslant y\sigma + \mu) \\ &= \int_{-\infty}^{y\sigma+\mu} f(x)dx \qquad (x = z\sigma + \mu \Longleftrightarrow z = \frac{x-\mu}{\sigma}) \\ &= \int_{-\infty}^{y} \sigma \cdot f(z\sigma + \mu)dz \\ &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{z^2}{2}} dz \end{aligned}$$

$\Rightarrow$ Dichte der Normalverteilung mit $\mu = 0, \sigma^2 = 1$, d.h. $Y \sim \mathcal{N}(0,1)$
heißt STANDARD-NORMALVERTEILUNG
($\varphi(x) \widehat{=} f(x)$ bei Normalverteilung)



Φ ist die Verteilungsfunktion φ' und ist nicht geschlossen darstellbar, liegt aber tabelliert vor.

Anwendungen:

- „Fehlergesetze" in der Physik
- asymptotische Verteilung einer Größe, die sich additiv aus vielen anderen zusammensetzt
  $\hookrightarrow$ zentraler Grenzwertsatz

Seien $(X_i)_{i \in \mathbb{N}}$ stochastisch unabhängig.

$$\underbrace{\frac{\frac{1}{n}\sum X_i - E(\frac{1}{n}\sum X_i)}{\sqrt{Var(\frac{1}{n}\sum X_i)}}}_{\text{Standardisiert}} \overset{n \to \infty}{\approx} \mathcal{N}(0,1)$$

Insbesonderes für $(X_i)_{i \in \mathbb{N}} iid,\ EX_i = \mu,\ Var(X_i) = \sigma^2$

$$\frac{\frac{1}{n}\sum X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{n \to \infty}{\approx} \mathcal{N}(0,1)$$

Für eine Zufallsvariable im $\mathbb{R}^n$ war die Verteilungsfunktion definiert in 6.18

$$\begin{aligned} F^X(x) &:= P^X((-\infty, x_1] \times \ldots \times (-\infty, x_n]) \\ &= P(X_1 \in (-\infty, x_1], \ldots, X_n \in (-\infty, x_n]) \end{aligned}$$

Die zugehörige Verteilungsfunktion der i-ten Randverteilung (Marginalverteilung) definiert in 6.10

$$\begin{aligned} F^{X_i}(x) &= P(\underbrace{\mathbb{R}\times\ldots\times\mathbb{R}}_{i-1}\times(-\infty,x]\times\mathbb{R}\times\ldots\times\mathbb{R}) \\ &= F^X(\underbrace{\infty,\ldots,\infty}_{i-1},x,\infty,\ldots,\infty) \end{aligned}$$

Definition 10.6 *Eine Funktion $f:\mathbb{R}^n\to\mathbb{R}$ heißt* (RIEMANN-) DICHTE *auf $\mathbb{R}^n$, falls gilt:*

- $f(x)\geqslant 0, \qquad \forall x\in\mathbb{R}^n$
- *f ist Riemann-integrierbar mit*
- $\int\limits_{-\infty}^{+\infty}\ldots\int\limits_{-\infty}^{+\infty} f(x_1,\ldots,x_n)dx_1\ldots dx_n = 1$

Satz 10.7 *Ist f eine Dichte über $\mathbb{R}^n$, so definiert*

$$F(x_1,\ldots,x_n) = \int\limits_{-\infty}^{x_1}\ldots\int\limits_{-\infty}^{x_n} f(y_1,\ldots,y_n)dy_1\ldots dy_n \qquad (x_1,\ldots,x_n)\in\mathbb{R}^n$$

eine stetige Verteilungsfunktion über $\mathbb{R}^n$.

Ist P das zugehörige Wahrscheinlichkeitsmaß über $(\mathbb{R}^n,\mathfrak{B}^n)$ heißt f eine Dichte von P.

Dann ist:

$$P(\bigtimes_{i=1}^{n}[a_i,b_i]) = \int\limits_{a_1}^{b_1}\ldots\int\limits_{a_n}^{b_n} f(y_1,\ldots,y_n)dy_1\ldots dy_n \qquad \forall a_i\leqslant b_i\in\mathbb{R}$$

Zurück zu Maßen mit Dichten im $\mathbb{R}^1$:
Bemerkung:

Die in den Kapiteln bisher eingeführten Begriffe waren allgemein gehalten, d.h. sie sind auf Maße mit Dichten anwendbar.

Sei $X:(\Omega,\mathfrak{A})\to(\Omega',\mathfrak{A}')$ eine Zufallsvariable.
Messbarkeit: $X^{-1}(A')\in\mathfrak{A} \quad \forall A'\in\mathfrak{A}'$ (siehe 6.4)

Speziell für $(\Omega',\mathfrak{A}')=(\mathbb{R},\mathfrak{B})$

$P^X(A') := P(X^{-1}(A'))$
ist Wahrscheinlichkeitsverteilung von X unter P (siehe 6.6)

$F^X(x) := P^X((-\infty,x])$
ist Verteilungsfunktion von X (siehe 6.16))

P^X ist durch F^X eindeutig bestimmt (siehe 6.17)

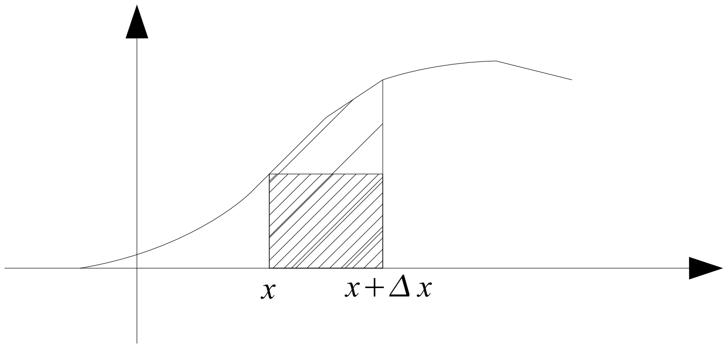
Bemerkung:

Sei $(\Omega, \mathfrak{A})$ ein messbarer Raum und $X : \Omega \to \mathbb{R}^k$ eine Abbildung mit
$X^{-1}((-\infty, x]) = \{\omega | \ X(\omega) \leqslant x\} \in \mathfrak{A} \qquad \forall x \in \mathbb{R}^k$ (Komponentenweise kleiner),
dann gilt:

$$X^{-1}(B) \in \mathfrak{A} \qquad \forall B \in \mathfrak{B}^k$$

d.h. Die Messbarkeit muß nur auf dem Erzeuger von $\mathfrak{B}^k$ nachgerechnet werden.

Nun zum Erwartungswert von Riemann-Dichten:



$$\begin{aligned} P(X \in (x, \Delta x)) &= \int_x^{x+\Delta x} f(y)dy \\ &\approx \Delta x \cdot f(x) \end{aligned}$$

d.h. der Beitrag zu einem durchschnittlichen Wert der Fläche: $x \cdot f(x)\Delta x$
erscheint sinnvoll: $\int\limits_{-\infty}^{+\infty} x f(x)dx$

Definition 10.8 *Sei f die Riemann-Dichte von X, so heißt*

$$EX := \int\limits_{-\infty}^{+\infty} x \cdot f(x)dx$$

Erwartungswert *von X (falls wohldefiniert) (vgl 7.1)*

Bemerkung:
Es gelten weiterhin nach dem Ersetzen von Summen durch Integrale:

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

  z.B. $g(x) := (X - EX)^2 \Rightarrow E(g \circ X) = Var(X)$

  $Var(X) = \int\limits_{-\infty}^{+\infty} (X - EX)^2 f(x)dx$

- Eigenschaften der Erwartungswerte (siehe 7.6)
- Eigenschaften der Varianz (siehe 7.13)

- Eigenschaften der Kovarianz (siehe 7.18)
- Ungleichung von Jensen (siehe 7.10)
- Ungleichung von Ljupunoff (siehe 7.12)
- Ungleichung von Markov und Tschebyschoff (siehe 8.4)

Beispiel 10.1 *1. Sei* $X \sim R(a,b)$

$$\begin{aligned}
EX &= \int_{-\infty}^{+\infty} x \cdot f(x)dx \\
&= \int_a^b x \cdot \frac{1}{b-a}dx \\
&= \frac{b^2-a^2}{2} \cdot \frac{1}{b-a} \\
&= \frac{b+a}{2}
\end{aligned}$$

$$\begin{aligned}
EX^2 &= \int_a^b x^2 \frac{1}{b-a}dx \\
&= \frac{1}{b-a} \cdot \frac{b^3-a^3}{3} \\
&= \frac{1}{3}(b^2+ab+a^2)
\end{aligned}$$

$$\begin{aligned}
Var(X) &= EX^2 - E^2X \\
&= \frac{1}{3}(b^2+ab+a^2) - \frac{(b+a)^2}{4} \\
&= \frac{(a-b)^2}{12} \qquad \textit{unabhängig von der Lage!}
\end{aligned}$$

2. $X \sim Exp(\lambda) \Rightarrow EX = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$

3. $X \sim Wei(\alpha,\beta) \Rightarrow EX = \alpha^{-\frac{1}{\beta}}\Gamma(\frac{1}{\beta}+1)$
$Var(X) = \alpha^{-\frac{2}{\beta}}\left(\Gamma(\frac{2}{\beta}+1) - \Gamma^2(\frac{1}{\beta}+1)\right)$

4. $X \sim \Gamma(b,p) \Rightarrow EX = \frac{p}{b}, Var(X) = \frac{p}{b^2}$

5. $X \sim \mathcal{N}(\mu,\sigma^2)$

$$\begin{aligned}
EX &= \int_{-\infty}^{+\infty} x \cdot f(x)dx \\
&= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} \frac{x}{\sigma} \cdot \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx \qquad (y := \frac{x-\mu}{\sigma})
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} (y\sigma + \mu) \cdot e^{-\frac{y^2}{2}} dy \\
&= \frac{\sigma}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} y \cdot e^{-\frac{y^2}{2}} dy + \mu \cdot \int_{-\infty}^{+\infty} \underbrace{\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}}_{\textit{Dichte von } \mathcal{N}(0,1)} dy \\
&= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\left[-e^{-\frac{y^2}{2}}\right]_{-\infty}^{+\infty}}_{=0} + \mu \cdot 1 \\
&= \mu
\end{aligned}$$

$$Var(X) = \ldots = \sigma^2$$

Beispiel 10.2 (Charakterisierungsresultat)
Typische Situation an einer (besetzten) Telefonzelle:
$\Rightarrow$ *die restliche Wartezeit ist „unabhängig“ von der bisherigen Wartezeit.*

Modellierung:
Zu $X \widehat{=}$ *„Wartezeit“,* $X \geqslant 0$

$$\hookrightarrow P(X > t + s \mid X > s) = P(X > t) \qquad \forall s, t \geqslant 0$$

Bezeichnung: „Gedächtnislosigkeit“ von P *(Alterungsfreiheit in anderen Modellen)*

$$\begin{aligned}
\Rightarrow \quad & \frac{P(X > t + s \wedge X > s)}{P(X > s)} = P(X > t) \\
\Longleftrightarrow \quad & \underbrace{P(X > t + s)}_{1-F(t+s)} = \underbrace{P(X > t)}_{1-F(t)} \cdot \underbrace{P(X > s)}_{1-F(s)} \qquad (\bar{F}(x) = 1 - F(x)) \\
\Longleftrightarrow \quad & \bar{F}(t + s) = \bar{F}(t) \cdot \bar{F}(s) \qquad \star
\end{aligned}$$

$\hookrightarrow$ *Funktionalgleichung für* $\bar{F}$
Voraussetzung: $\lambda := -\log P(X > 1) = -\log \bar{F}(1) \overset{!}{>} 0$
$\Rightarrow \bar{F}(1) = e^{-\lambda}$ *Seien* $m, n \in \mathbb{N}$*, dann folgt*

$$\begin{aligned}
\bar{F}(1) &= \bar{F}(\underbrace{\frac{1}{n} + \ldots + \frac{1}{n}}_{n\textit{-mal}}) \\
&\overset{\star}{=} \left(\bar{F}(\frac{1}{n})\right)^n \\
\Rightarrow \bar{F}\left(\frac{1}{n}\right) &= (\underbrace{\bar{F}(1)}_{e^{-\lambda}})^{\frac{1}{n}} \\
&= e^{-\frac{\lambda}{n}}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \bar{F}\left(\frac{m}{n}\right) &= \bar{F}(\underbrace{\frac{1}{n}+\ldots+\frac{1}{n}}_{m\text{-}mal}) \\
&\stackrel{\star}{=} \left(\bar{F}\left(\frac{1}{n}\right)\right)^m \\
&= \left(e^{-\frac{\lambda}{n}}\right)^m \\
&= e^{-\lambda\frac{m}{n}} \\
\Rightarrow F\left(\frac{m}{n}\right) &= 1-e^{-\lambda\frac{m}{n}}
\end{aligned}$$

Also ist $\bar{F}(q) = e^{-\lambda q} \qquad \forall q \in \mathbb{Q}^+$
Aus der Monotonie von $\bar{F}$ *folgt* $\forall x > 0:$

$$\underbrace{\inf_{q<x} \bar{F}(q)}_{\to e^{-\lambda q}} \quad \geqslant \quad \bar{F}(x) \quad \geqslant \quad \underbrace{\sup_{q>x} \bar{F}(q)}_{\to e^{-\lambda q}}$$

$\Rightarrow F(x) = 1 - e^{\lambda x} \qquad \forall x > 0,$ *d.h.* $X \sim Exp(\lambda)$

Anwendung von Dichten z.B. bei der Modellierung in der $\hookrightarrow$ Zuverläßigkeitstheorie
$\Rightarrow$ Beschreibung des Ausfallverhaltens einer Anlage

Bemerkung:

Betrachte für $x, \varepsilon > 0$

$$\begin{aligned}
P(X \leqslant x+\varepsilon)|\ X > x) &\,\widehat{=}\, \text{„Restlebensdauer kleiner } \varepsilon \text{ unter der Bedingung, dass die Lebensdauer} > x \\
&= \frac{P(x < X \leqslant x+\varepsilon)}{P(X > x)}
\end{aligned}$$

Voraussetzung: $\exists \lim\limits_{\varepsilon > 0+} P(X \leqslant x+\varepsilon|\ X > x)$
Dann gilt:

$$\frac{1}{\varepsilon}\frac{P(x < X \leqslant x+\varepsilon)}{P(X > x)} = \underbrace{\frac{F(x+\varepsilon)-F(x)}{\varepsilon}}_{\to F'(x)=f(x)} \cdot \frac{1}{1-F(x)}$$

d.h.

$$P(X \leqslant x+\varepsilon|\ X > x) \approx \varepsilon \frac{f(x)}{1-F(x)}$$

Bezeichnung 10.9 $h = \dfrac{f}{1-F}$ *heißt* Ausfallrate

Beispiel 10.3

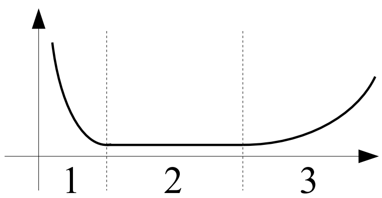
1. $X \sim Exp(\lambda) \Rightarrow h(x) = \dfrac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda = konst.$

2. $X \sim Wei(\alpha, \beta)$

$$\Rightarrow h(x) = \frac{\alpha\beta x^{\beta-1} e^{-\alpha x^\beta}}{e^{-\alpha x^\beta}} = \alpha\beta x^{\beta-1} \qquad , \alpha, \beta > 0$$

$$\Rightarrow h = \begin{cases} \nearrow & , \beta > 1 \quad \widehat{=} \ \textit{Abnutzung} \\ konst & , \beta = 1 \quad \widehat{=} \ \textit{keine Alterung} \\ \searrow & , \beta < 1 \quad \widehat{=} \ \textit{System wird stabiler} \Rightarrow \textit{Anfangsprobleme} \end{cases}$$

Auch gewünscht:



Badewannenverteilung
Ausfallrate in *3 Phasen*:

1. „Kinderkrankheiten“
2. Nutzungsproblem
3. Abnutzung, Verschleiß

$\hookrightarrow$ Andere Verteilungsfunktion notwendig

10.6 Ergänzungen zu stetigen Wahrscheinlichkeitsverteilungen

d.h. Wahrscheinlichkeitsverteilung mit Riemann-Dichten.

Bezeichnung 10.10
$\hookrightarrow$ *Verteilungsfunktion wurde eingeführt als Integral mit oberer Grenze als Argument*
Die so eingeführte Verteilungsfunktion hat (sogar) die Eigenschaft der
ABSOLUTEN STETIGKEIT.

Sprich: „X hat ABSOLUT STETIGE VERTEILUNG
d.h. die Dichte von X ist gegeben durch $F'(x) = f(x)$

Bemerkung:

Die Zufallsvariablen $X_1, \ldots, X_n : (\Omega, \mathfrak{A}, P) \to (\mathbb{R}, \mathfrak{B})$ sind stochastisch unabhängig

$$\Longleftrightarrow F^{(X_1, \ldots, X_n)}(x_1, \ldots, x_n) = F^{X_1}(x_1) \cdot \ldots \cdot F^{X_n}(x_n) \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

d.h. $P(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_{i=1}^{n} P(X_i \leqslant x_i) \quad \forall x_i \in \mathbb{R}$
ist notwendig und hinreichend für die stochastische Unabhängigkeit von $X_1, \ldots, X_n$

Nun zum Zusammenhang zwischen Randdichten und stochastischer Unabhängigkeit

Lemma 10.11
Sei $X = (X_1, \dots, X_n)$ ein absolut stetiger Zufallsvektor mit Dichtefunktion f^X.
Dann gilt für die i-te Randdichte

$$f^{X_i}(x_i) = \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty}}_{n-1} f^X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \qquad x \in \mathbb{R}$$

Ferner gilt:
$f^X(x_1, \dots, x_n) = \prod_{i=1}^{n} f^{X_i}(x_i) \quad \forall x_i \in \mathbb{R}$
$\Longleftrightarrow X_1, \dots, X_n$ *stochastisch unabhängig mit Dichten* $f^{X_1}, \dots, f^{X_n}$

Beweis:

$$\begin{aligned} F^{X_i} &\overset{10.5\ \&10.7}{=} \overset{1}{\int_{-\infty}^{+\infty}} \dots \overset{i-1}{\int_{-\infty}^{+\infty}} \overset{i}{\int_{-\infty}^{x}} \overset{i+1}{\int_{-\infty}^{+\infty}} \dots \overset{n}{\int_{-\infty}^{+\infty}} f^X(y) dy_1 \dots dy_n \\ \Rightarrow f^{X_i}(x) &= \left(F^{X_i}\right)'(x) \end{aligned}$$

Der Rest ergibt sich aus der vorigen Bemerkung:

$$\begin{aligned} f^X(x_1, \dots, x_n) &= \prod_{i=1}^{n} f^{X_i}(x_i) \\ \Rightarrow F^X(x_1, \dots, x_n) &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \prod_{i=1}^{n} f^{X_i}(y) dy_1 \dots dy_n \\ &= \prod_{i=1}^{n} \underbrace{\int_{-\infty}^{x_i} f^{X_i}(y_i) dy_i}_{F^{X_i}(x_i)} \end{aligned}$$

$\Rightarrow X_1, \dots, X_n$ stochastisch unabhängig

Die Umkehrung ist klar.

Nun zu Transformationen von Zufallsvariablen und deren Verteilungen.
Fragen:

- $X_1, \dots, X_n$ *iid* $\sim$ *Exp*(λ)
  $\Rightarrow \sum_{i=1}^{n} X_i \sim ?$
- Wie ist die gemeinsame Verteilung von $(X_1, X_1 + X_2, \dots, \sum X_i)$?
  (Komponenten sind nicht stochastisch unabhängig)

Satz 10.12 (Transformationssatz für Dichten)
Sei $X = (X_1, \dots, X_n)$ ein absolut stetiger Zufallsvektor auf $(\Omega, \mathfrak{A}, P)$ mit der Dichte f^X

- *Es existiert eine offene Menge*
  $M = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n | \ f^X(x_1, \dots, x_n) \neq 0 \quad \forall (x_1, \dots, x_n) \in M^C\right\}$

- *Sei* $T : (\mathbb{R}^n, \mathfrak{B}^n) \to (\mathbb{R}^n, \mathfrak{B}^n)$ *eine messbare Abbildung mit*
  1. $\widetilde{T} = T|_M$ *ist injektiv*
     ($T|_M$ *ist die Einschränkung von* T *auf die Menge* M*)*
  2. *alle partiellen Ableitungen von* $\widetilde{T}$ *sind stetig auf* M
  3. *die sog. Funktionaldeterminante erfüllt*
     $det \left(\frac{\delta \widetilde{T}_i}{\delta x_j} \right)_{1 \leqslant i,j \leqslant n} \neq 0 \quad \forall x_j \in \mathbb{R}$

Dann folgt, dass der Zufallsvektor $Y = \widetilde{T}(X)$ *ebenso absolut stetig ist mit der Dichte:*

$$\begin{aligned} f^Y(y_1, \ldots, y_n) &= \left| det \left(\frac{\delta \widetilde{T}_i}{\delta x_j} \right)_{1 \leqslant i,j \leqslant n} \middle| \widetilde{T}^{-1}(y_1, \ldots, y_n) \right|^{-1} \cdot f^X(\widetilde{T}^{-1}(y_1, \ldots, y_n)) \cdot \mathcal{I}_{\widetilde{T}(M)}(y_1, \ldots, y_n) \\ &= \left| det \left(\frac{\delta \widetilde{T}_i^{-1}}{\delta y_j} \right)_{1 \leqslant i,j \leqslant n} \right| \cdot f^X(\widetilde{T}^{-1}(x_1, \ldots, x_n)) \cdot \mathcal{I}_{\widetilde{T}(M)}(x_1, \ldots, x_n) \end{aligned}$$

Beweis siehe Literatur.

Korrolar 10.13 *Sei* (X_1, X_2) *ein absolut stetiger Zufallsvektor mit der Dichtefunktion* $f^{(X_1,X_2)}$*, dann ist* $Y = X_1 + X_2$ *absolut stetig mit der Dichte*

$$f^Y(y) = \int_{-\infty}^{+\infty} f^X(t, y-t)dt \qquad y \in \mathbb{R}$$

Beweis:

Betrachte die Abbildung T mit $T(x_1, x_2) = (x_1, x_1 + x_2)$,
dann ist $T^{-1}(y_1, y_2) = (y_1, y_2 - y_1)$

$$\begin{aligned} \Rightarrow det \left(\frac{\delta \widetilde{T}_i}{\delta x_j} \right)_{1 \leqslant i,j \leqslant n} &= det \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \\ &= 1 \star \end{aligned}$$

$$\Rightarrow f^{T(X)}(y_1, y_2) = \frac{1}{1\star} \cdot f^X(y_1, y_2 - y_1)$$

Die Dichte von $X_1 + X_2$ ist die zweite Randdichte von $T(X)$,also

$$f^Y(y) = \int_{-\infty}^{+\infty} f^X(y_1, y - y_1)dy \qquad \square$$

Speziell gilt:
Bemerkung:

Sind X_1 und X_2 stochastisch unabhängig, dann ist

$$f^{(X_1,X_2)}(x_1,x_2) = f^{X_1}(x_1) \cdot f^{X_2}(x_2) \qquad \forall x_1, x_2 \in \mathbb{R}$$

und damit

$$f^{X_1+X_2}(y) = \int\limits_{-\infty}^{+\infty} f^{X_1}(t) \cdot f^{X_2}(y-t)dt \qquad y \in \mathbb{R}$$

10.7 Faltung von stetigen Wahrscheinlichkeitsverteilungen

Bezeichnung 10.14 *Die Verteilung von $X_1 + X_2$ heißt* Faltung *von X_1 und X_2 (Vgl. 6.15 für diskrete Wahrscheinlichkeitsverteilung)*

Beispiel 10.4

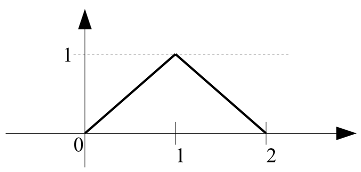
1. *X_1, X_2 seien stochastisch unabhängigiid $\sim R(0,1)$ und $y \in [0,2]$*

$$\begin{aligned} f^{X_1+X_2}(y) &= \int\limits_{-\infty}^{+\infty} f^{X_1}(t) \cdot f^{X_2}(y-t)dt \\ &= \int_0^1 \mathcal{I}_{(0,1)}(t) \cdot \mathcal{I}_{(0,1)}(y-t)dt \end{aligned}$$

$$\begin{aligned} \Rightarrow\ & t \in (0,1) \ \wedge\ y-t \in (0,1) \\ & 0 < t < 1 \wedge 0 < y-t < 1 \\ & 0 < t < 1 \wedge y-1 < t < y \quad \Rightarrow \quad \begin{cases} \int_0^1 \mathcal{I}_{(0,y)}(t)dt & 0 \leqslant y \leqslant 1 \\ \int_0^1 \mathcal{I}_{(y-1,1)}(t)dt & 1 \leqslant y \leqslant 2 \end{cases} \\ & \qquad = \begin{cases} \int_0^y dt \quad = y & 0 \leqslant y \leqslant 1 \\ \int_{y-1}^1 dt \ = 2-y & 1 \leqslant y \leqslant 2 \end{cases} \end{aligned}$$

Klar ist: $f^{X_1+X_2}(y) = 0 \qquad \forall y \notin [0,2]$

Bezeichnung 10.15 *$X_1 + X_2$ besitzt eine* Dreiecksverteilung.



2. *$X_1 \sim \Gamma(\alpha, p_1)$, $X_2 \sim \Gamma(\alpha, p_2)$ stochastisch unabhängig(vgl 10.4)*
   $\Rightarrow X_1 + X_2 \sim \Gamma(\alpha, p_1 + p_2)$

Bezeichnung 10.16 *Die Klasse der Γ-Verteilungen (bei festem Parameter) ist* FALTUNGSSTABIL, *d.h. die Faltung führt nicht aus der Klasse der Verteilung heraus.*

Speziell: $p_1 = p_2 = 1 \hookrightarrow Exp(\alpha)$

$$X_1, \dots, X_n iid Exp(\alpha) \Rightarrow \sum_{i=1}^{n} X_i \sim \Gamma(\alpha, n)$$

heißt ERLANG-VERTEILUNG

3. $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ $i = 1, 2$ *stochastisch unabhängig (vgl. 10.5)*

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

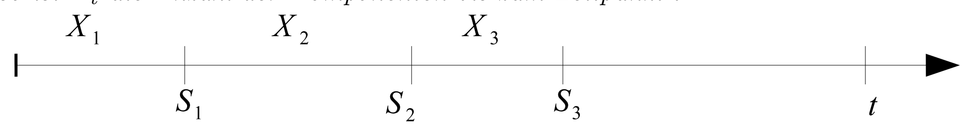
Summen von (stochastisch unabhängig) Zufallsvariablen spielen wichtige Rollen in vielen Bereichen der Stochastik:
$\hookrightarrow$ Statistik, Erneuerungstheorie, Warteschlangen, Versicherungsmathematik

Definition 10.17 *Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von stochastisch unabhängig iid. nicht-negativen Zufallsvariablen.*
Die Folge der Partialsummen $(\underbrace{\sum_{i=1}^{n} X_i}_{=:S_n})_{n\in\mathbb{N}}$ *mit* $S_0 = 0$
heißt ERNEUERUNGSPROZESS.
Für jedes $t \geqslant 0$ wird die Zufallsvariable N_t definiert durch

$$N_t := \sup(n \in \mathbb{N}_0,\ S_n \leqslant t) \qquad (= \sum_{i=1}^{\infty} \mathcal{I}_{\{S_i \leqslant t\}})$$

$(N_t)_{t\geqslant 0}$ *heißt* ERNEUERUNGSZÄHLPROZESS

Beispiel 10.5 *Seien X Lebensdauern von Komponenten und S_n die Gesammtlebensdauer, so ist N_t die Anzahl der Komponenten bis zum Zeitpunkt t*



Es gilt: $S_n \leqslant t \iff N_t \geqslant n \qquad \forall t > 0 \forall n \in \mathbb{N}_0$

Lemma 10.18 *Aus der Situation aus vorigem Beispeil mit $X_i \sim Exp(\lambda) \quad \forall i \in \mathbb{N}$ gilt*

$$N_t \sim po(\lambda \cdot t) \qquad \forall t > 0$$

BEWEIS:

$$\begin{aligned} P(N_t = k) &= P(N_t \geqslant k) - P(N_t \geqslant k+1) \\ &= P(\underbrace{S_k}_{\Gamma(\lambda,k)} \leqslant t) - P(\underbrace{S_{k+1}}_{\Gamma(\lambda,k+1)} \leqslant t) \qquad (10.21) \\ &= \int_0^t \frac{\lambda^k}{(k-1)!} \cdot e^{-\lambda x} \cdot x^{n-1} \ - \int_0^t \frac{\lambda^{k+1}}{k!} \cdot e^{-\lambda x} \cdot x^k dx \end{aligned}$$

$$\begin{aligned}
\text{(partielle Ableitung)} \quad &= \left[\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda x}\frac{x^k}{k}\right]_0^t + \frac{\lambda^k}{(k-1)!}\cdot\int_0^t \lambda e^{-\lambda x}\frac{x^k}{k}dx \ - \dots \\
&= \left[\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda x}\frac{x^k}{k}\right]_0^t \\
&= \frac{\lambda^k}{k!}\cdot e^{-\lambda t}\cdot t^k \\
&= \frac{(\lambda\cdot t)^k}{k!}e^{-\lambda t}
\end{aligned}$$

Bezeichnung 10.19 *Der sogenannte* STOCHASTISCHE PROZESS *in vorigem Lemma heißt* POISSON-PROZESS
$\hookrightarrow$ *Erneuerungstheorie, Warteschlangentheorie, Versicherungstheorie*

Bemerkung:

Gegeben sei 10.18

$$\begin{aligned}
E(N_t) &= \lambda\cdot t \qquad \forall t>0 \\
\rightarrow E(N_{t+\alpha}-N_t) &= \lambda(t+\alpha)-\lambda\cdot t = \lambda\cdot\alpha \\
&= \frac{\alpha}{EX_1}
\end{aligned}$$

Entspricht der Anzahl der Erneuerungen im Intervall $[t, t+\alpha]$

11 Grundlagen der Simulation

Simulation:

Werkzeug zur Analyse von Zufallsphänomenen bzw. Situationen, in denen analytische Behandlung zu aufwendig/unmöglich ist.
z.B. Laufzeitanalyse von Algorithmen, Ausfallverhalten von (beliebigen) Systemen

Grundlage: gleichverteilte PSEUDOZUFALLSZAHLEN
$\hookrightarrow$ Zahlengenerator

Erzeugung von sogenannten Pseudozufallszahlen:
deterministische und reproduzierbare Folge $(x_i)_{i\in\mathbb{N}} \subset [0,1]$
$\hookrightarrow$ „stetig gleichverteilt“
$\hookrightarrow$ wegen Genauigkeit nur diskret approximierbar !

Ziel:

Realisierung von Pseudozufallszahlen, die von stochastisch unabhängigen Zufallsvariablen $(X_i)_i \sim R[0,1]$ nicht unterscheidbar sind.
$\hookrightarrow$ Gütekriterien für Zufallszahl.

Gegeben sei ein Grundraum $\Omega_m := \{\frac{0}{m}, \frac{1}{m}, \dots, \frac{m-1}{m}\}$ $\quad m >> 1$ mit diskreter Gleichverteilung.

Lemma 11.1 *Für $[u,v] \subset [0,1]$ gilt:*

$$|P_m(\{\omega \in \Omega_m|\ u \leqslant w \leqslant v\}) - (v-u)| \leqslant \frac{1}{m}$$

Wobei $v-u$ die Wahrscheinlichkeit nach der Rechteckverteilung für das Intervall $[u,v]$ ist.

Beweis:

$$\begin{aligned} &\forall u < v\ \exists i,j \in \mathbb{Z} \text{ mit } 0 \leqslant i \leqslant j \leqslant m-1 \\ &\text{und } \frac{i}{m} \leqslant u \leqslant \frac{i+1}{m}, \frac{j}{m} \leqslant v \leqslant \frac{j+1}{m} \end{aligned}$$

$$\begin{aligned} \Rightarrow P_m(\{\omega \in \Omega_m|\ u \leqslant w \leqslant v\}) &= \begin{cases} \frac{j-i+1}{m} & ,u = \frac{i}{m} \\ \frac{j-i}{m} & ,u > \frac{i}{m} \end{cases} \\ &= \begin{cases} \frac{j-i}{m} \leqslant v-u \leqslant \frac{j+1-i}{m} & ,u = \frac{i}{m} \\ \frac{j-i-1}{m} \leqslant v-u \leqslant \frac{j-i}{m} & ,u > \frac{i}{m} \end{cases} \end{aligned}$$

Das heißt, das Wahrscheinlichkeitmaß P_m (mit diskreter Gleichverteilung) approximiert für großes m eine stetige Gleichverteilung auf $[0,1]$.

Beschreibung einer n-maligen, unabhängigen, zufälligen Auswahl einer Zahl aus Ω_m durch den Grundraum

$$\Omega_m^n := \{\omega = (\omega_1, \dots, \omega_n)|\ \omega_j \in \Omega_m, j \in \{1, \dots, n\}\}$$

und die Gleichverteilung

$$P_m^n(\{\omega\}) := \prod_{j=1}^{n} P_m(\{\omega_j\}) = \left(\frac{1}{m}\right)^n \qquad ,\omega = (\omega_1, \dots, \omega_n)$$

Lemma 11.2 *Sind $r_1, \dots, r_n, s_1, \dots, s_n \in [0,1]$ mit $|r_j - s_j| \leqslant \varepsilon,\ \forall j \in \{1, \dots, n\}$, so folgt*

$$\left|\prod_{j=1}^{n} r_j - \prod_{j=1}^{n} s_j\right| \leqslant n \cdot \varepsilon$$

Beweis durch vollständige Induktion:

$n = 1$ Klar

$$\begin{aligned} n \to n+1: \quad & \left|\prod_{j=1}^{n+1} r_j - \prod_{j=1}^{n+1} s_j\right| \\ =& \left|\left(\prod_{j=1}^{n} r_j - \prod_{j=1}^{n} s_j\right) \cdot r_{n+1} + \left(\prod_{j=1}^{n} s_j\right) \cdot (r_{n+1} - s_{n+1})\right| \\ \leqslant& \left|\left(\prod_{j=1}^{n} r_j - \prod_{j=1}^{n} s_j\right) \cdot r_{n+1}\right| + \left|\left(\prod_{j=1}^{n} s_j\right) \cdot (r_{n+1} - s_{n+1})\right| \end{aligned}$$

$$\begin{aligned} &\overset{IV}{\leqslant} \underbrace{r_{n+1}}_{\leqslant 1} \cdot n\varepsilon + \underbrace{\left(\prod_{j=1}^{n} s_j\right)}_{\leqslant 1} \underbrace{(r_{n+1} - s_{n+1})}_{\leqslant 1} \\ &\leqslant (n+1) \cdot \varepsilon \end{aligned}$$

Bemerkung: Anwendung von 11.2 auf

$$r_j := P_m(\{\omega_j \in \Omega_m | \ u_j \leqslant \omega_j \leqslant v_j\}) \text{ und } s_j := r_j - u_j$$

ergibt mit 11.1 und $P_m^n(\times B_j) = \prod P_m(B_j)$ (Unabhängigkeit)

$$\left| P_m^n(\{\omega \in \Omega_m^n | \ u_j \leqslant \omega_j \leqslant r_j \ , 1 \leqslant j \leqslant n\}) - \prod_{j=1}^{n} (r_j - u_j) \right| \leqslant \frac{n}{m}$$

Das heißt:

> Das Wahrscheinlichkeitmaß P_m^n nähert sich bei festem n und wachsendem m dem Modell der stetigen Gleichverteilung auf dem n-dimensionalen Einheitswürfel $[0,1]^n$ an.
>
> Diese Gleichverteilung ordnet jedem „Rechteck“ $\bigtimes_{i=1}^{n} [u_i, v_i]$
> sein Volumen $\prod_{i=1}^{n} (v_i - u_i)$ zu.

Notwendig sind weiterhin Güte-Kriterien für Pseudozufallszahlen, die von Zufallsgeneratoren erzeugt wurden.
$\hookrightarrow$ Statistik

11.1 Der lineare Kongruenzgenerator

Bezeichnung 11.3
Sei $m \in \mathbb{N}_0$ *(Modul),* $a \in \mathbb{N}_0$ *(Faktor),*
$b \in \mathbb{N}_0$ *(Inkrement),* $z_0 \in \mathbb{N}_0, z_0 \leqslant m - 1$ *(Anfangsglied).*

Kongruenzschema:

$$z_{j+1} \equiv (a \cdot z_j + b)(mod\ m) \qquad , j \in \mathbb{N}_0$$

Klar ist: $0 \leqslant z_j \leqslant m - 1 \quad \forall j \in \mathbb{N}_0$

Normierung:
$x_j := \frac{z_j}{m}, j \in \mathbb{N}_0$ *liefert eine Folge* $(x_n)_{n \in \mathbb{N}_0} \subset [0,1]$

Beispiel 11.1 (linearer Kongruenzgenerator)
Sei $m = 100, a = 18, b = 11, z_0 = 40$

$$\begin{aligned} x_0 &= \frac{40}{100} = 0,4 \\ z_1 &= 18 \cdot 40 + 11 = 731 \equiv 31 \\ &\rightarrow x_1 = 0,31 \end{aligned}$$

$$\begin{aligned} z_2 &= 18 \cdot 31 + 11 = 569 \equiv 69 \\ &\rightarrow x_2 = 0,69 \\ z_3 &= 18 \cdot 69 + 11 = 1253 \equiv 53 \\ &\rightarrow x_3 = 0,53 \end{aligned}$$

Bemerkung: zu obigem Beispiel.
Wegen $z_j \in \{0, \ldots, m-1\}, j \in \mathbb{N}_0$ kann jeder lineare Kongruenzgenerator mit Modul m höchstens m verschiedene Zufallszahlerzeugen.
$\Rightarrow$ also zunächst m groß wählen !

Aber in dem Beispiel ergibt sich weiterhin: $z_4 = 65,\ z_5 = 81,\ z_6 = 69 = z_2$

Hier enstehen also nur 6 verschiedene Zahlen !
D.h. ein Generator mit Periode der Länge 4
$(65, 53, 65, 81)$

Ziel:

Durch Wahl der Paramater die maximal mögliche Periodenlänge m sicherstellen!

Satz 11.4 *Für $b \geqslant 1$ wird die maximal mögliche Periodenlänge genau dann erreicht, wenn:*

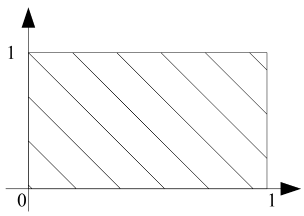
1. *b ist teilerfremd zu m*
2. *Jede Primzahl, die m teilt, teilt auch $a-1$*
3. *Ist m durch 4 teilbar, so muß auch $a-1$ durch 4 teilbar sein*

$\hookrightarrow$ *Zahlentheorie*

Regel 11.5 *Auch bei maximaler Periodenlänge dürfen nicht alle möglichen Pseudozufallszahlen für eine Simulation „verbraucht“ werden. Etwa wäre stets die letzte Zahl vorhersagbar.*

Bemerkung:

Es gibt auch prinzipielle Schwächen linearer Kongruenzgeneratoren.
$\hookrightarrow$ Gitterstruktur
d.h. Paare $(x_0, x_1), (x_2, x_3), \ldots$ können (relativ zu m) auf verhältnismäßig wenigen, parallelen Bahnen liegen.



Auswege:

1. Verwendung von wenigen Punkten relativ zu m in einer Simulation (Zur Vermeidung Systematischer Fehler, etwa Realisierungsfehler des Generators)

2. Verwendung „besserer“ Zufallsgeneratoren
   $\hookrightarrow$ Siehe weiterführende Literatur

Beispiel 11.2 *Ein Zufallsexperiment*
Erzeugung von 2^n Zahlen zur empirischen Überprüfung der Ergebnisse einer binären Suche

Situation der Simulation

Gegeben sei ein Experiment, bei dem Ereignis E_j mit der Wahrscheinlichkeit p_j eintritt
$(1 \leqslant j \leqslant s, \sum_{j=1}^{s} p_j = 1)$
$\hookrightarrow$ Erzeugung von Pseudozufallszahlen x_i
$\rightarrow$ feststellen, in welchem der disjunkten Intervalle
$[0, p_1), [p_1, p_1 + p_2), \ldots, [p_1 + \ldots + p_{s-1}, 1]$
die Zahl x_i liegt. Ist

$$x_i \in \left[\sum_{k=1}^{j-1} p_k \ , \ \sum_{k=1}^{j} p_k \right),$$

so ordnet man zu: Das Experiment liefert Ereignis E_j

Beispiel 11.3 *Laplace-Experiment mit s Ausgängen*
$\rightarrow p_1 = \ldots = p_s = \frac{1}{s}$
Anhand der Zufallszahl wird der Ausgang des Experiments durch $\lfloor x \cdot s \rfloor + 1$ simuliert, denn die Bedingung

$$\frac{j-1}{s} \leqslant x \leqslant \frac{j}{s} \quad \textit{ist äquivalent zu} \quad j = \lfloor x \cdot s \rfloor + 1$$

Grundsätzlich: **Vorsicht bei Simulationen !**

- Viele Durchläufe machen.
- verschiedene Generatoren verwenden.
- verschiedene Startwerte benutzen

Nun zu Zufallszahl aus anderen Verteilungen.
Ausgangspunkt:

Zufallsvariablen $Z_1, Z_2, \ldots iid \sim R[0, 1]$
bzw. deren Repräsentation durch Pseudozufallszahlen.

Ziel: Zufallszahl aus der Verteilung mit Verteilungsfunktion F

Beispiel 11.4 $Z \sim R[0, 1]$

1. *$X = a + (b - a) \cdot Z$ ist $R[a, b]$ verteilt*
   $\hookrightarrow$ *entsprechende Transformation der Pseudozufallszahlen*

$$\begin{aligned} P(X \leqslant x) &= P(a + (b-a) \cdot Z \leqslant x) \\ &= P\left(Z \leqslant \frac{x-a}{b-a}\right) \\ &= P\left(\frac{x-a}{b-a}\right) \end{aligned}$$

2. $X = \begin{cases} 1 & , Z \leqslant p \\ 0 & , Z > p \end{cases}$ *ist* $b(1,p)$ *verteilt.*

$$P(X = 1) = P(Z \leqslant p) = p$$

d.h. die Erzeugung von $b(1,p)$ *verteilten Pseudozufallszahlen durch Vergleich mit* p*.*
$\hookrightarrow$ *allgemeines Vorgehen ?*

Verfahren 11.6 (zur Erzeugung von Zufallszahl nach Verteilung mit Verteilungsfunktion F)
Voraussetzung: F *ist streng monoton auf dem Träger*

$$\begin{array}{rcl} \textit{Zuordnung: } X & := & F^{-1}(Z) \\ \Rightarrow X & \sim & F \\ P(X \leqslant x) & = & P(F^{-1}(Z) \leqslant x) \\ & = & P(Z \leqslant F(x)) \\ & = & F(x) \qquad , \forall x \end{array}$$

Also folgt aus einer gegebenen Folge von Zufallszahl $(x_i)_{i \in \mathbb{N}}$ *aus* $P(0,1)$
$(F^{-1}(x_i))_{i \in \mathbb{N}}$ *ist Folge von Zufallszahl aus* F

Beispiel 11.5 *Sei* $F(x) = 1 - e^{-\lambda x}, \lambda > 0, x > 0$ *exponentialverteilt und* $Z \sim R[0,1]$

$$\begin{array}{rcl} \hookrightarrow F^{-1}(t) & = & -\frac{1}{\lambda} \ln(1-t) \qquad , t \in [0,1) \\ \Rightarrow X & = & F^{-1}(Z) \\ & = & -\frac{1}{\lambda} \ln(1-Z) \end{array}$$

Ist wieder exponentialverteilt.

$\hookrightarrow$ *Näheres in den mannigfaltigen Büchern*
zu Simulations und Erzeugung von Pseudozufallszahlen

12 Einführung in die Statistik

Grundaufgabe: Aus Beobachtungen Rückschlüße auf die (ganz oder teilweise unbekannte) Wahrscheinlichkeitsverteilung ziehen.

Ausgangspunkt:

> Stochastisches Modell mit Zufallsvariablen $X_1, \ldots, X_n : (\Omega, \mathfrak{A}, P) \to (\mathbb{R}, \mathfrak{B}, P^X)$ *iid* und Realisation $x_1, \ldots, x_n$ von $X_1, \ldots, X_n$ (Stichprobe vom Umfang n)

Ziel: Aussagen über P^X

Beispiel 12.1 $X_1, \ldots, X_n$ *iid* $\sim \mathcal{N}(\mu, \sigma^2)$, μ *unbekannt*

$$\hookrightarrow \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \textit{als Schätzwert für } \mu$$

Alle möglichen Aussagen bzw. Entscheidungen werden zusammengefasst zu der Entscheidungsmenge $\mathcal{D}$

Beispiel 12.2

1. *Bekannt:* $X_i \sim Exp(\lambda)$, λ *unbekannt,* $EX_i = \frac{1}{\lambda}$

   Frage: Ist $EX_i < \frac{1}{\lambda_0}$ *oder* $EX_i \geqslant \frac{1}{\lambda_0}$, λ_0 *gegeben ?*
   $\hookrightarrow$ *2 Entscheidungen:* $d_0 : \lambda > \lambda_0$ *oder* $d_1 : \lambda \leqslant \lambda_0$
   d.h. $\mathcal{D} = \{d_0, d_1\}$

2. *Bekannt:* $X_i \sim b(n, p)$, n *bekannt,* p *unbekannt.*

   $\hookrightarrow$ *bei Schätzung von* p *ist* $\mathcal{D} = [0, 1]$

Antworten werden beschrieben durch:

Definition 12.1 *Die Menge aller möglichen Realisationen ist* $\mathfrak{X}^n = (X(\Omega))^n$.
Jede messbare Abbildung $\delta : \begin{cases} \mathfrak{X}^n & \to & \mathcal{D} \\ x = (x_1, \ldots, x_n) & \mapsto & d \end{cases}$
heißt STATISTISCHE ENTSCHEIDUNGSFUNKTION *(SEF)*

Nun eine Formalisierung von „Parameter ist unbekannt“.

Bezeichnung 12.2 *Eine Klasse* $\mathcal{P}$ *von Wahrscheinlichkeitsverteilungen heißt* PARAMETRISCH, *falls*

$$\exists \Theta \subset \mathbb{R}^k, \quad \textit{Abbildung } h : \begin{cases} \Theta & \to & \mathcal{P} \\ \vartheta & \mapsto & P_\vartheta \end{cases}, \quad h \textit{ bijektiv}$$

(Identifizierbarkeit)

Beispiel 12.3 *für* Θ

$$\begin{aligned} P &= \{b(n,p) |\ 0 \leqslant p \leqslant 1\},\ n\ fest \\ \to \Theta &= [0,1] \\ \\ P &= \{Exp(\lambda) |\ \lambda > 0\} \\ \to \Theta &= (0,\infty) \\ \\ P &= \{\mathcal{N}(\mu,\sigma^2) |\ \mu \in \mathbb{R}, \sigma > 0\} \\ \to \Theta &= \mathbb{R} \times \mathbb{R}^{>0} \\ & \quad d.h.\ \vartheta = (\vartheta_1, \vartheta_2)\ mit \\ & \quad \vartheta_1 = \mu,\ \vartheta_2 = \sigma^2 \end{aligned}$$

Bezeichnung 12.3 *Gegeben sei eine parametrische Klasse $\mathcal{P} = \{P_\vartheta |\ \vartheta \in \Theta\}$ von Wahrscheinlichkeitsverteilungen.*
Eine SEF (statistische Entscheidungsfunktion) $\delta : \mathfrak{X}^n \to \Theta$ heißt SCHÄTZFUNKTION. *(Parameterschätzung, Punktschätzung)*

Beispiel 12.4

1.

$$\begin{aligned} P &= \{b(1,p) |\ 0 \leqslant p \leqslant 1\} \\ \mathfrak{X}^n &= \{0,1\}^n \\ \Theta &= [0,1] \\ &= \mathcal{D} \end{aligned}$$

mögliche Schätzfunktionen für p:

$$\begin{aligned} \delta_1(x) &= x_1 \\ \delta_2(x) &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &\vdots \end{aligned}$$

2. $P = \{R[a,b] |\ a < b,\ b \in \mathbb{R}\},\ a\ fest$
mögliche Schätzfunktion für b:
$\delta(x) := \max\{x_1, \ldots, x_n\}$

3. $P = \{\mathcal{N}(\mu,\sigma^2) |\ \mu \in \mathbb{R}, \sigma > 0\},\ \Theta = \mathbb{R} \times \mathbb{R}^{>0}$
mögliche Schätzfunktionen für (μ, σ^2):

$$\begin{aligned} \delta_1(x) &= \Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{=\bar{x}}\ ,\ \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}_{empirische\ Varianz}\Big) \\ \delta_2(x) &= \Big(\frac{x_1 + x_2}{2}\ ,\ \underbrace{\max\{x_1, \ldots, x_n\} - \min\{x_1, \ldots, x_n\}}_{Spannweite}\Big) \end{aligned}$$

Allgemeine Fragen:

- Wahl, Bestimmung, Herleitung von Schätzfunktionen
- „bestmögliche“ Auswahl, Optimalitätskriterien
- Eigenschaften, dazu:

Definition 12.4
Gegegeben sei eine parametrische Klasse $\mathcal{P}$ von Wahrscheinlichkeitsverteilungen, ein STICHPROBENRAUM $\mathfrak{X}^n$*, ein Parameterraum* $\Theta = \mathcal{D} \subset \mathbb{R}$ *und eine Menge* $\Delta = \{\delta |\ \delta : \mathfrak{X}^n \to \Theta\}$ *von Schätzfunktionen.*

$\delta \in \Delta$ *nennt man* ERWARTUNGSTREU*, falls gilt:*

$$\begin{aligned} E_\vartheta(\delta(x)) &= \vartheta \qquad \forall \vartheta \in \Theta \\ &= \int \delta(x) \cdot f_\vartheta(x) dx \quad \textit{bzw.} \\ & \quad \sum \delta(x_i) \cdot p_\vartheta(x_i) \end{aligned}$$

(„Im Mittel liefert die Schätzfunktion den wahren Wert“)

Fortsetzung des vorigen Beispiels:

1. $E_\vartheta \delta_1(x) = E_\vartheta X_1 = \vartheta \quad \forall \vartheta \in \Theta$ [2]
   d.h. δ_1 ist erwartungstreu

   $E_\vartheta \delta_2(x) = E_\vartheta \left(\frac{1}{n}\sum_{i=1}^n x_i\right) = \vartheta$
   $\Rightarrow \delta_2$ ist erwartungstreu

   Aber, mit

$$\begin{aligned} Var(x_i) &= \sigma^2 \\ Var(\delta_1(x)) &= Var(x_1) \\ &= \sigma^2 \\ Var(\delta_2(x)) &= Var\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \\ &= \frac{\sigma^2}{n} \qquad ! \end{aligned}$$

2. Zunächst die Verteilungsfunktion von $\max\{X_1, \ldots, X_n\} =: X_{(n)}$ bestimmen

$$\begin{aligned} P\left(X_{(n)} \leqslant x\right) &= P(X_1 \leqslant x, \ldots, X_n \leqslant x) \\ &\overset{unabh}{=} \prod_{i=1}^n P(X_i \leqslant x) \\ &\overset{iid}{=} \left(P(X_1 \leqslant x)\right)^n \end{aligned}$$

[2] stimmt das ?

$$
\begin{aligned}
f^{X_1}(x) &= \begin{cases} \frac{1}{b-a} & , x \in [a,b] \\ 0 & , \text{ sonst} \end{cases} \\
F^{X_1}(x) &= \begin{cases} 0 & , x \leqslant a \\ \frac{x-a}{b-a} & , x \in [a,b] \\ 1 & , x \geqslant b \end{cases} \\
\Rightarrow P(X_{(n)} \leqslant x) &= \left(\frac{x-a}{b-a}\right)^n =: G(x) \quad , x \in [a,b] \\
\text{mit Dichte } g(x) &= \left(\frac{x-a}{b-a}\right)^{n-a} \cdot \frac{n}{b-a}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow EX_{(n)} &= \int_a^b x \cdot \frac{n}{b-a} \cdot \left(\frac{x-a}{b-a}\right)^{n-a} dx \\
&= \frac{n}{(b-a)^n} \cdot \int_a^b x \cdot (x-a)^{n-1} dx \\
&= \frac{n}{(b-a)^n} \int_0^{b-a} (x+a) \cdot x^{n-1} dx \\
&= \frac{n}{(b-a)^n} \cdot \left[\frac{x^{n+1}}{n+1} + a\frac{x^n}{n}\right]_0^{b-a} \\
&= \ldots \\
&= \frac{n}{n+1} b + \frac{a}{n+1}
\end{aligned}
$$

d.h. δ ist nicht erwartungstreu, aber mit $n \to \infty$
„asymptotisch erwartungstreu“

Weiterhin: a ist bekannt und

$$
\begin{aligned}
\widetilde{\delta} \text{ definiert durch} \quad \widetilde{\delta}(x) &= \frac{n+1}{n} \cdot \left(\delta(x) - \frac{a}{n+1}\right) \\
E\widetilde{\delta}(x) &= \frac{n+1}{n} \cdot \left(E\delta(x) - E\frac{a}{n+a}\right) \\
&= \frac{n+a}{n} \left(\frac{n}{n+a} b + \frac{a}{n+1} - \frac{a}{n+1}\right) \\
&= b
\end{aligned}
$$

d.h. $\widetilde{\delta}$ ist erwartungstreu !

3. Seien

$$
\begin{aligned}
\bar{X} = \delta^{(1)}(X) &= \frac{1}{n} \sum_{i=1}^n X_i \quad \text{und} \\
\delta^{(2)}(X) &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2
\end{aligned}
$$

$$
\begin{aligned}
\text{dann ist } \delta_1(X) &= (\delta^{(1)}(X), \delta^{(2)}(X)) \\
\Rightarrow E\delta^{(1)}(X) &= \mu \\
\Rightarrow E\delta^{(2)}(X) &= \frac{1}{n-1}\sum_{i=1}^{n} E(X_i - \bar{X})^2 \\
\text{zunächst:} & \\
E(X_i - \bar{X})^2 &= EX_i^2 - 2E(X_i \cdot \bar{X}) + E\bar{X}^2 \\
&= EX_i^2 - 2 \cdot \frac{1}{n} \cdot \sum_{j=1}^{n} \underbrace{E(X_i \cdot X_j)}_{EX_i \cdot EX_j \text{ für } j \neq i} + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n} E(X_j \cdot X_l) \\
&\overset{iid}{=} EX_i^2 - 2\frac{1}{n}\Big((n-1)E^2X_1 + EX_i^2\Big) + \frac{1}{n^2}\Big(n(n-1)E^2X_1 + n \cdot EX_1^2\Big) \\
&= EX_1^2 \underbrace{\left(1 - \frac{2}{n} + \frac{n}{n^2}\right)}_{\frac{n-1}{n}} + E^2X_1\left(-\frac{2(n-1)}{n} + \frac{n(n-1)}{n}\right) \\
&= \frac{n-1}{n}EX_1^2 - E^2X_1 \qquad \text{Stimmt das ?} \\
&= \frac{n-1}{n} VarX_1 = \sigma^2 \\
\Rightarrow E\delta^{(2)}(X) &= \frac{1}{n-1} \cdot n \cdot \frac{n-1}{n} \cdot \sigma^2 \\
&= \sigma^2
\end{aligned}
$$

Weitere Optimalitätskriterien siehe Mathematische Statistik.

Nun zur Methode zur Erzeugung einer SEF.

Bezeichnung 12.5 *Gegeben sei $\mathcal{P} = \{P_\vartheta |\ \vartheta \in \Theta\}$ auf $\mathfrak{X}^n$*
mit Riemann- bzw. Lebesque-Dichten f_ϑ , $\vartheta \in \Theta$
oder diskrete Wahrscheinlichkeitsverteilung, so heißt

$$
L = \begin{cases} (\mathfrak{X}^n, \Theta) \rightarrow [0,1] \\ (x, \vartheta) \mapsto \begin{cases} f_\vartheta(x) & \textit{kontinuierlicher Fall} \\ P_\vartheta(X = x) & \textit{diskreter Fall} \end{cases} \end{cases}
$$

Likelihood-funktion *zur Realisation X.*
(Funktion des Parameters ϑ bei festem x)

Gesucht:

$$
\widehat{\vartheta} \in \Theta \ \text{ mit } \ L(x, \widehat{\vartheta}) = \sup_{\vartheta \in \Theta} L(x, \vartheta)
$$

heißt Maximum-Likelihood-Schätzfunktion (MLS)
für ϑ (bei Vorliegen von x)

Idee:

> Liegt eine Beobachtung x vor, dann wähle denjenigen Parameter $\widehat{\vartheta}$ als Schätzwert für das unbekannte ϑ, unter dem die vorliegende Beobachtung x die größte Eintrittswahrscheinlichkeit hat.
> (vgl. Beispiel 2.6 bei hypergeometrischer Verteilung)

Beispiel 12.5

1. *$X_1, \ldots, X_n$ iid $\sim b(1,p)$*

$$\begin{aligned}
\Rightarrow L(x,p) &\overset{iid}{=} \prod_{i=1}^{n} p^{X_i} \cdot (1-p)^{1-X_i} \\
&= p^{\sum X_i}(1-p)^{n-\sum X_i} \\
\Rightarrow \frac{d}{dp} \ln L(X,p) &= \frac{d}{dp}\Big(\sum X_i \ln p + (n - \sum X_i) \ln(1-p)\Big) \\
&= \frac{1}{p}\sum X_i - \frac{1}{1-p}(n - \sum X_i) \overset{!}{=} 0 \\
\Rightarrow (1-\widehat{p})\sum X_i &= \widehat{p}(n - \sum X_i) \\
\widehat{p} &= \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \textit{(hinreichende Bedingung!)}
\end{aligned}$$

d.h. das arithmetische Mittel ist MLS

2. *Seien $X_1, \ldots, X_n$ iid $\sim Exp(\lambda)$*

$$\begin{aligned}
\textit{auch hier:} \qquad f_\vartheta^X(x) &= \prod_{i=1}^{n} f_\vartheta^{X_i}(x_i) \qquad (6.18) \\
\Rightarrow L(X,\lambda) &= \prod_{i=1}^{n} \lambda \cdot e^{-\lambda X_i} \cdot \mathcal{I}_{(0,\infty)}(x_i) \\
&= \lambda^n \exp(-\lambda \sum X_i) \cdot \mathcal{I}_{(0,\infty)}(\min X_i) \\
\Rightarrow \widehat{\lambda} &= \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} \\
E\widehat{\lambda} &= \frac{n}{n-1} \cdot \frac{1}{\lambda}
\end{aligned}$$

Bezeichnung 12.6 *Sei $\mathcal{P} = \{P_\vartheta |\ \vartheta \in \Theta\}$ mit $\mathcal{P} = \mathcal{P}^0 + \mathcal{P}^1$ (disjunkt)*
Eine SEF $\delta : \mathfrak{X}^n \mapsto \mathcal{D} = \{d_0, d_1\}$
heißt (nichtrandomisierter) TEST

Interpretation: Entscheidung zwischen zwei Alternativen (vgl. Bsp 12.2)
$\hookrightarrow$ Zerlegung des Beobachtungsraums

$$\begin{aligned}
\mathfrak{X}^n &= H + A \\
\text{mit } \delta(X) &= \begin{cases} d_0 & \text{für } x \in H \quad ,\text{Hypothese} \\ d_1 & \text{für } x \in A \quad ,\text{Alternative} \end{cases} \\
\text{d.h. } H &= d^{-1}(d_0) \\
A &= d^{-1}(d_1)
\end{aligned}$$

Beispiel 12.6

$$\begin{aligned}
\mathcal{P} &= \{b(1,p)|\ p \in [0,1]\} \\
\mathfrak{X}^n &= \{0,1\}^n \\
\mathcal{P}^0 &= \{b(1,p)|\ 0 \leqslant p \leqslant p_0\} \qquad 0 < p_0 < 1 \textit{ fest} \\
\mathcal{P}^1 &= \{b(1,p)|\ p_0 < p \leqslant 1\}
\end{aligned}$$

Gesucht ist ein Test $\delta : \mathfrak{X}^n \mapsto \{d_0, d_1\}$
d.h. gesucht ist eine Zerlegung $H + A = \mathfrak{X}^n$
Intuitiv fällt man diese Entscheidung aufgrund einer Stichprobe x vom Umfang n für

$\left\{ \begin{array}{l} p \leqslant p_0 \\ p > p_0 \end{array} \right\}$ *, falls Anzahl der Einsen in x* $\left\{ \begin{array}{l} \text{„klein“} \\ \text{„groß“} \end{array} \right.$ *Ansatz:*

$H = \{x|\ \sum x_i \leqslant c\},\ A = \{x|\ \sum x_i > c\}$
mit $0 \leqslant c \leqslant n$*, so dass*

$$\delta(x) = \begin{cases} d_0 & ,\textit{falls } \sum x_i \leqslant c \\ d_1 & ,\textit{falls } \sum x_i > c \end{cases}$$

Die Wahl von c ergibt sich aus der Fehlerbetrachtung

Bezeichnung 12.7

Fehler 1.Art: $p \in P^0$ *ablehnen, obwohl richtig*

Fehler 2.Art: $p \in P^0$ *annehmen, obwohl falsch*

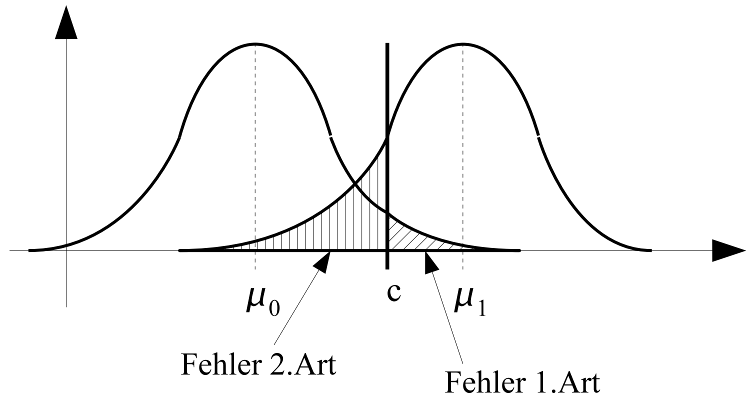
		„wahr“ P^0	„wahr“ P^1
Entscheidung:	P^0	*kein Fehler*	*Fehler 2.Art*
	P^1	*Fehler 1.Art*	*kein Fehler*

Beispiel 12.7

$$\begin{aligned}
\textit{Klasse } \mathcal{P} &= \{\mathcal{N}(\mu_0, \sigma^2), \mathcal{N}(\mu_1, \sigma^2)\} \quad ,\mu_0 < \mu_1 \\
\textit{Zerlegung } \mathcal{P}^i &= \{\underbrace{\mathcal{N}(\mu_i, \sigma^2)}_{P_i}\} \quad ,i = 0,1
\end{aligned}$$

$$\begin{array}{rcl}\mathcal{P}^0, \mathcal{P}^1 & & \textit{heißen Hypothesen} \\ \mathcal{P}^0 & : & \textit{Nullhypothese} \\ \mathcal{P}^1 & : & \textit{Gegenhypothese, Alternative}\end{array}$$

Sei $n = 1$, d.h. eine Beobachtung:



$p \in P_0$ ablehnen, falls $x > c$.
Gesucht ist eine Zerlegung $\mathbb{R} = H + A$

$$\textit{Intuitiv: } H = (-\infty, c], \qquad A = (c, \infty)$$

$$\textit{Test: } \delta(x) = \begin{cases} d_0 & , x \leqslant c \\ d_1 & , x > c \end{cases}$$

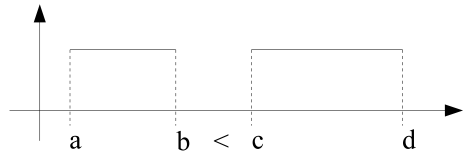
Fehlerwahrscheinlichkeiten:

$$\begin{array}{lrcl} \textit{1.Art} & P_0(X > c) & = & P_0(A) \\ \textit{2.Art} & P_1(X \leqslant c) & = & P_1(H) = 1 - P_1(A) \end{array}$$

Problem:

Im Allgemeinen ist es nicht möglich beide Fehlerwahrscheinlichkeiten simultan zu minimieren

Ausnahme:



Hier sind keine Fehler möglich. (Kaum interessanter Fall)

Ausweg:

Unsymmetrische Vorgehensweise
Vorgabe der Fehlerwahrscheinlichkeit 1.Art $P_0(A) \leqslant \alpha$
α heißt (Signifikanz-) NIVEAU *des Tests*
(Test zum Niveau α)

Dann wähle unter den möglichen Tests zum Niveau α denjenigen mit minimalem Fehler 2.Art
Zu Optimalitätskriterien und Eigenschaften siehe
$\hookrightarrow$ *mathematische Statistik.*

Aufgaben:

- *allgemeine Struktur eines Tests*
- *Wahl der sogenannten Teststatistik*
- *Konvergenz in n, d.h. Anzahl der nötigen Beobachtungen*

Beispiel 12.8 *Gegeben sei folgende Situation:*

$$\begin{aligned} X_1,\dots,X_n \text{ iid} &\sim \mathcal{N}(\mu,\sigma^2) \\ \Rightarrow \quad \sum_{i=1}^n X_i &\sim \mathcal{N}(n\cdot\mu, n\cdot\sigma^2) \\ \text{Weiterhin:} \quad aX_1 &\sim \mathcal{N}(a\cdot\mu, a^2\cdot\sigma^2) \\ \Rightarrow \quad \frac{1}{n}\sum_{i=1}^n &\sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \end{aligned}$$

Standardisierung:

$$Z := \frac{X_1-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

Standardisierung, falls σ^2 unbekannt:

$$\hookrightarrow \sigma \text{ wird geschätzt durch} \quad S := \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X})^2}$$

Die Verteilung von $Y := \dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ *heißt* T-VERTEILUNG *mit $n-1$ Freiheitsgraden.*
Dichte von Y:

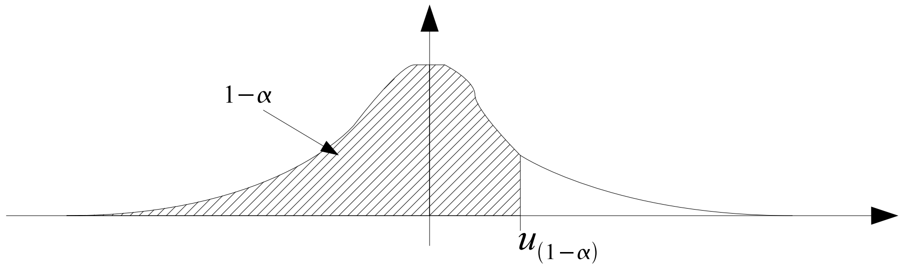
$$f^Y(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{\pi n}} \cdot \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

12.1 Tests bei Normalverteilung

Verfahren 12.8 (Ein-stichproben-Gaußtest)
Seien $X_1,\dots,X_n$ iid $\sim \mathcal{N}(\mu,\sigma^2)$, σ^2 bekannt und $\alpha \in (0,1)$ gegeben.
Test: $H:\mu \leqslant \mu_0$ gegen $A:\mu>\mu_0$ $\quad$ (μ_0 fest)

$$\text{Entscheidung:} \qquad \bar{X} \begin{Bmatrix} > \\ \leqslant \end{Bmatrix} \mu_0 + u_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad H \begin{cases} \text{ablehnen} \\ \text{annehmen} \end{cases}$$

$u_{1-\alpha}$ ist das sogenannte $1-\alpha$-Quantil von $\mathcal{N}(0,1)$



Sei $\widetilde{\mu} \leqslant \mu_0$

$$\begin{aligned}\textit{Fehler 1.Art:} \quad P_{\widetilde{\mu}}\left(\bar{X} > \mu_0 + u_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right) &= 1 - P_{\mu}\Bigg(\underbrace{\frac{\bar{X}-\widetilde{\mu}}{\frac{\sigma}{\sqrt{n}}}}_{\sim\mathcal{N}(0,1)} \leqslant \frac{\mu_0-\widetilde{\mu}}{\frac{\sigma}{\sqrt{n}}} + u_{1-\alpha}\Bigg)\\ &= 1-\Phi\left(\frac{\mu_0-\widehat{\mu}}{\frac{\sigma}{\sqrt{n}}}+u_{1-\alpha}\right)\\ &\leqslant 1-\underbrace{\Phi(u_{1-\alpha})}_{1-\alpha}\\ &= \alpha\end{aligned}$$

d.h. für jedes $\widetilde{\mu} \leqslant \mu_0$ Test zum Niveau α[3]
Entsprechend:

$H: \mu \geqslant \mu_0$ gegen $A: \mu < \mu_0$ (μ_0 fest)

$$\textit{Entscheidung:} \quad \bar{X}\left\{\begin{array}{c}<\\ \geqslant\end{array}\right\}\mu_0 - u_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}} \quad\Rightarrow\quad H_0\left\{\begin{array}{l}\textit{ablehnen}\\ \textit{annehmen}\end{array}\right.$$

Und schließlich:

$H: \mu = \mu_0$ gegen $A: \mu \neq \mu_0$ (Zweiseitiger Test)

$$\textit{Entscheidung:} \quad \bar{X}\left\{\begin{array}{c}>\\ \leqslant\end{array}\right\}\mu_0+u_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\left\{\begin{array}{c}\textit{oder} \; <\\ \textit{und} \; \geqslant\end{array}\right\}\mu_0-u_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}} \quad\Rightarrow\quad H_0\left\{\begin{array}{l}\textit{ablehnen}\\ \textit{annehmen}\end{array}\right.$$

Beispiel 12.9 *Füllmengen von 1l-Flaschen*
Modell: Realisation von $\mathcal{N}(\mu,\sigma^2)$
Test: $H: \mu \leqslant 1000$ gegen $A: \mu > 1000$

$n = 16$ Beobachtungen, $\bar{X} = 999$, $\sigma = 5$, $\alpha = 0{,}05$, $u_{1-\alpha} = 1{,}645$ (tabelliert)
Wegen

$$\bar{X} \leqslant 1000 + 1{,}645\cdot\frac{5}{\sqrt{16}} = 1002{,}06$$

wird H_0 angenommen.

[3] Subjekt-Prädikat-Objekt?

Verfahren 12.9 (Einstichproben-t-Test)
Wie Einstichproben-Gaußtest, jedoch σ^2 unbekannt

$\hookrightarrow$ *mathematische Statistik: Durchführung wie Gaußtest, jedocht σ durch S ersetzen und $u_{1-\alpha}$ durch $t_{1-\alpha}$, das $1-\alpha$-Quantil der t-Verteilung.*

Verfahren 12.10 (Zwei-Stichproben-Gaußtest)
Situation:

$$\left.\begin{array}{rl} X_1,\ldots,X_m & iid \sim \mathcal{N}(\mu,\sigma^2) \\ Y_1,\ldots,Y_n & iid \sim \mathcal{N}(\nu,\sigma^2) \end{array}\right\} \text{ stochastisch unabhängig}$$

μ,ν unbekannt, σ^2 bekannt

Test:
$$\begin{array}{llllll} H: & \mu = \nu & \text{gegen} & A: & \mu \neq \nu & \text{zum Niveau } \alpha \text{ bzw.} \\ H: & \mu-\nu = 0 & \text{gegen} & A: & \mu-\nu \neq 0 & \text{zum Niveau } \alpha \end{array}$$

Betrachte:

$$\begin{aligned} T &:= \bar{X} - \bar{Y} \\ VarT &= Var\bar{X} + Var\bar{Y} \\ \Rightarrow \quad T &= \mathcal{N}\left(\mu-\nu, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right) \\ \Rightarrow \quad \widetilde{T} &:= \frac{T}{\sigma \cdot \sqrt{\frac{m+n}{m\cdot n}}} \sim \begin{cases} \mathcal{N}(0,1) & \text{, bei } H \\ \mathcal{N}(\ldots,1) & \text{, bei } A \end{cases} \end{aligned}$$

Daraus folgt:

H ablehnen, falls $\widetilde{T} > u_{1-\alpha}$
vergleiche Einstichproben-Gaußtest mit Stichprobe vom Umfang 1 und $\mu_0 = 0$

Bemerkung:

- Zwei-(Mehr-)Stichproben-t-Test
- Vergleich von Varianzen
  $\hookrightarrow$ Varianzanalyse
- Tests ohne konkrete Verteilungsannahme
  $\hookrightarrow$ Nicht-parametrische Statistik
- Überprüfen der Voraussetzungen
  - Test auf Normalverteilung
  - Test auf stochastisch unabhängig

THE END