Universität Regensburg

Einführung in die Wahrscheinlichkeitsrechnung und Statistik

Vorlesung von Dr. Martin Kreuzer im SS 1998

geTEXt von

Alexander Zellner

Inhaltsverzeichnis

1 Kombinatorik 3

2 Laplace-Wahrscheinlichkeiten 7

3 Bedingte Wahrscheinlichkeiten 12

4 Klassische Wahrscheinlichkeitsverteilungen 18

5 Diskrete Wahrscheinlichkeitsräume 24

6 Diskrete Zufallsvariable 29

7 Erwartungswerte 33

8 Standardabweichung 38

9 Erzeugende Funktionen 45

10 Allgemeine Wahrscheinlichkeitsräume 49

11 Stetige Wahrscheinlichkeitsverteilungen 54

12 Stetige Zufallsvariable 65

13 Grenzwertsätze 77

14 Schätztheorie 88

15 Testtheorie 96

Übungsaufgaben

Blatt 1: Kombinatorik 105

Blatt 2: Bedingte Wahrscheinlichkeiten 107

Blatt 3: Klassische Wahrscheinlichkeitsverteilungen 109

Blatt 4: Diskrete Wahrscheinlichkeitsräume 112

Blatt 5: Erwartungswerte und Standardabweichungen **114**

Blatt 6: Erzeugende Funktionen **116**

Blatt 7: Stetige Wahrscheinlichkeitsverteilungen **119**

Blatt 8: Stetige Zufallsvariable **122**

Blatt 9: Erwartungswerte stetiger Zufallsvariablen **124**

Blatt 10: Grenzwertsätze **126**

Blatt 11: Schätztheorie **128**

Blatt 12: Testtheorie **130**

1 Kombinatorik

Definition 1.1

1. Für $n \in \mathbb{N}_+$ schreiben wir $n! := 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$ und nennen es $\boldsymbol{n}$ ***Fakultät***.

   Wir setzen $0! := 1$.

2. Sei $n \in \mathbb{N}_+$ und M eine Menge mit n Elementen.

   Eine bijektive Abbildung $\psi : M \to M$ heißt eine ***Permutation von*** $\boldsymbol{M}$.

   Im Fall $M = \{1, \ldots, n\}$ schreiben wir auch $\psi = \begin{pmatrix} 1 & 2 & \ldots & n \\ \psi(1) & \psi(2) & \ldots & \psi(n) \end{pmatrix}$.

Satz 1.2 *Ist $n \in \mathbb{N}$ und M eine Menge mit n Elementen, so gibt es genau $n!$ Permutationen von M.*

Beweis:
Sei $p(n)$ die Anzahl der Permutationen von M. Zuerst zeigen wir, dass $p(n)$ auch die Anzahl der Permutationen von $\{1, 2, \ldots, n\}$ ist.
Da M genau n Elemente hat, können wir eine bijektive Abbildung $\varphi : M \to \{1, 2, \ldots, n\}$ wählen.
Für jede Permutation von ψ von M ist die Abbildung $\varphi \circ \psi \circ \varphi^{-1}$:
$\{1, \ldots, n\} \xrightarrow{\varphi^{-1}} M \xrightarrow{\psi} M \xrightarrow{\varphi} \{1, \ldots, n\}$ eine Permutation von $\{1, \ldots, n\}$. Bei der Zuordnung $\psi \mapsto \varphi \circ \psi \circ \varphi^{-1}$ entsprechen sich die Permutationen von M und $\{1, \ldots, n\}$ eineindeutig.
Also ist $p(n)$ auch die Zahl der Permutationen von $\{1, \ldots, n\}$.
Mit vollständiger Induktion nach n zeigen wir nun $p(n) = n!$.
Induktionsbeginn n=0:
Die Menge $M = \emptyset$ besitzt genau eine Permutation, nämlich $\psi = id_M : \emptyset \to \emptyset$.
Induktionsschluss von $n-1$ auf n für $n > 0$:
Für den Wert $\psi(n)$ gibt es n Möglichkeiten.
Für ein festes $m = \psi(n)$ können wir eine bijektive Abbildung $\varphi : \{1, \ldots, n\} \setminus \{\psi(n)\} \to \{1, \ldots, n-1\}$ wählen.
Bei der Zuordnung $\psi \mapsto \varphi \circ \psi|_{\{1,\ldots,n-1\}}$ entsprechen sich die Permutationen ψ von M mit $\psi(n) = m$ und die Permutationen von $\{1, \ldots, n-1\}$ eineindeutig.
Also gibt es $p(n-1) = (n-1)!$ solcher Permutationen und insgesamt gilt $p(n) = n \cdot p(n-1) = n \cdot (n-1)! = n!$. $\square$

Beispiel 1.3 Sei $n \in \mathbb{N}_+$.

1. Hat man n verschiedene Objekte, so gibt es $n!$ Möglichkeiten, diese (in einer Reihe) anzuordnen.

2. Enthält eine Urne n verschiedene Kugeln und zieht man diese nacheinander ohne Zurücklegen, so gibt es genau $n!$ mögliche Versuchsergebnisse.

Definition 1.4 Seien $m, n \in \mathbb{N}$. Dann heißt die Zahl

$\binom{m}{n} = \begin{cases} \frac{m!}{(m-n)!n!} & \text{falls} \quad m \geq n \\ 0 & \text{falls} \quad m < n \end{cases}$ der ***Binomialkoeffizient „m über n" oder „n aus m"***.

Satz 1.5 *Seien $m, n \in \mathbb{N}$ und sei M eine Menge mit m Elementen. Dann gibt es genau $\binom{m}{n}$ Teilmengen von M, die n Elemente besitzen.*

Beweis:

Da die Behauptung im Fall $m \leq n$ klar ist, können wir o.E. $m > n$ annehmen.
Auch für $n = 0$ ist die Behauptung offensichtlich richtig. Sei also $1 \leq n \leq m$.
Sei Σ die Menge aller Tupel $(x_1, \ldots, x_n) \in M^n = \underbrace{M \times \ldots \times M}_{n\text{-mal}}$ mit $x_i \neq x_j$ für $i \neq j$.
Für die Wahl von x_1 gibt es m Möglichkeiten.
Für die Wahl von x_2 in $M \setminus \{x_1\}$ gibt es dann $n - 1$ Möglichkeiten, etc. .
Induktiv ergibt sich, dass Σ genau $m \cdot (m-1) \cdot \ldots \cdot (m-n+1)$ Elemente besitzt.
Jedes Tupel $(x_1, \ldots, x_n) \in \Sigma$ liefert eine Teilmenge $N = \{x_1, \ldots, x_n\}$.
Zwei Tupel liefern genau dann die gleiche Teilmenge $N = \{x_1, \ldots, x_n\} = \{y_1, \ldots, y_n\}$, wenn sich $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in \Sigma$ nur durch eine Permutation unterscheiden.
Also liefern je $n!$ Tupel dieselbe Menge N und es gibt $\frac{m(m-1)\cdot\ldots\cdot(m-n+1)}{n!} = \frac{m!}{n!(m-n)!} = \binom{m}{n}$ verschiedene Teilmengen $N \subset M$ mit n Elementen. $\square$

Korollar 1.6 *Seien $m, n \in \mathbb{N}_+$ und sei $(M, <)$ eine vollständig geordnete Menge (d.h. je zwei Elemente sind vergleichbar), die aus m Elementen besteht.*

1. *Es gibt genau $\binom{m}{n}$ streng monoton steigende Folgen $x_1 < x_2 < \ldots < x_n$ in M.*
2. *Es gibt genau $\binom{m+n-1}{n}$ monoton steigende Folgen $x_1 \leq x_2 \leq \ldots \leq x_n$ in M.*

Beweis:

1. Jede n-elementige Teilmenge $N = \{x_1, \ldots, x_n\} \subset M$ kann auf genau eine Art so geordnet werden, dass $x_1 < x_2 < \ldots < x_n$ gilt.
   Umgekehrt liefert jede Folge $x_1 < \ldots < x_n$ in M eine n-elementige Teilmenge $N = \{x_1, \ldots, x_n\}$.
   Also gibt es genau soviele Folgen $x_1 < \ldots < x_n$ in M, wie es n-elementige Teilmengen von M gibt.
2. Wir ordnen die Elemente von $M = \{x_1, \ldots, x_m\}$ so dass $x_1 < \ldots < x_m$ gilt. Dann ist jede bijektive Abbildung $M \to \{1, \ldots, m\}$ mit der Relation „$<$" verträglich und wir brauchen die Behauptung nur im Fall $M = \{1, \ldots, n\}$ zu zeigen.
   Jetzt entspricht jede monotone Folge $y_1 \leq \ldots \leq y_n$ in M eineindeutig einer

streng monotonen Folge $y_1 < y_2 + 1 < y_3 + 2 < \ldots < y_n + (n-1)$ in $\{1, \ldots, m+n-1\}$.
Die Behauptung folgt nun aus 1. .

$\square$

Definition 1.7 Sei $r \geq 1$ und seien $m, n_1, \ldots, n_r \in \mathbb{N}$ mit $n_1 + \ldots + n_r = m$. Dann heißt die Zahl $\binom{m}{n_1, \ldots, n_r} := \frac{m!}{n_1! \cdot \ldots \cdot n_r!}$ der ***Multinomialkoeffizient „m über $n_1, \ldots, n_r$“***.
Offenbar gilt: $\binom{m}{n_1} = \binom{m}{n_1, m-n_1}$.

Satz 1.8 *Sei $r \geq 1$ und seien $m, n_1, \ldots, n_r \in \mathbb{N}$ mit $n_1 + \ldots + n_r = m$.*
Eine Menge M besitze m Elemente.

1. *Es gibt genau $\binom{m}{n_1, \ldots, n_r}$ Möglichkeiten, die Menge M so in paarweise disjunkte Teilmengen $N_1, \ldots, N_r$ zu unterteilen, dass N_i für jedes $i \in \{1, \ldots, r\}$ genau n_i Elemente besitzt.*
2. *Unterteilt man $M = N_1 \dot\cup \ldots \dot\cup N_r$ wie in 1., und betrachtet man die Elemente jeder Teilmenge N_i als gleich, so kann man die Elemente von M auf genau $\binom{m}{n_1, \ldots, n_r}$ viele verschiedene Arten anordnen.*

Beweis:

1. Für die Auswahl der Teilmenge N_1 gibt es $\binom{m}{n_1}$ Möglichkeiten.
   Für die Auswahl $N_2 \subset M \setminus N_1$ gibt es dann $\binom{m-n_1}{n_2}$ Möglichkeiten, etc.
   Schließlich gibt es $\binom{m-n_1-\ldots-n_{r-1}}{n_r} = \binom{n_r}{n_r} = 1$ Möglichkeit für die Auswahl von $N_r \subset M \setminus (N_1 \cup \ldots \cup N_{r-1})$.
   Insgesamt erhalten wir $\binom{m}{n_1} \cdot \binom{m-n_1}{n_2} \cdot \ldots \cdot \binom{m-n_1-\ldots-n_{r-1}}{n_r} =$
   $= \frac{m!}{n_1!(m-n_1)!} \cdot \frac{(m-n_1)!}{n_2!(m-n_1-n_2)!} \cdot \ldots \cdot \frac{(m-n_1-\ldots-n_r)}{n_r!0!} =$
   $= \frac{m!}{n_1! \cdot \ldots \cdot n_r!} = \binom{m}{n_1, \ldots, n_r}$ Möglichkeiten für die gewünschte Unterteilung.
2. Von den $m!$ Anordnungen der Menge M sind $n_1! \cdot \ldots \cdot n_r!$ als gleich zu betrachten, da sie sich nur um Permutationen von $N_1, \ldots, N_r$ unterscheiden.
   Also gibt es $\frac{m!}{n_1! \cdot \ldots \cdot n_r!} = \binom{m}{n_1, \ldots, n_r}$ viele verschiedene Anordnungen.

$\square$

Beispiel 1.9 Befinden sich in einer Urne m Kugeln, und zwar n_1 von der Farbe 1, n_2 von der Farbe 2, ... , n_r von der Farbe r (mit $m = n_1 + \ldots + n_r$) und zieht man diese nacheinander ohne Zurücklegen, so gibt es $\binom{m}{n_1, \ldots, n_r}$ viele mögliche Versuchsergebnisse.

Notation: M Menge, $\mathfrak{P}(M)$ =Potenzmenge von M=Menge aller Teilmengen von M

Satz 1.10 *(Siebformel)*

Seien M eine Menge, $r, n \in \mathbb{N}_+$ und $N_1, \ldots, N_r \subset M$ Teilmengen.

Ferner sei eine Funktion $f : \mathfrak{P}(M) \to \mathbb{R}$ gegeben mit $f(A \cup B) = f(A) + f(B)$ für disjunkte Teilmengen $A, B \subset M$.

Schließlich sei

$S_i = \{x \in M \mid x \in N_i$ für genau n Indizes $1 \leq i \leq r\}$.

Dann gilt:

$$f(S_n) = \sum_{i=n}^{r} (-1)^{i-n} \binom{i}{n} \sum_{\substack{\{j_1,\ldots,j_i\} \subset \{1,\ldots,r\} \\ i\text{-elementige Teilmenge}}} f(N_{j_1} \cap \ldots \cap N_{j_i})$$

Beweis:

Beh. 1: $B, A_1, \ldots, A_n \subset M$, $\overline{A}_i := M \setminus A_i$ für $i = 1, \ldots, n$. Dann ist

$$f(B \cap \overline{A}_1 \cap \ldots \cap \overline{A}_n) = \sum_{i=0}^{n} (-1)^i \sum_{\substack{\{j_1,\ldots,j_i\} \subset \{1,\ldots,n\} \\ i\text{-elementige Teilmenge}}} f(B \cap A_{j_1} \cap \ldots \cap A_{j_i})$$

Beweis durch Induktion nach n.

$n = 1$: $f(B \cap \overline{A}_1) = f(B) - f(B \cap A_1)$

$n - 1 \to n$:

$$f(B \cap \overline{A}_1 \cap \ldots \cap \overline{A}_n) \overset{\text{Fall } n=1}{=}$$

$$f(B \cap \overline{A}_1 \cap \ldots \cap \overline{A}_{n-1}) - f(B \cap A_n \cap \overline{A}_1 \cap \ldots \cap \overline{A}_{n-1}) \overset{\text{I.V.}}{=}$$

$$\sum_{i=0}^{n-1} (-1)^i \sum_{\substack{\{j_1,\ldots,j_i\} \subset \{1,\ldots,n-1\} \\ i\text{-elementige Teilmenge}}} f(B \cap A_{j_1} \cap \ldots \cap A_{j_i}) -$$

$$\sum_{i=0}^{n-1} (-1)^i \sum_{\substack{\{j_1,\ldots,j_i\} \subset \{1,\ldots,n\} \\ i\text{-elementige Teilmenge}}} f(B \cap A_{j_1} \cap \ldots \cap A_{j_i} \cap A_n) =$$

$$\sum_{i=0}^{n-1} (-1)^i \sum_{\substack{\{j_1,\ldots,j_i\} \subset \{1,\ldots,n-1\} \\ i\text{-elementige Teilmenge}}} f(B \cap A_{j_1} \cap \ldots \cap A_{j_i}) -$$

$$\sum_{i=1}^{n} (-1)^{i+1} \sum_{\substack{\{j_1,\ldots,j_{i-1}\} \subset \{1,\ldots,n\} \\ (i-1)\text{-elementige Teilmenge}}} f(B \cap A_{j_1} \cap \ldots \cap A_{j_{i-1}} \cap A_n) =$$

$$\sum_{i=0}^{n-1} (-1)^i \sum_{\substack{\{j_1,\ldots,j_i\} \subset \{1,\ldots,n\} \\ i\text{-elem. Teilm., die } n \text{ nicht enthält}}} f(B \cap A_{j_1} \cap \ldots \cap A_{j_i}) +$$

$$\sum_{i=1}^{n} (-1)^i \sum_{\substack{\{j_1,\ldots,j_i\} \subset \{1,\ldots,n\} \\ i\text{-elem. Teilm., die } n \text{ enthält}}} f(B \cap A_{j_1} \cap \ldots \cap A_{j_i}) =$$

$$\sum_{i=0}^{n}(-1)^i \sum_{\substack{\{j_1,\dots,j_i\}\subset\{1,\dots,n\}\\ i\text{-elem. Teilm., die } n \text{ nicht enthält}}} f(B\cap A_{j_1}\cap\ldots\cap A_{j_i})+$$

$$\sum_{i=0}^{n}(-1)^i \sum_{\substack{\{j_1,\dots,j_i\}\subset\{1,\dots,n\}\\ i\text{-elem. Teilm., die } n \text{ enthält}}} f(B\cap A_{j_1}\cap\ldots\cap A_{j_i}) =$$

$$\sum_{i=0}^{n}(-1)^i \sum_{\substack{\{j_1,\dots,j_i\}\subset\{1,\dots,n\}\\ i\text{-elementige Teilmenge}}} f(B\cap A_{j_1}\cap\ldots\cap A_{j_i})$$

Beh. 2: $S_n = \dot{\bigcup}_{\substack{\{j_1,\dots,j_n\}\subset\{1,\dots,r\}\\ n\text{-elementige Teilmenge}}} N_{j_1}\cap\ldots\cap N_{j_n}\cap\overline{N}_{j_{n+1}}\cap\ldots\cap\overline{N}_{j_r}$ mit $\{j_{n+1},\ldots,j_r\} := \{1,\ldots,r\}\setminus\{j_1,\ldots,j_n\}$

Beh. 3: $f(S_n) = \sum_{i=n}^{r}(-1)^{i-n}\binom{i}{n} \sum_{\substack{j_1,\dots,j_i\subset\{1,\dots,r\}\\ i\text{-elementige Teilmenge}}} f(N_{j_1}\cap\ldots\cap N_{j_i})$

Beweis:

$$f(S_n) \overset{\text{Beh.2}}{=} \sum_{\substack{\{j_1,\dots,j_n\}\subset\{1,\dots,r\}\\ n\text{-elementige Teilmenge}}} f(N_{j_1}\cap\ldots\cap N_{j_n}\cap\overline{N}_{j_{n+1}}\cap\ldots\cap\overline{N}_{j_r}) \overset{\text{Beh.1}}{=}$$

$$\sum_{\substack{\{j_1,\dots,j_n\}\subset\{1,\dots,r\}\\ n\text{-elementige Teilm.}}} \sum_{i=n}^{r}(-1)^{i-n} \sum_{\substack{\{k_1,\dots,k_{i-n}\}\subset\{j_{n+1},\dots,j_r\}\\ (i-n)\text{-elem. Teilm.}}} f(N_{j_1}\cap\ldots\cap N_{j_n}\cap N_{k_1}\cap\ldots\cap N_{k_{i-n}}) =$$

$$\sum_{i=n}^{r}(-1)^{i-n} \sum_{\substack{\{j_1,\dots,j_n\}\subset\{1,\dots,r\}\\ n\text{-elementige Teilm.}}} \sum_{\substack{\{k_1,\dots,k_{i-n}\}\subset\{j_{n+1},\dots,j_r\}\\ (i-n)\text{-elem. Teilm.}}} f(N_{j_1}\cap\ldots\cap N_{j_n}\cap N_{k_1}\cap\ldots\cap N_{k_{i-n}}) =$$

$$\sum_{i=n}^{r}(-1)^{i-n}\binom{i}{n} \sum_{\substack{\{l_1,\dots,l_i\}\subset\{1,\dots,r\}\\ i\text{-elem. Teilm.}}} f(N_{l_1}\cap\ldots\cap N_{l_i}) \qquad \square$$

Beispiel 1.11 Sei M eine endliche Menge und seien $N_1, N_2, N_3 \subset M$.

Dann gilt:

$\#S_1 = \#N_1 + \#N_2 + \#N_3 - 2\#(N_1\cap N_2) - 2\#(N_1\cap N_3) - 2\#(N_2\cap N_3) + 3\#(N_1\cap N_2\cap N_3)$

2 Laplace-Wahrscheinlichkeiten

Definition 2.1

1. Ein ***endlicher Wahrscheinlichkeitsraum*** (Ω, P) besteht aus einer endlichen Menge Ω und einer Abbildung $P : \mathfrak{P}(\Omega) \to [0,1]$ von der Potenzmenge $\mathfrak{P}(\Omega)$ in die Menge $[0,1] := \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$, so dass die folgenden Rechenregeln gelten:
   1. $P(\Omega) = 1$ (Normiertheit)

2. $P(A \cup B) = P(A) + P(B)$ für disjunkte Teilmengen $A, B \in \mathfrak{P}(\Omega)$ (Additivität)

2. Ist (Ω, P) ein endlicher Wahrscheinlichkeitsraum, so verwenden wir die folgenden Sprechweisen:
$\Omega =$ ***Ergebnisraum / Stichprobenraum=Menge aller möglichen Versuchsergebnisse***
$A \subset \Omega$ heißt ein ***Ereignis***.
$\mathfrak{P}(\Omega) =$ ***Menge aller Ereignisse***
$P =$ ***Wahrscheinlichkeitsverteilung*** oder ***Wahrscheinlichkeitsmaß***

3. Ein endlicher Wahrscheinlichkeitsraum heißt
Laplace-Wahrscheinlichkeitsraum, wenn P gegeben ist durch $P(\{\omega\}) = \frac{1}{\#\Omega}$ für alle $\omega \in \Omega$.
In diesem Fall heißt P die ***Gleichverteilung auf Ω***.
Für $A \subset \Omega$ heißt $P(A) = \frac{\#A}{\#\Omega}$ die ***Laplace-Wahrscheinlichkeit*** des Ereignisses A.

Beispiel 2.2 (Das Würfeln)

1. Das Werfen eines Würfels wird beschrieben durch den folgenden Laplace-Wahrscheinlichkeitsraum:
$\Omega = \{1, 2, 3, 4, 5, 6\}$
$P : \mathfrak{P}(\Omega) \to [0, 1]$ Gleichverteilung
$A =$„es wurde eine gerade Zahl gewürfelt“$= \{2, 4, 6\}$
$P(A) = \frac{\#A}{\#\Omega} = \frac{3}{6} = 0,5$

2. Das gleichzeitige Werfen zweier Würfel wird beschrieben durch $\Omega = \{(i, j) \mid 1 \leq i, j \leq 6\} = \{1, \ldots, 6\}^2$ und $P : \mathfrak{P}(\Omega) \to [0, 1]$ Gleichverteilung. Für $i = 2, \ldots, 12$ sei A_i, das Ereignis „die Augensumme ist i“. Dann gilt:

i	2	3	4	5	6	7	8	9	10	11	12
$P(A_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Beispiel 2.3 (Die Glücksspirale 1971)

Jeder Teilnehmer (Los) erhält eine 7-stellige Losnummer.
Es wurden 70 Kugeln (je 7 mit den Ziffern $0, \ldots, 9$) in eine Urne gegeben, und dann werden gleichzeitig 7 Kugeln gezogen.
Dies ergibt die Losnummer des Gewinners.

Mathematische Beschreibung als Laplace-Experiment:
Mache alle Kugeln unterscheidbar (70 Farben $f_1, \dots, f_{70}$)
Ω =geordnete 7-er Stichproben aus $\{f_1, \dots, f_{70}\}$ ohne Zurücklegen=
$= \{(f_{i_1}, \dots, f_{i_7}) \mid i_1, \dots, i_7 \in \{1, \dots, 70\}$ paarweise verschieden$\}$
$\#\Omega = 70 \cdot 69 \cdot 68 \cdot 67 \cdot 66 \cdot 65 \cdot 64$

1. A_1 =„Losnummer 1111111 gewinnt"
   $\#A_1 = 7!$, $P(A_1) = \frac{7!}{70 \cdot \ldots \cdot 64} \approx 0,83 \cdot 10^{-9}$
2. A_2 =„Losnummer 1234567 gewinnt"
   $\#A_2 = 7^7$, $P(A_2) = \frac{7^7}{70 \cdot \ldots \cdot 64} \approx 0,136 \cdot 10^{-6}$
   Es gilt $\frac{P(A_2)}{P(A_1)} \approx 163,4$.

Beispiel 2.4 In einem Raum befinden sich 45 Personen. Mit welcher Wahrscheinlichkeit haben mindestens 2 davon am selben Tag Geburtstag ?
Annahmen: Es gebe 365 Geburtstage und alle seien gleich wahrscheinlich.
$\Omega = \{1, \dots, 365\}^n$, $\#\Omega = 365^n$
A_n :=„mind. 2 Komponenenten eines Tupels $(x_1, \dots, x_n) \in \Omega$ sind gleich."
$\Omega \setminus A_n$ =„alle Tupel mit paarweise verschiedenen Komponenten"
$\#(\Omega \setminus A_n) = 365 \cdot \ldots \cdot (365 - n + 1)$
$P(A_n) = 1 - P(\Omega \setminus A_n) = 1 - \frac{365 \cdot \ldots \cdot (365-n+1)}{365^n}$
z.B. $P(A_{45}) = 0,94097590$

Beispiel 2.5 (Lotto 6 aus 49)
Versuchsablauf: vgl. jeden Samstag Abend im Fernsehen.
<u>Modell I:</u>
geordnete 6-er Stichproben aus $\{1, \dots, 49\}$ ohne Rücklegen.
$\Omega = \{(x_1, \dots, x_6) \in \{1, \dots, 49\}^6 \mid x_i \neq x_j \text{ für } i \neq j\}$
$\#\Omega = 49 \cdot 48 \cdot \ldots \cdot 44$
<u>Modell II:</u>
ungeordnete 6-er Stichproben aus $\{1, \dots, 49\}$ ohne Rücklegen
Ω =Menge aller 6-elementigen Teilmengen von $\{1, \dots, 49\}$
$\#\Omega = \binom{49}{6}$
Sei A das Ereignis „es wurde $\{1, 2, 3, 4, 5, 6\}$ gezogen"
Modell I: $P(A) = \frac{6!}{49 \cdot \ldots \cdot 44}$
Modell II: $P(A) = \frac{1}{\binom{49}{6}}$

Wahrscheinlichkeiten bestimmter Ereignisse:

$A_6 :=$„6 Richtige“ $P(A_6) = \frac{1}{\binom{49}{6}} \approx 7,1511 \cdot 10^{-8}$

$A_5^* :=$„5 Richtige mit ZZ“ $\#(A_5^*) = \binom{6}{5}$ $P(A_5^*) \approx 0,429 \cdot 10^{-6}$

$A_5 =$„5 Richtige ohne ZZ“ $\#(A_5) = \binom{6}{5} \cdot 42$ $P(A_5) \approx 0,180 \cdot 10^{-4}$

$A_4 :=$„4 Richtige“ $\#(A_4) = \binom{6}{4}\binom{42}{2}$ $P(A_4) \approx 0,924 \cdot 10^{-3}$

$A_5 :=$„3 Richtige“ $\#(A_3) = \binom{6}{3}\binom{43}{3}$ $P(A_3) \approx 0,01765$

Satz 2.6 *(Die hypergeometrische Verteilung)*

Eine Urne enthalte $n \in \mathbb{N}_+$ Kugeln, von denen m schwarz und $n-m$ weiß sind. Wir ziehen l Kugeln ohne Zurücklegen, wobei $1 \leq l \leq n$ gelte.

Für $0 \leq k \leq \min\{m,l\}$ ist die Wahrscheinlichkeit des Ereignisses $A_k =$„es wurden genau k schwarze Kugeln gezogen“ gegeben durch $P(A_k) = \frac{\binom{m}{k}\binom{n-m}{l-k}}{\binom{n}{l}}$.

Beweis:
Wir betrachten die Kugeln als unterscheidbar: $\{1, \ldots, n\}$
Der Ergebnisraum ist $\Omega = \{A \in \mathfrak{P}(\{1, \ldots, n\}) \mid \#A = l\}$ und es gilt $\#\Omega = \binom{n}{l}$.
Für ein Ergebnis in A_k müssen wir k schwarze Kugeln aus den m schwarzen Kugeln und genau $l-k$ weiße Kugeln aus den $n-m$ weißen auswählen.
$\Rightarrow \#(A_k) = \binom{m}{k}\binom{n-m}{l-k}$
$\Rightarrow P(A_k) = \frac{\binom{m}{k}\binom{n-m}{l-k}}{\binom{n}{l}}$ $\square$

Beispiel 2.7 (3 Asse beim Skat)

Beim Skat erhält jeder der 3 Spieler 10 Karten und 2 bleiben verdeckt (Skat).

Mit welcher Wahrscheinlichkeit hat ein Spieler genau 3 Asse ?

Es genügt, die Wahrscheinlichkeit zu bestimmen, dass Spieler i 3 Asse hat (Ereignis A_i)

$\Omega =$10-elementige Teilmengen von $\{1, \ldots, 32\}$

Es gibt 4 Asse und 28 Nicht-Asse.

Nach 2.6 gilt:
$P(A_i) = \frac{\binom{4}{3}\binom{28}{7}}{\binom{32}{10}} = \frac{66}{899} \approx 0,0734$
$P(A_1 \cup A_2 \cup A_3) = 3P(A_1) = \frac{198}{899} \approx 0,220$

Satz 2.8 *(Die Binomialverteilung)*

In der Situation von 2.6 ziehen wir l Kugeln mit Zurücklegen, wobei $l \in \mathbb{N}_+$ sei.

Für $k = 0, \ldots, l$ ist die Wahrscheinlichkeit des Ereignisses $A_k =$„es wurden

genau k schwarze Kugeln gezogen" gegeben durch
$P(A_k) = \binom{l}{k} \left(\frac{m}{n}\right)^k \left(\frac{n-m}{n}\right)^{l-k}$

Beweis:
Wir betrachten die Kugeln wieder als unterscheidbar: $\{1, \ldots, n\}$
Der Ergebnisraum ist $\Omega = \{1, \ldots, n\}^l$ und $\#\Omega = n^l$.
Für ein Ergebnis in A_k müssen wir aus den l Ziehungen diejenigen k auswählen, bei denen eine schwarze Kugeln gezogen wird ($\binom{l}{k}$ Möglichkeiten).
Bei den k Ziehungen schwarzer Kugeln müssen wir je eine der m schwarzen Kugeln auswählen (m^k Möglichkeiten) und bei den $l-k$ Ziehungen weißer Kuglen, je eine der $n-m$ weißen Kugeln ($(n-m)^{l-k}$ Möglichkeiten). Insgesamt folgt:
$\#(A_k) = \binom{l}{k} m^k (n-m)^{l-k}$ und
$P(A_k) = \binom{l}{k} \frac{m^k(n-m)^{l-k}}{n^l} = \binom{l}{k} \left(\frac{m}{n}\right)^k \left(\frac{n-m}{n}\right)^{l-k}$ $\square$

Satz 2.9 *Sei (Ω, P) ein endlicher Wahrscheinlichkeitsraum und seien $A_1, \ldots, A_m \in \mathfrak{P}(\Omega)$.*
Dann gilt: $P(A_1 \cup \ldots \cup A_m) = T_1 - T_2 + T_3 - \ldots + \cdots + (-1)^{m+1} T_m$, wobei
$$T_k = \sum_{\substack{\{j_1,\ldots,j_k\} \subset \{1,\ldots,m\} \\ k\text{-elementige Teilmenge}}} P(A_{j_1} \cap \ldots \cap A_{j_k}).$$

Beweis:
Vorbemerkung: Für $i \in \mathbb{N}_+$ beliebig gilt $\sum\limits_{j=0}^{i} (-1)^j \binom{i}{j} = 0$.
Beweis durch Induktion nach i.
$i = 1$: klar!
Induktionsschluss $i-1 \to i$.
$$\sum_{j=0}^{i} (-1)^j \binom{i}{j} = (-1)^i + 1 + \sum_{j=1}^{i-1} (-1)^j \binom{i}{j} \overset{\text{Pascalsches Dreieck}}{=}$$
$$(-1)^i + 1 + \sum_{j=1}^{i-1} (-1)^j \left(\binom{i-1}{j} + \binom{i-1}{j-1} \right) = (-1)^i + 1 + \sum_{j=1}^{i-1} (-1)^j \binom{i-1}{j-1} + \sum_{j=1}^{i-1} (-1)^j \binom{i-1}{j} =$$
$$(-1)^i - \sum_{j=0}^{i-2} (-1)^j \binom{i-1}{j} + \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \overset{\text{I.V}}{=} (-1)^i - \sum_{j=0}^{i-2} (-1)^j \binom{i-1}{j} =$$
$$- \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \overset{\text{I.V}}{=} 0$$
Wende die Siebformel 1.10 mit $f := P$ an:
$$P(A_1 \cup \ldots \cup A_m) = \sum_{j=1}^{m} P(S_j) = \sum_{j=1}^{m} \sum_{i=j}^{m} (-1)^{i-j} \binom{i}{j} T_i = \sum_{1 \le j \le i \le m} (-1)^{i-j} \binom{i}{j} T_i =$$
$$\sum_{i=1}^{m} \sum_{j=1}^{i} (-1)^{i-j} \binom{i}{j} T_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{i} (-1)^j \binom{i}{j} \right) T_i (-1)^i =$$

$$\sum_{i=1}^{m}\Big(-1+\underbrace{\sum_{j=0}^{i}(-1)^j\binom{i}{j}}_{=0 \text{ nach Vorbemerkung}}\Big)T_i(-1)^i=\sum_{i=1}^{m}(-1)^{i+1}T_i \qquad \Box$$

Beispiel 2.10 (2 Asse beim Skat) Mit welcher Wahrscheinlichkeit hat mindestens einer der 3 Spieler beim Skat genau 2 Asse ?

Ω =„alle Anordnungen der 32 Karten (Platz 1-10=Spieler 1, 11-20=Spieler 2, 21-30=Spieler 3, 31-32=Skat)“

A_i =„Spieler i hat genau 2 Asse (vgl. Beispiel 2.7)“

$P(A_i)=\frac{\binom{4}{2}\binom{28}{8}}{\binom{32}{10}}=\frac{2079}{7192}\approx 0,289$

Bestimmung von $P(A_i\cap A_j)$ (für $i\neq j$)

Für die Plätze der Asse gibt es $\binom{10}{2}^2$ Möglichkeiten.

Für die Anordnung der Asse gibt es 4! Möglichkeiten.

Für die Anordnung der restlichen Karten gibt es 28! Möglichkeiten.

$P(A_i\cap A_j)=\frac{\binom{10}{2}^2 4!28!}{32!}=\frac{405}{7192}\approx 0,056$

Es folgt:

$P(A_1\cup A_2\cup A_3)=\sum\limits_i P(A_i)-\sum\limits_{i\neq j}P(A_i\cap A_j)=\frac{5022}{7192}\approx 0,698$

3 Bedingte Wahrscheinlichkeiten

Beispiel 3.1 (Schafkopfen) Beim Schafkopfen („mit der Kurzen“$\cong$bayer. Tarockkarten ohne 7er und 8er) (4 Spieler à 6 Karten)

Hat ein Spieler 2 Unter („Wenzen“), was ist dann die Wahrscheinlichkeit dafür, dass die anderen beiden Unter „zusammenstehen“ (d.h. dass ein anderer Spieler die restlichen beiden Unter besitzt) ?

Ω =„18! Anordnungen der Karten von Spieler 2-4 (Spieler 2$\cong$Nr. 1-6 usw.)“

A_i = „Spieler i hat die restlichen 2 Unter“ ($i=2,3,4$)

$\#A_i=\binom{6}{2}2!16!$

$P(A_2\cup A_3\cup A_4)=3P(A_i)=3\cdot\frac{30}{17\cdot 18}\approx 0,294$

Definition 3.2 Sei (Ω,P) ein Wahrscheinlichkeitsraum. Seien $A,B\subset\Omega$ Ereignisse und es gelte $P(B)\neq 0$.

Dann heißt $P(A|B):=\frac{P(A\cap B)}{P(B)}$ die ***bedingte Wahrscheinlichkeit von A bei***

gegebenem B (oder ***die Wahrscheinlichkeit von A unter der Bedingung B***).

Beispiel 3.3 (Fortsetzung von 3.1)
$\Omega =$„24! Anordnungen aller Karten“,
$A_i =$„Spieler i hat genau 2 Unter.“ $(i = 2, 3, 4)$
$B =$„Spieler 1 hat genau 2 Unter.“
$\#B = \binom{6}{2}\binom{18}{2}4!20!$
$P(B) = \frac{15\cdot 9\cdot 17\cdot 24}{24\cdot 23\cdot 22\cdot 21} \approx 0,216$
$\#(A_i \cap B) = \binom{6}{2}\binom{6}{2}4!20!$
$A := A_2 \mathbin{\dot\cup} A_3 \mathbin{\dot\cup} A_4$
$\#(A \cap B) = \#(A_2 \cap B) + \#(A_3 \cap B) + \#(A_4 \cap B)$
$P(A \cap B) = 3\frac{\binom{6}{2}^2 4!20!}{24!} \approx 0,0635$
$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{45}{153} \approx 0,294$

Bemerkung 3.4 *Sei (Ω, P) ein Wahrscheinlichkeitsraum.*

1. *Sind $A, B \subset \Omega$ Ereignisse und $P(B) \neq 0$, so gilt $P(A \cap B) = P(B) \cdot P(A|B)$. Ist z.B. das Ereignis B zeitlich vor A beobachtbar, so ist $P(A|B)$ oft leichter zu bestimmen als $P(A \cap B)$.*
2. *Sind $A_1, \dots, A_n \subset \Omega$ Ereignisse mit $P(A_1 \cap \ldots \cap A_{n-1}) \neq 0$, so folgt mit vollständiger Induktion:*
   $P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1})$

Beispiel 3.5 Was ist die Wahrscheinlichkeit dafür, dass beim Schafkopf jeder Spieler genau ein Ober („Dame“) besitzt ?
$\Omega =$„24! Anordnungen der Karten“
$A_i :=$„Spieler i hat genau einen Ober.“
$P(A_1) = \frac{\binom{4}{1}\binom{20}{5}}{\binom{24}{6}}$
$P(A_2|A_1) = \frac{\binom{3}{1}\binom{15}{5}}{\binom{18}{6}}$ Es sind genau 3 Ober unter den restlichen 18 Karten.
$P(A_3|A_1 \cap A_2) = \frac{\binom{2}{1}\binom{10}{5}}{\binom{12}{6}}$
$P(A_4|A_1 \cap A_2 \cap A_3) = 1$
$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) = \frac{36\cdot 36}{23\cdot 22\cdot 21} \approx 0,122$

Satz 3.6 *Sei (Ω, P) ein Wahrscheinlichkeitsraum.*

1. *Ist $B \subset \Omega$ ein Ereignis mit $P(B) \neq 0$, so ist $\tilde{P} : \mathfrak{P}(\Omega) \to [0,1]$, def. durch $A \mapsto P(A|B)$, wieder ein Wahrscheinlichkeitsmaß auf Ω.*
2. *Sind $B_1, \dots, B_n \subset \Omega$ Ereignisse mit $\Omega = B_1 \dot\cup \dots \dot\cup B_n$, so gilt für jedes Ereignis $A \subset \Omega$ die Gleichung $P(A) = \sum\limits_{\{i|P(B_i)\neq 0\}} P(B_i) \cdot P(A|B_i)$.*
3. *(Formel von Bayes 1713)*
   Unter den Voraussetzungen von 2. gilt für jedes Ereignis $A \subset \Omega$ mit $P(A) \neq 0$ und für jeden Index $i \in \{1, \dots, n\}$ mit $P(B_i) \neq 0$ die Formel:
   $$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum\limits_{\{j|P(B_j)\neq 0\}} P(B_j) \cdot P(A|B_j)}$$
4. *Sind $B_1, \dots, B_n \subset \Omega$ paarweise disjunkte Ereignisse mit $P(B_i) \neq 0$ für $i = 1, \dots, n$ und ist $A \subset \Omega$ ein Ereignis mit $P(A|B_1) = P(A|B_2) = \dots = P(A|B_n)$, so gilt $P(A|B_1 \cup \dots \cup B_n) = P(A|B_i)$.*

Beweis:

1. Es gilt $\tilde{P}(\Omega) = P(\Omega|B) = \frac{P(\Omega\cap B)}{P(B)} = 1$
   Sind $A_1, A_2 \subset \Omega$ disjunkt, so gilt $\tilde{P}(A_1 \cup A_2) = \frac{P((A_1\cup A_2)\cap B)}{P(B)} = \frac{P(A_1\cap B)+P(A_2\cap B)}{P(B)} = \tilde{P}(A_1) + \tilde{P}(A_2)$.
2. Aus $A = (A \cap B_1) \dot\cup \dots \dot\cup (A \cap B_n)$ folgt:
   $P(A) = \sum\limits_{i=1}^{n} P(A \cap B_i) = \sum\limits_{\{i|P(B_i)\neq 0\}} P(B_i) \cdot P(A|B_i)$
3. folgt aus 2. und $P(B_i \cap A) = P(B_i) \cdot P(A|B_i)$
4. Das folgt aus $P(A|B_i) \cdot P(B_1 \cup \dots \cup B_n) = P(A|B_i) \cdot \sum\limits_{j=1}^{n} P(B_j) = \sum\limits_{j=1}^{n} P(A|B_j) \cdot P(B_j) = \sum\limits_{j=1}^{n} P(A \cap B_j) = P(A \cap (B_1 \cup \dots \cup B_n)) = P(A|B_1 \cup \dots \cup B_n) \cdot P(B_1 \cup \dots \cup B_n)$

$\square$

Beispiel 3.7 (Test für eine seltene Krankheit)
Eine Krankheit komme bei 0,5% der Bevölkerung vor.
Ein Test liefere bei 99% der Kranken eine Reaktion, aber auch bei 2% der Gesunden.
Mit welcher Wahrscheinlichkeit ist jemand, bei dem die Reaktion auftritt wirklich krank ?

$\Omega =$„Patient ist gesund oder krank“

$B_1 =$„Patient ist krank“

$B_2 =$„Patient ist gesund“

$\Omega = B_1 \cup B_2$

$A =$„Test liefert Reaktion.“

$P(B_1) = 0,005,\ P(B_2) = 0,995,\ P(A|B_1) = 0,99,\ P(A|B_2) = 0,02$

$P(B_1|A) = \frac{P(B_1)\cdot P(A|B_1)}{P(B_1)\cdot P(A|B_1)+P(B_2)\cdot P(A|B_2)} = \frac{0{,}005\cdot 0{,}99}{0{,}005\cdot 0{,}99+0{,}995\cdot 0{,}02} = \frac{495}{2485} \approx 0,2$

Von den Personen, bei denen eine Reaktion auftritt, sind nur 20% wirklich krank.

Definition 3.8 Sei (Ω, P) ein Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \subset \Omega$ heißen ***unabhängig***, wenn gilt $P(A \cap B) = P(A) \cdot P(B)$.

Beispiel 3.9 Die 21 Schüler einer Klasse machen einen Rechentest. Es ergibt sich folgende Tabelle:

	bestanden	nicht bestanden
Jungen	6	3
Mädchen	8	4

Sind die Ereignisse

$A =$„Testperson ist ein Mädchen“ und

$B =$„Test wurde bestanden“ unabhängig ?

$P(A) = \frac{12}{21} = \frac{4}{7},\ P(B) = \frac{14}{21} = \frac{2}{3}$

$P(A \cap B) = \frac{8}{21} = P(A) \cdot P(B)$

d.h. A und B sind unabhängig.

Satz 3.10 *(Charakterisierung unabhängiger Ereignisse)*

Sei (Ω, P) ein Wahrscheinlichkeitsraum und seien $A, B \subset \Omega$ Ereignisse.

Es gelte $0 < P(B) < 1$.

Dann sind die folgenden Bedingungen äquivalent:

1. *Die Ereignisse A und B sind unabhängig.*
2. *Es gilt $P(A|B) = P(A)$.*
3. *Es gilt $P(A|B) = P(A|\overline{B})$, wobei $\overline{B} = \Omega \setminus B$.*

Beweis:

1.$\Rightarrow$2.

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

2.$\Rightarrow$3.
$P(A|\overline{B}) = \frac{P(A\cap\overline{B})}{P(\overline{B})} = \frac{P(A)-P(A\cap B)}{1-P(B)} = \frac{P(A)-P(B)P(A|B)}{1-P(B)} = \frac{P(A|B)-P(B)P(A|B)}{1-P(B)} = \frac{P(A|B)(1-P(B))}{1-P(B)} = P(A|B)$
3.$\Rightarrow$1.
$P(A\cap B) = P(A) - P(A\cap\overline{B}) = P(A) - P(\overline{B})P(A|\overline{B}) =$
$P(A)-(1-P(B))P(A|B) = P(A)-(1-P(B))\frac{P(A\cap B)}{P(B)} = P(A)+P(A\cap B)-\frac{P(A\cap B)}{P(B)}$
Also folgt: $P(B)P(A\cap B) = P(A)P(B) + P(B)P(A\cap B) - P(A\cap B)$
$\Rightarrow P(A\cap B) = P(A)P(B)$ $\square$

Bemerkung: Man beachte, dass A und B unabhängig sein können, selbst wenn das Eintreten von A davon abhängt, ob B vorliegt.

Beispiel 3.11 Ein idealer Würfel wird zweimal nacheinander geworfen.

$\Omega = \{1,\dots,6\}^2$

$A =$„Die Summe der Augenzahlen ist gerade."

$B =$„Der zweite Wurf liefert eine gerade Augenzahl."

$\#\Omega = 36$, $\#A = 18$, $\#B = 18$, $\#(A\cap B) = 9$

$P(A) = P(B) = \frac{1}{2}$, $P(A\cap B) = \frac{1}{4} = P(A)P(B)$

$\Rightarrow A, B$ unabhängig

Definition 3.12 Sei (Ω, P) ein Wahrscheinlichkeitsraum, sei I eine Menge und für jedes $i\in I$ sei $A_i\subset\Omega$ ein Ereignis.

Die Ereignisse $\{A_i\}_{i\in I}$ heißen ***unabhängig***, wenn für alle endlichen Teilmengen $J\subset I$ gilt:

$P\left(\bigcup_{j\in J} A_j\right) = \prod_{j\in J} P(A_j)$ (Produktformel)

Beispiel 3.13 Eine ideale Münze werde dreimal hintereinander geworfen.

$\Omega = \{\text{Kopf,Zahl}\}^3$

$A :=$„Es wurde mindestens zweimal Kopf geworfen."

$B :=$„Der erste Wurf liefert Kopf."

$C :=$„Der zweite und der dritte Wurf liefern das gleiche Ergebnis."

$\#\Omega = 8$, $\#A = 4$, $\#B = 4$, $\#(A\cap B) = 3$, $\#(A\cap C) = 2$, $\#(B\cap C) = 2$, $\#(A\cap B\cap C) = 1$

$P(A\cap B\cap C) = P(A)P(B)P(C)$, aber $P(A\cap B) = \frac{3}{8} \neq \frac{1}{4} = P(A)P(B)$,

d.h. A, B, C sind nicht unabhängig.

Satz 3.14 *(Eigenschaften von Familien unabhängiger Ereignisse)*
Sei (Ω, P) *ein Wahrscheinlichkeitsraum,* I *eine Menge und* $\{A_i\}_{i\in I} \subset \mathfrak{P}(\Omega)$ *eine Familie von Ereignissen.*
Dann gilt:

1. *Genau dann ist* $\{A_i\}_{i\in I}$ *unabhängig, wenn* $\{A_j\}_{j\in J}$ *unabhängig ist für jede endliche Teilmenge* $J \subset I$.
2. *Ist* $\{A_i\}_{i\in I}$ *unabhängig und* $B \subset \Omega$ *mit* $P(B) \in \{0, 1\}$*, so ist* $\{A_i\}_{i\in I} \cup \{B\}$ *unabhängig.*
3. *Ist* $\{A_i\}_{i\in I}$ *unabhängig und* $B_i \in \{A_i, \Omega\backslash A_i, \emptyset, \Omega\}$ *für jedes* $i \in I$*, so ist* $\{B_i\}_{i\in I}$ *unabhängig.*
4. *Ist* I *endlich, so ist* $\{A_i\}_{i\in I}$ *genau dann unabhängig, wenn für jede Menge* $\{B_i\}_{i\in I}$ *mit* $B_i \in \{A_i, \Omega \setminus A_i\}$ *die Produktformel* $P\left(\bigcup_{i\in I} B_i\right) = \prod_{i\in I} P(B_i)$ *gilt.*

Beweis:

1. klar!
2. Ist $P(B) = 0$, so gilt $P\left(\bigcap_{j\in J} A_j \cap B\right) = 0$ für $J \subset I$ endlich.

   Dann gilt (*): $P\left(\bigcap_{j\in J} A_j \cap B\right) = \left(\prod_{j\in J} P(A_j)\right) P(B)$, da auf beiden Seiten null steht.
   Sei $P(B) = 1$. Es genügt Gleichung (*) für $J \subset I$ endlich zu beweisen.
   Aus $\bigcap_{j\in J} A_j = \left(\bigcap_{j\in J} A_j \cap B\right) \dot\cup \left(\bigcap_{j\in J} A_j \setminus B\right)$ und $\bigcap_{j\in J} A_j \setminus B \subset \Omega \setminus B$ folgt
   $P\left(\bigcap_{j\in J} A_j \cap B\right) = P\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} P(A_j) = \left(\prod_{j\in J} P(A_j)\right) P(B)$
3. Für endliche Teilmengen $J \subset I$ müssen wir die Produktformel $P\left(\bigcap_{j\in J} B_j\right) = \prod_{j\in J} P(B_j)$ beweisen.
   o.E. sei $J = \{1, \ldots, n\}$
   Wir schließen mit vollständiger Induktion nach $m = \#\{j \mid B_j = \Omega \setminus A_j\}$.
   $m = 0$: klar nach 2. und der Voraussetzung
   $m > 0$: o.E. sei $B_n = \Omega \setminus A_n$.
   Es gilt: $P\left(\bigcap_{j=1}^{n} B_j\right) = P\left(\bigcap_{j=1}^{n-1} B_j \cap (\Omega \setminus A_n)\right) =$

$$P\left(\bigcap_{j=1}^{n-1} B_j\right) - P\left(\bigcap_{j=1}^{n-1} B_j \cap A_n\right) \overset{\text{I.V.}}{=} \left(\prod_{j=1}^{n-1} P(B_j)\right) \underbrace{(1 - P(A_n))}_{P(B_n)} = \prod_{j=1}^{n} P(B_j)$$

4. Ist $\{A_i\}_{i\in I}$ unabhängig, so gilt $P\left(\bigcap_{i\in I} B_i\right) = \prod_{i\in I} P(B_i)$ nach 3.

   Wir zeigen $P\left(\bigcap_{j\in J} B_j\right) = \prod_{j\in J} P(B_j)$ für $J \subset I$ mit absteigender Induktion nach $\#J =: m$.
   o.E. sei $I = \{1, \ldots, n\}$
   $m = n$: gilt nach Voraussetzung
   $m < n$: o.E. sei $J = \{1, \ldots, m\}$.
   Nach Induktionsvoraussetzung gilt:
   $P\left(\bigcap_{j=1}^{m} B_j \cap B_{m+1}\right) = \left(\prod_{j=1}^{m} P(B_j)\right) P(B_{m+1})$, sowie
   $P\left(\bigcap_{j=1}^{m} B_j \cap (\Omega \setminus B_{m+1})\right) = \left(\prod_{j=1}^{m} P(B_j)\right) (1 - P(B_{m+1}))$
   Addieren der Gleichungen liefert $P\left(\bigcap_{j=1}^{m} B_j\right) = \prod_{j=1}^{m} P(B_j)$

□

4 Klassische Wahrscheinlichkeitsverteilungen

Notation: Ist (Ω, P) ein Wahrscheinlichkeitsraum und $\omega \in \Omega$, so ist $P(\omega) := P(\{\omega\})$.

Satz 4.1

Sei $n \in \mathbb{N}_+$ und seien $(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ Wahrscheinlichkeitsräume.
Dann ist die Abbildung $P : \mathfrak{P}(\Omega_1 \times \ldots \times \Omega_n) \to [0, 1]$, def. durch
$\omega := (\omega_1, \ldots, \omega_n) \mapsto P_1(\omega_1) \cdot \ldots \cdot P_n(\omega_n)$ und $A \mapsto \sum_{(\omega_1,\ldots,\omega_n)\in A} P(\omega)$, eine Wahrscheinlichkeitsverteilung auf $\Omega_1 \times \ldots \times \Omega_n$.

Beweis:
Es gilt
$P(\Omega_1 \times \ldots \times \Omega_n) = \sum_{(\omega_1,\ldots,\omega_n)\in\Omega_1\times\ldots\times\Omega_n} P(\omega) = \sum_{\omega_1\in\Omega_1} \ldots \sum_{\omega_n\in\Omega_n} P_1(\omega_1) \cdot \ldots \cdot P_n(\omega_n) =$
$\left(\sum_{\omega_1\in\Omega_1} P_1(\omega_1)\right) \cdot \ldots \cdot \left(\sum_{\omega_n\in\Omega_n} P_n(\omega_n)\right) = 1 \cdot \ldots \cdot 1 = 1$
Additivität gilt nach Definition. □

Definition 4.2

Sei $n \in \mathbb{N}_+$,seien $(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ Wahrscheinlichkeitsräume und seien $\Omega := \Omega_1 \times \ldots \times \Omega_1$, sowie $P : \mathfrak{P}(\Omega) \to [0,1]$, die in Satz 4.1 definierte Abbildung.

1. Der Wahrscheinlichkeitsraum (Ω, P) heißt das ***Produkt der Wahrscheinlichkeitsräume*** $\boldsymbol{(\Omega_1, P_1), \ldots, (\Omega_n, P_n)}$.

   Man sagt auch (Ω, P) sei ein ***Produktraum*** und P heißt die ***Produktverteilung***.

2. Ist $i \in \{1, \ldots, n\}$ und $A_i \subset \Omega_i$, so schreiben wir $\{X_i \in A_i\} := \Omega_1 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots \times \Omega_n$.

   Im Fall $A_i = \{\omega_i\}$ schreiben wir auch $\{X_i = \omega_i\} := \Omega_1 \times \ldots \times \Omega_{i-1} \times \{\omega_i\} \times \Omega_{i+1} \times \ldots \times \Omega_n$.

3. Ein Ereignis $A \subset \Omega$ ***hängt nur von den Koordinaten in*** $\boldsymbol{I = \{i_1, \ldots, i_d\} \subset \{1, \ldots, n\}}$ ***ab***, wenn es ein Ereignis $B \subset \Omega_{i_1} \times \ldots \times \Omega_{i_d}$ gibt, so dass $A = \{(\omega_1, \ldots, \omega_n) \in \Omega \mid (\omega_{i_1}, \ldots, \omega_{i_d}) \in B\}$.

Satz 4.3 *(Eigenschaften von Produkträumen)*

Sei $n \in \mathbb{N}_+$, seien $(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ Wahrscheinlichkeitsräume und sei (Ω, P) ihr Produkt.

1. *Ist $i \in \{1, \ldots, n\}$ und $A_i \subset \Omega_i$, so gilt $P(\{X_i \in A_i\}) = P_i(A_i)$.*
2. *Sind $A_i \subset \Omega_i$ Ereignisse für $i = 1, \ldots, n$, so gilt*
   $P\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = P(A_1 \times \ldots \times A_n) = \prod_{i=1}^{n} P_i(A_i) = \prod_{i=1}^{n} P(\{X_i \in A_i\})$.
3. *Ist $\{1, \ldots, n\} = I_1 \,\dot\cup\, \ldots \,\dot\cup\, I_m$ und ist $A_j \subset \Omega$ für jedes $j \in \{1, \ldots, m\}$ ein Ereignis, das nur von den Koordinaten in I_j abhängt, so sind die Ereignisse $A_1, \ldots, A_m$ unabhängig.*

Beweis:

1. $P(\{X_i \in A_i\}) = \sum\limits_{\omega_1 \in \Omega_1} \cdots \sum\limits_{\omega_{i-1} \in \Omega_{i-1}} \sum\limits_{\omega_i \in A_i} \sum\limits_{\omega_{i+1} \in \Omega_{i+1}} \cdots \sum\limits_{\omega_n \in \Omega_n} P_1(\omega_1) \cdot \ldots \cdot P_n(\omega_n) = \left(\sum\limits_{\omega_i \in A_i} P_i(\omega_i)\right) \prod\limits_{j \neq i} \left(\sum\limits_{\omega_j \in \Omega_J} P_j(\omega_j)\right) = P_i(A_i)$
2. $\bigcap\limits_{i=1}^{n} \{X_i \in A_i\} = A_1 \times \ldots \times A_n$ und $P(A_1 \times \ldots \times A_n) = \sum\limits_{\omega_1 \in A_1} \ldots \sum\limits_{\omega_n \in A_n} P_1(\omega_1) \cdot \ldots \cdot P_n(\omega_n) = \left(\sum\limits_{\omega_1 \in A_1} P_1(\omega_1)\right) \cdot \ldots \cdot \left(\sum\limits_{\omega_n \in A_n} P_n(\omega_n)\right) = P_1(A_1) \cdot \ldots \cdot P_n(A_n)$

   Zusammen mit 1. ergibt sich die Behauptung.

3. Sei o.E. $I_1 = \{1, \ldots, d_1\}$, $I_2 = \{d_1 + 1, \ldots, d_2\}$ usw. (kann man durch Umnummerieren der Koordinaten erreichen)
   Dann ist $A_i = \prod_{k=1}^{d_{i-1}} \Omega_k \times B_i \times \prod_{k=d_i+1}^{n} \Omega_k$ mit geeigneten $B_i \in \prod_{k \in I_i} \Omega_k$ und es gilt $A_1 \cap \ldots \cap A_n = B_1 \times \ldots \times B_m$.
   $P(A_1 \cap \ldots \cap A_n) = P(B_1 \times \ldots \times B_m) \overset{2.}{=} \prod_{j=1}^{m} P_{I_j}(B_j) \overset{2.}{=} \prod_{j=1}^{m} P(A_j)$, wobei P_{I_j} die Produktverteilung der Wahrscheinlichkeitsräume (Ω_l, P_l) mit $l \in I_j$ ist.
   Haben wir nur eine Teilmenge $\{A_j\}_{j \in J}$ mit $J \subsetneq \{1, \ldots, m\}$, so setzen wir $A_j := \Omega$ für $j \in \{1, \ldots, m\} \setminus J$ und es gilt
   $P\left(\bigcap_{j \in J} A_j\right) = P\left(\bigcap_{j \in J} A_j \cap \bigcap_{j \in \{1,\ldots,m\} \setminus J} A_j\right) = \prod_{j \in J} P(A_j) \prod_{j \in \{1,\ldots,m\} \setminus J} P(A_j) = \prod_{j \in J} P(A_j)$
   d.h. $\{A_i\}_{i \in \{1,\ldots,m\}}$ ist unabhängig.

□

Beispiel 4.4 Gegeben sie eine Zahl $p \in [0, 1]$ und ein $n \in \mathbb{N}_+$. Für $i = 1, \ldots, n$ sei (Ω_i, P_i) der Wahrscheinlichkeitsraum $\Omega_i = \{0, 1\}$ mit $P_i(1) = p$ und $P_i(0) = 1 - p$.
Dann heißt der Produktraum (Ω, P) auch der ***n-te Bernoulli-Raum*** (oder der ***Bernoulli-Raum vom Umfang n***) und P heißt die ***Bernoulli-Verteilung***.
Ist $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ und $k = \#\{i \mid w_i = 1\}$, so gilt $P(\omega) = p^k(1 - p)^{n-k}$

Beispiel 4.5 (Die Binomialverteilung)
Sei $n \in \mathbb{N}_+$ und $\tilde{\Omega} = \{0, 1\}^n$ der Bernoulli-Raum mit Bernoulli-Verteilung $\tilde{P}$.
Für $k = 0, \ldots, n$ sei $A_k := \{(\omega_1, \ldots, \omega_n) \in \tilde{\Omega} \mid \omega_i = 1$ für genau k $i \in \{1, \ldots, n\}\}$.
Dann gilt: $\tilde{P}(A_k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
Offenbar ist $\tilde{\Omega} = A_0 \,\dot\cup\, \ldots \,\dot\cup\, A_n$.
Betrachte nun $\Omega = \{0, \ldots, n\}$ und $P : \mathfrak{P}(\Omega) \to [0, 1]$, def. durch
$\{k\} \mapsto \binom{n}{k} p^k (1 - p)^{n-k}$.
Dann definiert $P =: b_{n,p}$ eine Wahrscheinlichkeitsverteilung auf Ω.
Sie heißt die ***Binomialverteilung mit Parametern n und p***.

Beispiel 4.6 Ein bestimmtes Medikament heilt eine gewisse Krankheit in 80% der Fälle. Es werde 10 Patienten verabreicht, die an dieser Krankheit leiden. Mit welcher Wahrscheinlichkeit werden mindestens 8 Patienten geheilt ?

Für $i = 1, \dots, 10$ sei $\Omega_i = \{0,1\}$ und $\{1\}$ =„Patient i werde geheilt.“ und $P_i(1) = 0,8$.
Sei $(\tilde{\Omega}, \tilde{P})$ der Produktraum und (Ω, P) der zugehörige binomialverteilte Wahrscheinlichkeitsraum.
Auf $(\Omega = \{1, \dots, 10\}, P = b_{10,0.8})$ gilt $\{k\}$ =„es werden genau k Patienten geheilt“ und die gesuchte Wahrscheinlichkeit ist
$P(8) + P(9) + P(10) = \binom{10}{8}(0,8)^8(0,2)^2 + \binom{10}{9}(0,8)^9(0,2) + \binom{10}{10}(0,8)^{10} \approx 0,678$

Beispiel 4.7 (Die Multinomialverteilung)
Seien $r, n \in \mathbb{N}_+$ und für $i = 1, \dots, n$ gelte $\Omega_i = \{1, \dots, r\}$, sowie $P_i : \mathfrak{P}(\Omega_i) \to [0,1]$, def. durch $\{j\} \mapsto p_j$ mit $p_1, \dots, p_r \geq 0$ und $p_1 + \ldots + p_r = 1$.
Sei $(\tilde{\Omega}, \tilde{P})$ der Produktraum.
Für $(\omega_1, \dots, \omega_n) \in \tilde{\Omega}$ gilt $\tilde{P}((\omega_1, \dots, \omega_n)) = p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}$ mit $k_i := \#\{j \mid w_i = j\}$ für $i = 1, \dots, r$.
Für $k_1, \dots, k_r \in \mathbb{N}$ mit $k_1 + \ldots + k_r = n$ betrachten wir
$A_{(k_1,\dots,k_r)} = \{(\omega_1, \dots, \omega_n) \in \tilde{\Omega} \mid \text{für } i = 1, \dots, r \text{ gilt } \#\{j \mid \omega_j = i\} = k_i\}$.
Nach der Definition der Multinomialkoeffizienten gilt $\tilde{P}\left(A_{(k_1,\dots,k_r)}\right) = \binom{n}{k_1,\dots,k_r} \cdot p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}$
Außerdem ist wiederum $\tilde{\Omega} = \dot{\bigcup}_{\substack{(k_1,\dots,k_r)\in\mathbb{N}^r \\ k_1+\ldots+k_r=n}} A_{(k_1,\dots,k_r)}$
Auf der Menge $\Omega = \{(k_1, \dots, k_r) \in \mathbb{N}^r \mid k_1 + \ldots + k_r = n\}$ erhalten wir durch $P : \mathfrak{P}(\Omega) \to [0,1]$, def. durch $\{(k_1, \dots, k_r)\} \mapsto \binom{n}{k_1,\dots,k_r} \cdot p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}$, damit eine Wahrscheinlichkeitsverteilung. Sie heißt ***Multinomialverteilung*** (oder ***Polynomialverteilung***) ***mit Parametern*** $\boldsymbol{n}$ ***und*** $\boldsymbol{p_1, \dots, p_r}$.

Beispiel 4.8 Mit welcher Wahrscheinlichkeit erhält man beim gleichzeitigen Werfen von 12 Würfeln jede Augenzahl genau 2 mal ?
Für $i = 1, \dots, 12$ sei $\Omega_i = \{1, \dots, 6\}$ und P_i =Gleichverteilung.
Ferner sei $(\tilde{\Omega}, \tilde{P})$ der Produktraum und $\Omega = \{(k_1, \dots, k_6) \in \mathbb{N}^6 \mid k_1 + \ldots + k_6 = 12\}$ der Wahrscheinlichkeitsraum mit Multinomialverteilung P.
Dann gilt:
$P((2,2,2,2,2,2)) = \binom{12}{2,2,2,2,2,2} \cdot \left(\frac{1}{6}\right)^2 \cdot \ldots \cdot \left(\frac{1}{6}\right)^2 = \frac{12!}{2^6 6^{12}} \approx 0,00344$

Beispiel 4.9 Eine Fussballmannschaft bestehe aus 4 Stürmern, 2 Mittelfeldspielern, 4 Verteidigern und 1 Torwart.

Man wählt aus 6 Mannschaften zufällig je einen Spieler aus.
$A :=$„2 Stürmer. 1 Mittelfeldspieler, 2 Verteidiger, 1 Torwart“
$B :=$„3 Stürmer, 3 Verteidiger“
Bestimme $P(A)$ und $P(B)$.
$n = 6$, $r = 4$
Für $i = 1, \dots, 6$ sei $\Omega_i = \{S, M, V, T\}$ und $P_i : \mathfrak{P}(\Omega_i) \to [0,1]$ sei def. durch $\{S\} \mapsto \frac{4}{11}$, $\{M\} \mapsto \frac{2}{11}$, $\{V\} \mapsto \frac{4}{11}$, $\{T\} \mapsto \frac{1}{11}$.
Ferner sei $(\tilde{\Omega}, \tilde{P})$ der Produktraum und (Ω, P) der zugehörige multinomialverteilte Wahrscheinlichkeitsraum.
Dann gilt $A = \{(2,1,2,1)\} \subset \Omega$ und $B = \{(3,0,3,0)\} \subset \Omega$ und
$P(A) = \binom{6}{2,1,2,1} \left(\frac{4}{11}\right)^2 \left(\frac{2}{11}\right) \left(\frac{4}{11}\right)^2 \left(\frac{1}{11}\right) \approx 0,0520$, sowie
$P(B) = \binom{6}{3,0,3,0} \left(\frac{4}{11}\right)^3 \left(\frac{4}{11}\right)^3 \approx 0,0462$

Beispiel 4.10 (Die geometrische Verteilung)
Sei $n \in \mathbb{N}_+$ und $\tilde{\Omega} = \{0,1\}^n$ der n-te Bernoulli-Raum mit Bernoulli-Verteilung $\tilde{P}$ zum Parameter p.
Für $k = 1, \dots, n$ sei $A_k = \{(\omega_1, \dots, \omega_n) \in \tilde{\Omega} \mid \omega_i = 0$ für $i = 1, \dots, k-1$ und $\omega_k = 1\}$, sowie $A_{n+1} = \{(0, \dots, 0)\}$.
Dann gilt $P(A_k) = (1-p)^{k-1}p$ für $k \leq n$, sowie $\tilde{\Omega} = \dot{\bigcup}_{k=1}^{n+1} A_k$
$P(A_{n+1}) = (1-p)^n$
Somit hat man auf $\Omega = \{1, \dots, n+1\}$ eine Wahrscheinlichkeitsverteilung $P : \mathfrak{P}(\Omega) \to [0,1]$, def. durch $\{k\} \mapsto \begin{cases} (1-p)^{k-1}p & \text{falls} \quad k \leq n \\ (1-p)^n & \text{falls} \quad k = n+1 \end{cases}$.
Sie heißt die ***n-te geometrische Verteilung***.

Beispiel 4.11 Ein Ü-Eiersammler hat bereits 9 der 10 aktuellen „TopTenTeddies“.
Wieviele Eier muss er durchschnittlich kaufen, wenn er bei 120 Inhalten die fehlende Figur mit 90% Wahrscheinlichkeit finden will ?
Sei n die Zahl der Eier, die er kaufen muss.
Für $i = 1, \dots, n$ gelte $\Omega_i = \{0,1\}$ mit $P_i(1) = \frac{1}{120}$.
Sei (Ω, P) der Wahrscheinlichkeitsraum mit der geometrischen Wahrscheinlichkeitsverteilung, der zu $\tilde{\Omega} := \Omega_1 \times \ldots \times \Omega_n$ gehört.
Dann gilt $P(k) = \left(\frac{119}{120}\right)^{k-1} \cdot \frac{1}{120}$ =die Wahrscheinlichkeit dafür, dass der Sammler

genau beim k-ten Ei fertig ist.
Es folgt: $P(1) + P(2) + \ldots + P(n) \overset{!}{\geq} 0,9 \Leftrightarrow$
$1 + \frac{119}{120} + \left(\frac{119}{120}\right)^2 + \ldots + \left(\frac{119}{120}\right)^{n-1} \geq 108 \Leftrightarrow$
$1 - \left(\frac{119}{120}\right)^n \geq \frac{108}{120} \Leftrightarrow$
$\left(\frac{119}{120}\right)^n \leq 0,1 \Leftrightarrow$
$n \geq \frac{\ln 0,1}{\ln \frac{119}{120}} \approx 275,16$
Antwort: Er muss mindestens 276 Eier kaufen, um mit 90% Wahrscheinlichkeit fertig zu sein.

Satz 4.12 *Sei $n \geq 2$, seien $(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ Wahrscheinlichkeitsräume und sei $\Omega = \Omega_1 \times \ldots \times \Omega_n$.*
Für $i = 2, \ldots, n$ sei $P_i(\omega_i \mid \omega_1, \ldots, \omega_{i-1})$ die Wahrscheinlichkeit dafür, dass im i-ten Versuch ω_i eintritt, wenn $\omega_1, \ldots \omega_{i-1}$ die Ergebnisse der Versuche $1, \ldots, i-1$ sind.

1. *Die Abbildung $P : \mathfrak{P}(\Omega) \to [0,1]$, def. durch $\{(\omega_1, \ldots, \omega_n)\} \mapsto P_1(\omega_1) \cdot P_2(\omega_2|\omega_1) \cdot \ldots \cdot P_n(\omega_n|\omega_1, \ldots, \omega_{n-1})$, definiert eine Wahrscheinlichkeitsverteilung auf Ω.*
2. *Für $(\eta_1, \ldots, \eta_n) \in \Omega$ gilt $P(\{X_1 = \eta_1\}) = P_1(\eta_1)$, sowie $P(\{X_i = \eta_i\} \mid \{X_1 = \eta_1\} \cap \ldots \cap \{X_{i-1} = \eta_{i-1}\}) = P_i(\eta_i \mid \eta_1, \ldots, \eta_{i-1})$ für $i = 2, \ldots, n$.*

Beweis:

1. $\sum\limits_{(\omega_1,\ldots,\omega_n)\in\Omega} P((\omega_1, \ldots, \omega_n)) = \sum\limits_{(\omega_1,\ldots,\omega_n)\in\Omega} P_1(\omega_1) \cdot \ldots \cdot P_n(\omega_n \mid \omega_1, \ldots, \omega_{n-1}) =$
$\left(\sum\limits_{\omega_1\in\Omega_1} \ldots \sum\limits_{\omega_{n-1}\in\Omega_{n-1}} P_1(\omega_1) \cdot \ldots \cdot P_{n-1}(\omega_{n-1} \mid \omega_1, \ldots, \omega_{n-2})\right) \cdot$
$\underbrace{\left(\sum\limits_{\omega_n\in\Omega_n} P_n(\omega_n \mid \omega_1, \ldots, \omega_{n-1})\right)}_{=1} = \ldots = \sum\limits_{\omega_1\in\Omega_1} P_1(\omega_1) = 1$
Additivität gilt nach Definition.
2. Für $i = 1, \ldots, n$ gilt:
$P(\{X_1 = \eta_1\} \cap \ldots \cap \{X_i = \eta_i\}) =$
$\sum\limits_{(\omega_{i+1},\ldots,\omega_n)\in\Omega_{i+1}\times\ldots\times\Omega_n} P((\eta_1, \ldots, \eta_i, \omega_{i+1}, \ldots, \omega_n)) =$
$\sum\limits_{\omega_{i+1}\in\Omega_{i+1}} \ldots \sum\limits_{\omega_n\in\Omega_n} P_1(\eta_1)P_2(\eta_2|\eta_1)\cdot\ldots\cdot P_i(\eta_i \mid \eta_1\cdot\ldots\cdot\eta_{i-1})\cdot P_{i+1}(\omega_{i+1} \mid \eta_1, \ldots, \eta_i)\cdot$

$\ldots \cdot P_n(\omega_n \mid \eta_1, \ldots, \eta_i, \omega_{i+1}, \ldots, \omega_{n-1}) =$
$\left(\sum\limits_{\omega_{i+1}\in\Omega_{i+1}} \cdots \sum\limits_{\omega_{n-1}\in\Omega_{n-1}} P_1(\eta_1) \cdot \ldots \cdot P_{n-1}(\omega_{n-1} \mid \eta_1, \ldots, \eta_i, \omega_{i+1}, \ldots, \omega_{n-2})\right) \cdot$
$\left(\sum\limits_{\omega_n\in\Omega_n} P_n(\omega_n \mid \eta_1, \ldots, \eta_i, \omega_{i+1}, \ldots, \omega_{n-1})\right) = \ldots =$
$P_1(\eta_1) \cdot \ldots \cdot P_i(\eta_i \mid \eta_1, \ldots, \eta_{i-1})$
Im Fall $i = 1$ liefert dies bereits die Behauptung.
Im Fall $i > 1$ gilt
$P(\{X_i = \eta_i\} \mid \{X_1 = \eta_1\} \cap \ldots \cap \{X_{i-1} = \eta_{i-1}\}) = \frac{P(\{X_1=\eta_1\}\cap\ldots\cap\{X_i=\eta_i\})}{P(\{X_1=\eta_1\}\cap\ldots\cap\{X_{i-1}=\eta_{i-1}\})} =$
$P_i(\eta_i \mid \eta_1, \ldots, \eta_{i-1})$

□

Definition 4.13
Sei $n \geq 2$, seien $(\Omega_1, P_1), \ldots, (\Omega_n, P_n)$ Wahrscheinlichkeitsräume, sei $\Omega = \Omega_1 \times \ldots \times \Omega_n$, und sei P die in 4.12 1. definierte Wahrscheinlichkeitsverteilung. Dann heißt (Ω, P) ein ***Stufenraum*** und P heißt eine ***bedingte Wahrscheinlichkeitsverteilung***.

Beispiel 4.14 (Urnenmodell von Polya)
Eine Urne enthält w weiße und s schwarze Kugeln.
Sei $c \in \mathbb{N}_+$. Nachdem eine Kugel gezogen wurde, wird sie und c weitere Kugeln ihrer Farbe zurückgelegt. Sukzessive wird so 4-mal gezogen.
Bestimme $P(w, s, w, w)$.
Verwende $\Omega = \Omega_1 \times \ldots \times \Omega_n$ mit $\Omega_i = \{w, s\}$.
Es gilt $P_i(w \mid \omega_1, \ldots, \omega_{i-1}) = \frac{w+\alpha c}{w+s+(i-1)c}$ mit $\alpha = \#\{j \in \{1, \ldots, i-1\} \mid \omega_j = w\}$, sowie $P_i(s \mid \omega_1, \ldots, \omega_{i-1}) = \frac{s+(i-1-\alpha)c}{w+s+(i-1)c}$.
Es ergibt sich:
$P((w, s, w, w)) = P_1(w) \cdot P_2(s|w) \cdot P_3(w|w, s) \cdot P_4(w|w, s, w) = \frac{w}{w+s} \cdot \frac{s}{w+s+c} \cdot \frac{w+c}{w+s+2c} \cdot \frac{w+2c}{w+s+3c}$

5 Diskrete Wahrscheinlichkeitsräume

Definition 5.1 Ein Paar (Ω, P) heißt ein ***diskreter Wahrscheinlichkeitsraum***, wenn Ω eine nichtleere höchstens abzählbar unendliche Menge ist und wenn $P : \mathfrak{P}(\Omega) \to [0, 1]$ eine Abbildung mit den folgenden Eigenschaften ist:
1. $P(\Omega) = 1$ (Normiertheit)

2. Ist $\{A_i\}_{i\in I} \subset \mathfrak{P}(\Omega)$ eine Familie paarweiser disjunkter Teilmengen von Ω, so ist $\sum\limits_{i=0}^{\infty} P(A_i)$ konvergent und es gilt:
$P\left(\dot{\bigcup\limits_{i\in\mathbb{N}}} A_i\right) = \sum\limits_{i=0}^{\infty} P(A_i)$ (Additivität)

Beispiel 5.2 Ein Bernoulli-Experiment werde unendlich oft wiederholt (mit Parameter $p \in]0,1]$).
Wir beobachten, bei welcher Wiederholung zum ersten Mal das Ergebnis 1 auftritt.
$\Rightarrow \Omega = \{1, 2, 3, \dots\} = \mathbb{N}_+$
Nach Beispiel 4.10 gilt $P(k) = (1-p)^{k-1}p$.
Die Abbildung $P : \mathfrak{P}(\Omega) \to [0,1]$, def. durch $\{k\} \mapsto (1-p)^{k-1}p$, definiert eine Wahrscheinlichkeitsverteilung auf Ω, denn
$\sum\limits_{k\geq 1}(1-p)^{k-1}p = p(1 + (1-p) + (1-p)^2 + \dots) = p\frac{1}{1-(1-p)} = 1$

Beispiel 5.3 (Die Pascal-Verteilung)
Sei $n, r \in \mathbb{N}_+$ und sei $p \in]0,1]$.
Ferner sei $(\tilde{\Omega}, \tilde{P})$ das Produkt $\tilde{\Omega} = \{0,1\}^n$ mit Bernoulli-Verteilung $\tilde{P}$.
Für $k = 0, \dots, n-r$ betrachte
$A_k = \{(\omega_1, \dots, \omega_n) \in \tilde{\Omega} \mid$ genau k Nullen gehen der r-ten Eins voraus $\} =$
$\{(\omega_1, \dots, \omega_n) \in \tilde{\Omega} \mid \omega_{r+k} = 1$ und $\#\{i < r+k \mid \omega_i = 0\} = k\}$
Es gilt $\tilde{P}(A_k) = \underbrace{\binom{k+r-1}{k}}_{\substack{\text{Mögl. für } k \text{ Nullen unter} \\ \text{den ersten } k+r-1 \text{ Komponenten}}} p^r(1-p)^k$
Offenbar gilt $\tilde{\Omega} = A_0 \,\dot{\cup} \dots \dot{\cup}\, A_{n-1} \,\dot{\cup}\, B$ mit $B = \{(\omega_1, \dots, \omega_n) \in \tilde{\Omega} \mid \#\{i \mid \omega_i = 1\} < r\}$.
Die Wahrscheinlichkeit $\tilde{P}(A_k)$ hängt nur von k und r ab, nicht von n, solange $n \geq k + r$ gilt.
Das Bernoulli-Experiment mit Parameter p werde nun unendlich oft wiederholt.
Wir beobachten, wie viele Nullen voraus gehen bis zum r-ten Mal das Ergebnis 1 auftritt.
$\Rightarrow \Omega = \{0, 1, 2, 3, \dots\} = \mathbb{N}$ und durch die Abbildung $P : \mathfrak{P} : \mathfrak{P}\Omega \to [0,1]$, def. durch $\{k\} \mapsto \binom{k+r-1}{k}p^r(1-p)^k$, wird eine Wahrscheinlichkeitsverteilung auf Ω definiert (siehe unten).

Sie heißt die ***Pascal-Verteilung*** (oder die ***negative Binomialverteilung***) ***auf*** $\boldsymbol{\Omega}$.

Setzt man $\binom{-r}{k} := \frac{(-r)(-r-1)\cdot\ldots\cdot(r-k+1)}{k!}$, so gilt $P(k) := \binom{-r}{k}(-1)^k p^r(1-p)^k$.

Beweis:
Additivität gilt nach Definition.
Noch zu zeigen: $\sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k = 1$
Beweis durch Induktion nach r:
Induktionsbeginn $r = 1$:
$\sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k = \sum\limits_{k=0}^{\infty} p(1-p)^k = p\sum\limits_{k=0}^{\infty}(1-p)^k = p\frac{1}{1-(1-p)} = 1$
Induktionsschluss $(r-1) \to r$ für $r \geq 2$:
$\sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k = p^r + \sum\limits_{k=1}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k \overset{\text{Pascalsches Dreieck}}{=}$
$p^r + \sum\limits_{k=1}^{\infty} \binom{k+r-2}{k} p^r(1-p)^k + \sum\limits_{k=1}^{\infty} \binom{k+r-2}{k-1} p^r(1-p)^k =$
$\sum\limits_{k=0}^{\infty} \binom{k+(r-1)-1}{k} p^r(1-p)^k + \sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^{k+1} =$
$p\sum\limits_{k=0}^{\infty} \binom{k+(r-1)-1}{k} p^{r-1}(1-p)^k + (1-p)\sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k \overset{\text{I.V.}}{=}$
$p + (1-p)\sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k$
$\Rightarrow p\sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k = p$
$\Rightarrow \sum\limits_{k=0}^{\infty} \binom{k+r-1}{k} p^r(1-p)^k = 1$ $\square$

Beispiel 5.4 (Die Poisson-Verteilung)
Sei $\lambda \in \mathbb{R}_+$. Auf der Potenzmenge von $\Omega = \mathbb{N}$ betrachten wir die von der Abbildung $P : \mathfrak{P}(\Omega) \to [0,1]$ definierte Wahrscheinlichkeitsverteilung $\{k\} \mapsto \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$.
Es gilt $\sum\limits_{k=0}^{\infty} \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} = \mathrm{e}^{\lambda}\mathrm{e}^{-\lambda} = 1$.
Sie heißt die ***Poisson-Verteilung auf*** $\mathbb{N}$.

Satz 5.5 *(Die Poisson-Verteilung als Grenzwert der Binomialverteilung)*
Sei $\lambda \in \mathbb{R}_+$ und für jedes $n \in \mathbb{N}$ sei $\Omega_n = \{0, \ldots, n\}$ der Wahrscheinlichkeitsraum mit der Binomialverteilung $P_n := b_{n,p}$ zum Parameter $p = \frac{\lambda}{n}$.
Dann gilt: $\lim\limits_{\substack{n\to\infty \\ n\geq k}} P_n(k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$, d.h. die Poisson-Verteilung ist der Grenzwert der Binomialverteilungen $b_{n,p:=\frac{\lambda}{n}}$ für $n \to \infty$.
Für $n \gg 0$ gilt die Näherungsformel $b_{n,p}(k) \approx \frac{(np)^k}{k!}\mathrm{e}^{-np}$.

Beweis:
Es gilt $\frac{\mathrm{d}}{\mathrm{d}x}(\ln(1-x))(0) = -1$
Also folgt $\ln(1-k) = -k + R(k)$, wobei R Restglied 1. Ordnung im Aufpunkt 0.
Mit $k = \frac{\lambda}{n}$ folgt für $n \to \infty$:
$\left(1 - \frac{\lambda}{n}\right)^n = \exp\left(n \ln\left(1 - \frac{\lambda}{n}\right)\right) = \exp\left(n\left(-\frac{\lambda}{n} + R\left(\frac{\lambda}{n}\right)\right)\right) =$
$\exp(-\lambda) \cdot \exp\left(\lambda \underbrace{\frac{R\left(\frac{\lambda}{n}\right)}{\left|\frac{\lambda}{n}\right|}}_{\substack{\to 0 \text{ für } n \to \infty \\ R \text{ Restglied in } 0}}\right) \to \mathrm{e}^{-\lambda}$ für $n \to \infty$

$\Rightarrow P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \overset{p=\frac{\lambda}{n}}{=} \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} =$
$\frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} =$
$= 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \ldots \cdot \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \to 1 \cdot \ldots \cdot 1 \cdot \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$ für $n \to \infty$
□

Beispiel 5.6 Sei $\lambda = 1$. Dann gilt:

k	$b_{10,\frac{1}{10}}(k)$	$b_{100,\frac{1}{100}}(k)$	Poisson
0	0,349	0,366	0,367
1	0,387	0,369	0,367
2	0,194	0,184	0,184
3	0,057	0,061	0,061

Beispiel 5.7 Ein Geiger-Müller-Zählrohr entdeckt ein von einer Strahlungsquelle ausgesandtes radioaktives Teilchen mit einer Wahrscheinlichkeit 10^{-4}.
Mit welcher Wahrscheinlichkeit entdeckt es von 30 000 emittierten Teilchen:

1. keines
2. mindestens 3

$\tilde{\Omega} = \{0,1\}^{30000}$ und $\tilde{P}$=Bernoulli-Verteilung zum Parameter $P = 10^{-4}$
Gesucht: $A = \{(0, \ldots, 0)\}$ und $B = \{(\omega_1, \ldots, \omega_{30000}) \in \tilde{\Omega} \mid \sum_i \omega_i \geq 3\}$
$\tilde{P}(A) = b_{30000,10^{-4}}(0) \approx \frac{3^0}{0!}\mathrm{e}^{-3} = \mathrm{e}^{-3} \approx 0,0498$
$\tilde{P}(B) = 1 - b_{30000,10^{-4}}(0) - b_{30000,10^{-4}}(1) - b_{30000,10^{-4}}(2) \approx 1 - \mathrm{e}^{-3} - \frac{3^1}{1!}\mathrm{e}^{-3} - \frac{3^2}{2!}\mathrm{e}^{-3} =$
$1 - \frac{17}{2}\mathrm{e}^{-3} \approx 0,5768$

Satz 5.8 *(Bernoullisches Gesetz der großen Zahlen)*
Sei $p \in [0,1]$. Für jedes $n \geq 0$ sei $\Omega_n = \{0,1\}^n$ der n-te Bernoulli-Raum zum Parameter p.
Für ein fest vorgegebenes $\varepsilon \in \mathbb{R}_+$ betrachten wir das Ereignis

$$A_n = \left\{(\omega_1, \dots, \omega_n) \in \Omega_n \mid \left|\underbrace{\frac{\omega_1 + \dots + \omega_n}{n}}_{\substack{\text{relative Häufigkeit} \\ \text{von } \omega_i = 1}} - p\right| > \varepsilon\right\}.$$

Dann gilt: $P(A_n) < \frac{1}{np(1-p)\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$

Insbesondere folgt:

$\lim\limits_{n\to\infty} P\left(\left\{(\omega_1, \dots, \omega_n) \in \Omega_n \mid \left|\frac{\omega_1+\dots+\omega_n}{n} - p\right| \leq \varepsilon\right\}\right) = 1$

Beweis:
Für alle $k \in \{1, \dots, n\}$ sei $B_{n,k} = \{(\omega_1, \dots, \omega_n) \in \Omega_n \mid \omega_1 + \dots + \omega_n = k\}$
Nach Beispiel 4.5 gilt $P(B_{n,k}) = \binom{n}{k} p^k (1-p)^{n-k}$
Weiter gilt $P(A_n) =$
$P(\{(\omega_1, \dots, \omega_n) \in \Omega_n \mid \omega_1 + \dots + \omega_n > n(p+\varepsilon)\})+$
$P(\{(\omega_1, \dots, \omega_n) \in \Omega_n \mid \omega_1 + \dots + \omega_n < n(p-\varepsilon)\}) =$
$\sum\limits_{k<n(p-\varepsilon)} P(B_{n,k}) + \sum\limits_{k>n(p+\varepsilon)} P(B_{n,k})$
Ist $k < n(p-\varepsilon)$, so gilt $np - k > n\varepsilon$, folglich $(np-k)^2 > n^2\varepsilon^2$ und falls $k > n(p+\varepsilon)$ gilt $(k-np)^2 > n^2\varepsilon^2$.
Also folgt $\frac{(np-k)^2}{n^2\varepsilon^2} > 1$ und

$P(A_n) < \sum\limits_{k<n(p-\varepsilon)} \frac{(np-k)^2}{n^2\varepsilon^2} P(B_{n,k}) + \sum\limits_{k>n(p+\varepsilon)} \frac{(np-k)^2}{n^2\varepsilon^2} P(B_{n,k}) \leq$

$\frac{1}{n^2\varepsilon^2} \sum\limits_{k\geq 0} (k-np)^2 \binom{n}{k} p^k (1-p)^{n-k} =$

$$\frac{1}{n^2\varepsilon^2} \underbrace{\sum_{k\geq 0} k^2 \binom{n}{k} p^k (1-p)^{n-k}}_{=:S_1} - \frac{2p}{n\varepsilon^2} \underbrace{\sum_{k\geq 0} k \binom{n}{k} p^k (1-p)^{n-k}}_{=:S_2} +$$

$$\frac{p^2}{\varepsilon^2} \underbrace{\sum_{k\geq 0} \binom{n}{k} p^k (1-p)^{n-k}}_{=:S_3 = 1 \text{ nach Bsp. } 4.5}$$

Für $k \geq 1$ gilt $k\binom{n}{k} = \frac{n\cdot\ldots\cdot(n-k+1)}{1\cdot\ldots\cdot(k-1)} = n\binom{n-1}{k-1}$ und folglich

$S_2 = \sum\limits_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} = \sum\limits_{k=0}^{n-1} n\binom{n-1}{k} p^{k+1} (1-p)^{n-1-k} =$

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

Für $k \geq 2$ gilt $k^2\binom{n}{k} = (k + k(k-1))\binom{n}{k} = k\binom{n}{k} + n(n-1)\binom{n-2}{k-2}$ und folglich

$S_1 = np(1-p)^{n-1} + \sum\limits_{k=2}^{n} k\binom{n}{k} p^k (1-p)^{n-k} + n(n-1) \sum\limits_{k=2}^{n} \binom{n-2}{k-2} p^k (1-p)^{n-k} =$

$$\underbrace{\sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}}_{=S_2 = np} + n(n-1) \underbrace{\sum_{k=0}^{n-2} \binom{n-2}{k} p^2 p^k (1-p)^{n-2-k}}_{=p^2(p+(1-p))^{n-2} = p^2} =$$

$np + n(n-1)p^2$

$\Rightarrow P(A_n) < \frac{1}{n^2\varepsilon^2}(np + n(n-1)p^2) - \frac{2p}{n\varepsilon^2}np + \frac{p^2}{\varepsilon^2} =$
$\frac{1}{\varepsilon^2}\left(\frac{p}{n} + p^2 - \frac{p^2}{n} - 2p^2 + p^2\right) = \frac{1}{n\varepsilon^2}p(1-p) \leq \frac{1}{4n\varepsilon^2}$ $\square$

6 Diskrete Zufallsvariable

Im folgenden sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum.

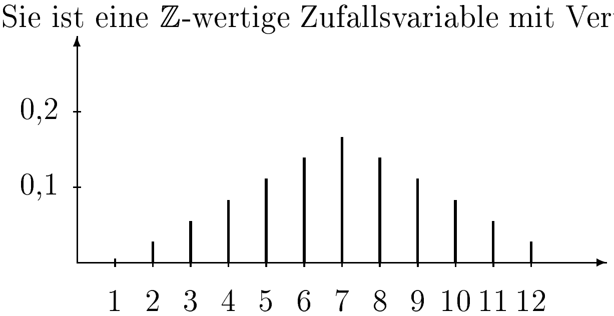
Definition 6.1

1. Sei M eine Menge. Eine Abbildung $X : \Omega \to M$ heißt ***M*-wertige diskrete Zufallsvariable auf Ω**.
2. Die Menge $W = X(\Omega)$ ist abzählbar und heißt die ***Wertemenge von X***.
3. Die Abbildung $P_X := P \circ X^{-1} : \mathfrak{P}(W) \to [0,1]$, def. durch $\{x\} \mapsto P(\{\omega \in \Omega \mid X(\omega) = x\}) = P \circ X^{-1}(\{\omega\})$ und $A \mapsto \sum\limits_{x \in A} P_X(\{x\})$, definiert eine Wahrscheinlichkeitsverteilung auf W.
   Sie heißt die ***Verteilung von X***.

Bemerkung 6.2

1. *Setzt man $P_X(\{x\}) = 0$ für $x \in M \setminus W$, so kann man P_X auch als Wahrscheinlichkeitsverteilung auf der (evtl. überabzählbar unendlichen) Menge M betrachten.*
2. *Für alle Teilmengen $A \subset M$ gilt $P_X(A) = P \circ X^{-1}(A)$.*
3. *Man kann die Verteilung P_X durch ein Stabdiagramm veranschaulichen.*

Beispiel 6.3 Sei $\Omega = \{1, \ldots, 6\}^2$ und P =Gleichverteilung der Wahrscheinlichkeitsraum, der das gleichzeitige Werfen zweier unterscheidbarer Würfel beschreibt.
Sei $X : \Omega \to \mathbb{Z}$ die Augensumme $(\omega_1, \omega_2) \mapsto \omega_1 + \omega_2$.
Sie ist eine $\mathbb{Z}$-wertige Zufallsvariable mit Verteilung



Notation:
1. $P(X = x) := P(\{\omega \in \Omega \mid X(\omega) = x\})$ für $x \in W$
2. $A \subset W$ Teilmenge $\Rightarrow P(X \in A) := P(\{\omega \in \Omega \mid X(\omega) \in A\})$
3. $X_i : \Omega \to M_i$ Zufallsvariable und $A_i \subset M_i$ $(i = 1, 2)$
   $\Rightarrow P(X_1 \in A_1, X_2 \in A_2) = P(\{\omega \in \Omega \mid X_1(\omega) \in A_1, X_2(\omega) \in A_2\})$

Definition 6.4 Sei $n \in \mathbb{N}_+$ und seien $X_i : \Omega \to M_i$ für $i = 1, \ldots, n$ Zufallsvariable.

Dann ist die durch $X = X_1 \times \ldots \times X_n : \Omega \to M_1 \times \ldots \times M_n$, def. durch $\omega \mapsto (X_1(\omega), \ldots, X_n(\omega))$, eine $M_1 \times \ldots \times M_n$-wertige Zufallsvariable auf Ω gegeben. Ihre Verteilung P_X heißt die ***gemeinsame Verteilung von $X_1, \ldots, X_n$***.

Beispiel 6.5 Sei $\Omega = \{0, 1\}^n$ der n-te Bernoulli-Raum zum Parameter $p \in [0, 1]$.
Wir definieren zwei Zufallsvariable

$X_1 : \Omega \to \mathbb{Z}$ $(\omega_1, \ldots, \omega_n) \mapsto \omega_1 + \ldots + \omega_n$ und

$$X_2 : \Omega \to \mathbb{Z}\ (\omega_1, \ldots, \omega_n) \mapsto \begin{cases} \min\{i \mid \omega_i = 1\} & \text{falls} \quad \omega_1 + \ldots + \omega_n > 0 \\ n + 1 & \text{falls} \quad \omega_1 = \ldots = \omega_n = 0 \end{cases}$$

Sei $X = X_1 \times X_2$ und P_X die gemeinsame Verteilung von X_1 und X_2.
Offenbar wird P_X vollständig beschrieben durch die Angabe von
$P_X(X_1 = a, X_2 = b)$ für $0 \leq a \leq n$ und $1 \leq b \leq n + 1$.

1. Es gilt $P_X(X_1 = 0, X_2 = n + 1) = (1 - p)^n$ und $P_X(X_1 = 0, X_2 = b) = 0$ für $1 \leq b \leq n$.
2. Ist $X_1(\omega) = a \geq 1$ für ein Tupel $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ und $X_2(\omega) = b$, so müssen genau $a - 1$ Einsen unter den Elementen $\omega_{b+1}, \ldots, \omega_n$ sein.
   $\Rightarrow P_X(X_1 = a, X_2 = b) = \binom{n-b}{a-1} p^a (1 - p)^{n-a}$ für $a \geq 1$.

Die Verteilung P_X läßt sich als Matrix darstellen, z.B. für $n = 3$, $p = \frac{1}{2}$.

$X_2 \setminus X_1$	0	1	2	3
1	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$	0
3	0	$\frac{1}{8}$	0	0
4	$\frac{1}{8}$	0	0	0

Definition 6.6 Sei I eine Menge.
Eine Familie von Zufallsvariablen $\{X_i\}_{i \in I}$ mit $X_i : \Omega \to M_i$ für $i \in I$ heißt ***unabhängig***, wenn für jede Wahl von Teilmengen $A_i \subset M_i$ für $i \in I$ gilt, dass die Ereignisse $\{X_i \in A_i\}_{i \in I}$ unabhängig sind.

Satz 6.7 *(Charakterisierung unabhängiger Zufallsvariablen)*

Sei $n \in \mathbb{N}_+$ und seien $X_i : \Omega \to M_i$ für $i = 1, \ldots, n$ Zufallsvariable. Dann sind die folgenden Bedingungen äquivalent:

1. *Die Zufallsvariablen $X_1, \ldots, X_n$ sind unabhängig.*
2. *Für alle Tupel $(x_1, \ldots, x_n) \in M_1 \times \ldots \times M_n$ gilt $P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$*
3. *Sind Teilmengen $A_i \subset M_i$ gegeben (für $i = 1, \ldots, n$), so gilt $P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$*
4. *Die Verteilung von $X := X_1 \times \ldots \times X_n$ ist das Produkt der Verteilungen $P_{X_1}, \ldots, P_{X_n}$, d.h. $P_X = P_{X_1} \times \ldots \times P_{X_n}$.*

Beweis:

1.⇒2.

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(\{X_1 \in \{x_1\}\} \cap \ldots \cap \{X_n \in \{x_n\}\}) =$$
$$\prod_{i=1}^{n} P(X_i \in \{x_i\}) = \prod_{i=1}^{n} P(X_i = x_i)$$

2.⇒3.

$$P(X_1 \in A_1, \ldots, X_n \in A_n) =$$
$$\sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} P(X_1 = x_1, \ldots, X_n = x_n) =$$
$$\sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) =$$
$$\left(\sum_{x_1 \in A_1} P(X_1 = x_1) \right) \cdot \ldots \cdot \left(\sum_{x_n \in A_n} P(X_n = x_n) \right) =$$
$$P(X_1 \in A_1) \cdot \ldots \cdot P(X_n \in A_n)$$

3.⇒4.

Zu zeigen ist, dass für alle $(x_1, \ldots, x_n) \in M_1 \times \ldots \times M_n$ gilt:

$$P_X(\{(x_1, \ldots, x_n)\}) = P_{X_1}(\{x_1\}) \cdot \ldots \cdot P_{X_n}(\{x_n\})$$

Mit Hilfe der Definition von P_X und mit $A_i = \{x_i\} \subset M_i$ berechnen wir:

$$P_X(\{(x_1, \ldots, x_n)\}) = P \circ X^{-1}(\{(x_1, \ldots, x_n)\}) =$$
$$P(X_1^{-1}(x_1) \times \ldots \times X_n^{-1}(x_n)) = P(X_1 \in \{x_1\}, \ldots, X_n \in \{x_n\}) =$$
$$\prod_{i=1}^{n} P(X_i \in \{x_i\}) = \prod_{i=1}^{n} P \circ X_i^{-1}(x_i) = \prod_{i=1}^{n} P_{X_i}(\{x_i\})$$

4.⇒3.

Die Ereignisse $\{X_i \in A_i\}_{i \in I}$ sind unabhängig, denn für $J \subset I$ endlich gilt:

Setze $B_j := A_j$ für $j \in J$ und $B_j := M_j$ für $j \in I \setminus J$ und erhalte:

$$P\left(\bigcap_{j \in J} \{X_j \in A_J\} \right) = P\left(\bigcap_{j \in I} \{X_j \in B_j\} \right) =$$
$$\sum_{(x_1, \ldots, x_n) \in B_1 \times \ldots \times B_n} P\left(\bigcap_{i \in I} \{X_j = x_j\} \right) = \sum_{(x_1, \ldots, x_n) \in B_1 \times \ldots \times B_n} P_X(\{(x_1, \ldots, x_n)\}) =$$

$$\sum_{(x_1,\dots,x_n)\in B_1\times\dots\times B_n} P_{X_1}(x_1)\cdot\ldots\cdot P_{X_n}(x_n) =$$
$$\left(\sum_{x_1\in B_1} P_{X_1}(x_1)\right)\cdot\ldots\cdot\left(\sum_{x_n\in B_n} P_{X_n}(x_n)\right) =$$
$$P(\{X_1\in B_1\})\cdot\ldots\cdot P(\{X_n\in B_n\}) = \prod_{j\in J} P(\{X_j\in A_j\}) \qquad \square$$

Korollar 6.8 *Sind $X_i:\Omega\to M_i$ unabhängige Zufallsvariable und sind $F_i:M_i\to N_i$ Abbildungen in Mengen N_i ($i=1,\dots,n$), so sind die Zufallsvariablen $F_1\circ X_1,\dots,F_n\circ X_n$ ebenfalls unabhängig.*

Beweis:
Für $(y_1,\dots,y_n)\in N_1\times\ldots\times N_n$ definieren wir $A_i := F_i^{-1}(\{y_i\})\subset M_i$.
Dann gilt
$$P(F_1\circ X_1 = y_1,\dots,F_n\circ X_n = y_n) = P(X_1\in A_1,\dots,X_n\in A_n) =$$
$$= \prod_{i=1}^n P(X_i\in A_i) = \prod_{i=1}^n P(F_i\circ X_i = y_i)$$
Nach dem Satz folgt die Behauptung. $\square$

Definition 6.9 Sei $X:\Omega\to\mathbb{R}$ eine reellwertige (diskrete) Zufallsvariable. Dann heißt die Funktion $F_X:\mathbb{R}\to[0,1]$, def. durch $x\mapsto P(X\le x) = P_X(]-\infty,x]) = P(\{\omega\in\Omega\mid X(\omega)\le x\})$ die ***Verteilungsfunktion von X***.

Satz 6.10 *(Eigenschaften der Verteilungsfunktion)*
Sei $X:\Omega\to\mathbb{R}$ eine reellwertige Zufallsvariable.

1. *Für $x\in\mathbb{R}$ gilt $F_X(x) = \sum_{\{y\in X(\Omega)\mid y\le x\}} P(X=y)$.*
2. *Die Funktion F_X ist monoton steigend, d.h. aus $x\le y$ folgt $F_X(x)\le F_X(y)$.*
3. *Für $a,b\in\mathbb{R}$ gilt $P(a<X\le b) = F_X(b)-F_X(a)$, sowie $P(X>a) = 1-F_X(a)$.*
4. *Für $a,b\in\mathbb{R}$ gilt $P(a\le X\le b) = F_X(b) - \lim_{\substack{x\overset{>}{\to}a\\ x\ne a}} F_X(x)$ und $P(a\le X) = 1-\lim_{\substack{x\overset{>}{\to}a\\ x\ne a}} F_X(x)$.*

Beweis:

1. folgt aus $\{X\le x\} = X^{-1}(\{y\in X(\Omega)\mid y\le x\})$
2. folgt sofort aus 1.
3. Es gilt $P(a<X\le b) = P(X\le b) - P(X\le a) = F_X(b)-F_X(a)$, sowie $P(x>a) = 1-P(X\le a)$.

4. Es gilt $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) + P(X = a) =$
$F_X(b) - (F_X(a) - P(X = a)) = F_X(b) - \sum\limits_{\{y \in X(\Omega) | y < a\}} P(X < y) =$
$F_X(b) - \lim\limits_{\substack{x \overset{>}{\to} a \\ x \neq a}} F_X(x)$
$P(a \leq X) = 1 - P(X < a) = 1 - (F_X(a) - P(X = a)) \overset{\text{wie oben}}{=}$
$1 - \lim\limits_{\substack{x \overset{>}{\to} a \\ x \neq a}} F_X(x)$

$\square$

Beispiel 6.11 „Mensch-ärgere-dich-nicht"
Die Zufallsvariable X soll die Anzahl der Versuche beschreiben, die ein Spieler braucht, bis er eine 6 würfelt.
$\Rightarrow \Omega = \mathbb{N}_+$ mit P =geometrische Verteilung zum Parameter $p = \frac{1}{6}$.
$\Rightarrow X : \mathfrak{P}(\Omega) \to [0,1]$ und $F_X(x) = P(X \leq x) = \sum\limits_{i=1}^{[x]} \frac{1}{6} \left(\frac{5}{6}\right)^{i-1} = \frac{1}{6} \frac{1 - \left(\frac{5}{6}\right)^{[x]}}{1 - \frac{5}{6}} =$
$1 - \left(\frac{5}{6}\right)^{[x]}$, wobei $[x] := \max\{z \in \mathbb{Z} \mid z \leq x\}$

7 Erwartungswerte

Sei (Ω, P) im Folgenden stets ein diskreter Wahrscheinlichkeitsraum.

Definition 7.1 Sei $X : \Omega \to \mathbb{R}$ eine reellwertige (diskrete) Zufallsvariable auf Ω. Konvergiert die Reihe $\sum\limits_{\omega \in \Omega} |X(\omega)| P(\omega)$, so heißt die Zahl $E(X) = \sum\limits_{\omega \in \Omega} X(\omega) P(\omega)$ der ***Erwartungswert von X***.
Für komplexwertiges X sei der Erwartungswert von X genauso definiert.

Bemerkung 7.2

1. Aus der absoluten Konvergenz der Reihe $\sum\limits_{\omega \in \Omega} X(\omega) P(\omega)$ folgt die Konvergenz jeder Umordnung.
Es ist daher egal, wie wir die Menge Ω abzählen.

2. Ist $X(\Omega) = \{x_1, x_2, \dots\}$ und existiert $E(X)$, so gilt $E(X) = \sum\limits_{i=1}^{\infty} x_i P(X_i = x_i)$, denn
$E(X) = \sum\limits_{\omega \in \Omega} X(\omega) P(\omega) = \sum\limits_{i=1}^{\infty} \sum\limits_{\{\omega \in \Omega | X(\omega) = x_i\}} X(\omega) P(\omega) =$
$\sum\limits_{i=1}^{\infty} x_i P(\{\omega \in \Omega \mid X(\omega) = x_i\})$

Beispiel 7.3 (Der „Tout“ beim Schafkopf)
In der Situation von Beispiel 3.1 („Schafkopf mit der Kurzen“) spielen wir zum Tarif „Zehnerl, Zwanzgerl mit Schuss“.
Der Spieler, der als Erster herauskommt, habe folgendes Blatt:
Eichel Unter, Herz Unter, 4 Asse
Soll er das Spiel als „Tout“ spielen, d.h. ansagen, dass er alle Stiche machen will?

1. als Tout:
   Es gibt zwei Ergebnisse:
   Stehen die beiden anderen Unter zusammen, so verliert der Spieler.
   $P(\text{„Verlust“}) = 0,294$, andernfalls gewinnt er, $P(\text{„Gewinn“}) = 0,706$.
2. als normales Spiel:
   Stehen die beiden Unter zusammen, so gewinnt der Spieler „Schneider“ mit $P(\text{„Schneider“}) = 0,294$ und andernfalls gewinnt er „Schneider Schwarz“mit $P(\text{„Schneider Schwarz“}) = 0,706$.
   Es gelte $\Omega = \{\text{Gewinn, Verlust}\}$ bzw. $\Omega = \{\text{Schneider, Schneider Schwarz}\}$.

Die Zufallsvariable $X : \Omega \to \mathbb{R}$ beschreibe den Geldbetrag in DM, den der Spieler gewinnt oder verliert.

1. Es gilt: $X(\text{„Gewinn“}) = 3 \cdot 0,80 = 2,40$
   (20 Pf. Spiel, 20 Pf Schn. Schw., Faktor 2 für „Tout“)
   (zusätzlich erhält der Spieler einen „Schuss“.)
   Also folgt: $E(X) = 0,706 \cdot 2,40 + 0,294 \cdot (-4,80) \approx 0,283$
2. Es gilt: $X(\text{„Schneider“}) = 3 \cdot 0,30 = 0,90$ und $X(\text{„Schn. Schw.“}) = 3 \cdot 0,40 = 1,20$
   Also folgt: $E(X) = 0,706 \cdot 1,2 + 0,294 \cdot 0,9 \approx 1,112$

Beispiel 7.4 Sei $A \subset \Omega$ ein Ereignis. Dann heißt die Zufallsvariable $X = 1_A : \Omega \to \mathbb{R}$, def. durch $\omega \mapsto \begin{cases} 1 & \text{falls} \quad \omega \in A \\ 0 & \text{falls} \quad \omega \in \Omega \setminus A \end{cases}$ die ***Indikatorfunktion*** oder ***charakteristische Funktion von A***.
Es gilt: $E(1_A) = 1 \cdot P(1_A = 1) + 0 \cdot P(1_A = 0) = P(A)$
Für endlich viele Ereignisse $A_1, \ldots, A_n \subset \Omega$ gilt:

1. $1_{A_i}^2 = 1_{A_i}$
2. $A_1, \ldots, A_n$ paarweise disjunkt $\Rightarrow 1_{A_1 \cup \ldots \cup A_n} = 1_{A_1} + \ldots + 1_{A_n}$

3. $1_{A_1\cap\ldots\cap A_n} = 1_{A_1} \cdot \ldots \cdot 1_{A_n}$
4. $1_{\Omega\setminus A_i} = 1 - 1_{A_i}$
5. $A_1, \ldots, A_n$ unabhängig $\Leftrightarrow 1_{A_1}, \ldots 1_{A_n}$ unabhängige Zufallsvariable

Definition 7.5 Sei R ein Ring und seien $X_1, X_2 : \Omega \to R$ R-wertige Zufallsvariable.

1. Durch $X_1 + X_2 : \Omega \to R$, def. durch $\omega \mapsto X_1(\omega) + X_2(\omega)$, wird eine Zufallsvariable auf Ω definiert, die ***Summe von X_1 und X_2 heißt***.
2. Durch $X_1 \cdot X_2 : \Omega \to R$, def. durch $\omega \mapsto X_1(\omega) \cdot X_2(\omega)$, wird eine Zufallsvariable auf Ω definiert, die ***Produkt von X_1 und X_2 heißt***.
3. Für jedes $r \in R$ wird durch $rX_1 : \Omega \to R$, def. durch $\omega \mapsto rX_1(\omega)$, eine Zufallsvariable auf Ω definiert.

Satz 7.6 *(Rechenregeln für Erwartungswerte)*
Seien $X, X_1, X_2 : \Omega \to \mathbb{C}$ komplexwertige Zufallsvariable, deren Erwartungswert existiere.

1. *Für jedes $\lambda \in \mathbb{C}$ existiert der Erwartungswert von λX und es gilt $E(\lambda X) = \lambda E(X)$.*
2. *Der Erwartungswert von $X_1 + X_2$ existiert und es gilt $E(X_1 + X_2) = E(X_1) + E(X_2)$.*
3. *Sind X_1 und X_2 unabhängig, so existiert der Erwartungswert von $X_1 \cdot X_2$ und es gilt $E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2)$.*
4. *Ist $Y : \Omega \to M$ eine Zufallsvariable mit Werten in einer Menge M und ist $F : M \to \mathbb{C}$ eine Abbildung, so existiert der Erwartungswert der komplexwertigen Zufallsvariablen $F \circ Y$ genau dann, wenn der Erwartungswert der Zufallsvariable $\tilde{F} = F|_{Y(\Omega)} : Y(\Omega) \to \mathbb{C}$ existiert.*
   In diesem Fall gilt $E(F \circ Y) = E(\tilde{F})$.
   Mit anderen Worten: Gilt $Y(\Omega) = \{y_1, y_2, \ldots\}$, so ergibt sich $E(F \circ Y) = \sum\limits_{i=1}^{\infty} F(y_i)P(Y = y_i)$.

Beweis:

1. Da $\sum\limits_{\omega\in\Omega} |X(\omega)|P(\omega)$ konvergiert, konvergiert auch $\sum\limits_{\omega\in\Omega} |\lambda X(\omega)|P(\omega) = \sum\limits_{\omega\in\Omega} |\lambda| \cdot |X(\omega)|P(\omega) = |\lambda| \sum\limits_{\omega\in\Omega} |X(\omega)|P(\omega)$.

Damit gilt
$E(\lambda X) = \sum_{\omega\in\Omega} (\lambda X)(\omega)P(\omega) = \sum_{\omega\in\Omega} \lambda \cdot X(\omega)P(\omega) = \lambda \sum_{\omega\in\Omega} X(\omega)P(\omega) = \lambda E(X)$.

2. Da $\sum_{\omega\in\Omega} |X_i(\omega)|P(w)$ für $i = 1, 2$ konvergieren, konvergiert auch
$\sum_{\omega\in\Omega} |(X_1 + X_2)(\omega)|P(\omega) = \sum_{\omega\in\Omega} |X_1(\omega) + X_2(\omega)|P(\omega) \leq$
$\sum_{\omega\in\Omega} |X_1(\omega)P(\omega) + \sum_{\omega\in\Omega} |X_2(\omega)|P(\omega)$ und es gilt
$E(X_1 + X_2) = \sum_{\omega\in\Omega} (X_1 + X_2)(\omega)P(\omega) =$
$\sum_{\omega\in\Omega} X_1(\omega)P(\omega) + \sum_{\omega\in\Omega} X_2(\omega)P(\omega) = E(X_1) + E(X_2)$.

3. Sei $X_1(\Omega) = \{x_1, x_2, x_3, \dots\}$ und $X_2(\Omega) = \{y_1, y_2, y_3, \dots\}$.
Dann gilt
$\sum_{\omega\in\Omega} |X_1(\omega)X_2(\omega)|P(\omega) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |x_i||y_j|P(X_1 = x_i, X_2 = y_j) =$
$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |x_i||y_j|P(X_1 = x_i)P(X_2 = y_j) =$
$\left(\sum_{i=1}^{\infty} |x_i|P(X_1 = x_i)\right)\left(\sum_{j=1}^{\infty} |y_j|P(X_2 = y_j)\right)$, d.h. der Erwartungswert
$E(X_1 \cdot X_2)$ existiert und erfüllt
$E(X_1 \cdot X_2) =$
$\sum_{\omega\in\Omega} X_1(\omega)X_2(\omega)P(\omega) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j P(X_1 = x_i, X_2 = y_j) =$
$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_i y_j P(X_1 = x_i)P(X_2 = y_j) =$
$\left(\sum_{i=1}^{\infty} x_i P(X_1 = x_i)\right)\left(\sum_{j=1}^{\infty} y_j P(X_2 = y_j)\right)$

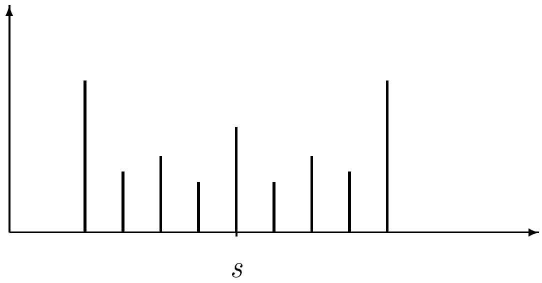
4. Der Erwartungswert $E(F \circ Y)$ existiert genau dann, wenn die Reihe
$\sum_{\omega\in\Omega} |F \circ Y(\omega)|P(\omega)$ konvergiert.
Der Erwartungswert von $E(\tilde{F})$ existiert genau dann, wenn die Reihe
$\sum_{x\in Y(\Omega)} |F(x)|\underbrace{P(\{\omega \in \Omega \mid Y(\omega) = x\})}_{P_Y(x)}$ konvergiert.
Diese beiden Bedingungen sind äquivalent, denn für $x \in Y(\Omega)$ gilt
$|F(x)|P(\{\omega \in \Omega \mid Y(\omega) = x\}) = \sum_{\{\omega\in\Omega \mid Y(\omega)=x\}} |F \circ Y(\omega)|P(\omega)$.
Existieren beide Erwartungswerte, so sind sie wegen
$E(F \circ Y) = \sum_{\omega\in\Omega} (F \circ Y)(\omega)P(\omega) = \sum_{x\in Y(\Omega)} F(x) \cdot (P \circ Y^{-1})(x) = E(\tilde{F})$ gleich.
Der Zusatz folgt aus
$E(F \circ Y) = \sum_{x\in Y(\Omega)} F(x) \cdot (P \circ Y^{-1})(x) = \sum_{i=1}^{\infty} F(y_i)P(Y = y_i)$.

□

Bemerkung 7.7 *Mit vollständiger Induktion folgt aus Satz 7.6 für komplexwertige $X_1, \ldots, X_n : \Omega \to \mathbb{C}$, deren Erwartungswerte existieren, die Gültigkeitkeit der folgenden Formeln:*

1. *Für $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ gilt*
   $E(\lambda_1 X_1 + \ldots + \lambda_n X_n) = \lambda_1 E(X_1) + \ldots + \lambda_n E(X_n)$
2. *Sind $X_1, \ldots, X_n$ unabhängig, so gilt $E(X_1 \cdot \ldots \cdot X_n) = E(X_1) \cdot \ldots \cdot E(X_n)$*

Korollar 7.8 *Eine reellwertige Zufallsvariable $X : \Omega \to \mathbb{R}$ besitze einen Erwartungswert und sei symmetrisch verteilt mit Symmetriepunkt $s \in \mathbb{R}$, d.h. es gelte $X(\Omega) = \{x_1, x_2, \ldots\} \dot\cup \{y_1, y_2, \ldots\}$ mit $x_i - s = s - y_i$ und mit $P(X = x_i) = P(X = y_i)$ für $i = 1, \ldots, m$.*



Dann gilt $E(X) = s$

Beispiel 7.9 (Die Erwartungswerte der klassischen Wahrscheinlichkeitsverteilungen)

1. Sei $X : \Omega \to \mathbb{R}$ eine binomialverteilte Zufallsvariable, d.h. P_X sei die Binomialverteilung zum Parameter $p \in [0,1]$ auf $X(\Omega) = \{0, \ldots, n\}$.
   Dann gilt $E(X) = \sum\limits_{x \in X(\Omega)} x \cdot P_X(x) = \sum\limits_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \overset{\text{Beweis von 5.8}}{=} np.$
2. Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ sei geometrisch verteilt zum Parameter $p \in]0,1]$, d.h. $X(\Omega) = \mathbb{N}_+$ und $P_X(k) = P(X = k) = (1-p)^{k-1} p$ für $k \in \mathbb{N}_+$.
   Dann ergibt sich:
   $E(X) = \sum\limits_{k=1}^{\infty} kp(1-p)^{k-1}$.
   Für $|x| < 1$ gilt $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$ und Ableiten liefert $\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \ldots$
   Einsetzen von $x = 1 - p$ liefert $\frac{1}{p^2} = 1 + 2(1-p) + 3(1-p)^2 + \ldots$
   Insgesamt ergibt sich:
   $E(X) = \frac{1}{p}$

3. In einer Urne sind w weiße und s schwarze Kugeln. Es werden $n \leq w + s$ Kugeln ohne Zurücklegen gezogen und es wird notiert, wieviele weiße Kugeln gezogen wurden.

   $\Rightarrow \Omega = \{\text{weiß,schwarz}\}^n$ und $X : \Omega \to \mathbb{R}$ ist def. durch $(\omega_1, \dots, \omega_n) \mapsto \#\{i \mid \omega_i = \text{weiß}\}$.

   Die Zufallsvariable X ist hypergeometrisch verteilt mit Parametern n, w, s, d.h. es gilt

   $X(\Omega) = \{0, \dots, n\}$ und $P(X = i) = \frac{\binom{w}{i}\binom{s}{n-i}}{\binom{w+s}{n}}$ für $i = 0, \dots, n$.

   Dann gilt:

   $E(X) = \sum\limits_{i=0}^{n} i \frac{\binom{w}{i}\binom{s}{n-i}}{\binom{w+s}{n}} = \frac{nw}{s+w}$

Beweis:
Sei $X_i : \Omega \to \mathbb{R}$ def. durch $X_i(\omega_1, \dots, \omega_n) = 1$, falls $\omega_i = \text{weiß}$ und $X_i(\omega_1, \dots, \omega_n) = 0$, falls $\omega_i = \text{schwarz}$.
Nun gilt offensichtlich $X = X_1 + \ldots + X_n$. Weiterhin gilt $E(X_1) = \frac{w}{w+s}$ und aus Symmetriegründen ebenso $E(X_i) = \frac{w}{w+s}$ für die anderen $i = 2, \dots, n$.
Folglich ist $E(X) = E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = \frac{nw}{w+s}$ nach 7.7. $\square$

8 Standardabweichung

Im Folgenden sei (Ω, P) stets ein diskreter Wahrscheinlichkeitsraum.

Lemma 8.1 *Seien $X, Y : \Omega \to \mathbb{C}$ zwei komplexwertige Zufallsvariable.*

1. *Existiert $E(X^2)$, so existieren auch $E(X)$ und $E((X - E(X))^2)$.*
2. *Existieren $E(X^2)$ und $E(Y^2)$, so existiert auch $E((X - E(X))(Y - E(Y)))$.*

Beweis:

1. Für $\omega \in \Omega$ gilt $(|X(\omega)| - 1)^2 \geq 0$, also $2|X(\omega)| \leq 1 + |X(\omega)|^2$.
   Es folgt $\sum\limits_{\omega \in \Omega} |X(\omega)|P(\omega) \leq \frac{1}{2} + \frac{1}{2} \sum\limits_{\omega \in \Omega} |X^2(\omega)|P(\omega)$, d.h. $E(X)$ existiert.
   Außerdem gilt für alle $\omega \in \Omega$ die Ungleichung
   $|X(\omega) - E(X)|^2 = |X^2(\omega) - 2X(\omega)E(X) + (E(X))^2| \leq |X^2(\omega)| + 2|X(\omega)| + |E(X)|^2$, so dass aus der Existenz von $E(X^2)$ und $E(X)$ auch die von $E((X - E(X))^2)$ folgt.
2. Für alle $\omega \in \Omega$ gilt
   $|(X(\omega) - E(X))(Y(\omega) - E(Y))| \leq$
   $|X(\omega)Y(\omega)| + |E(Y)||X(\omega)| + |E(X)||Y(\omega)| + |E(X)E(Y)| \leq$
   $\frac{1}{2}|X(\omega)|^2 + \frac{1}{2}|Y(\omega)|^2 + |E(Y)||X(\omega)| + |E(X)||Y(\omega)| + |E(X)E(Y)|$ und hieraus folgt die Behauptung.

$\square$

Definition 8.2 Seien $X, Y : \Omega \to \mathbb{R}$ zwei reellwertige Zufallsvariable, für die $E(X^2)$ und $E(Y^2)$ existieren.

1. Die Zahl $\mathrm{Var}(X) := E((X - E(X))^2) \in [0, \infty[$ heißt die ***Varianz*** (oder der ***zweite zentrale Moment***) ***von*** $\boldsymbol{X}$.
2. Die Zahl $\sigma(X) := \sqrt{\mathrm{Var}(X)}$ heißt die ***Standardabweichung*** (oder die ***Streuung***) ***von*** $\boldsymbol{X}$.
3. Die Zahl $\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))$ heißt die ***Kovarianz von*** $\boldsymbol{X}$ ***und*** $\boldsymbol{Y}$.
4. Sei $\sigma(X)\sigma(Y) \neq 0$. Dann heißt die Zahl $\rho(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$ der ***Korrelationskoeffizient von*** $\boldsymbol{X}$ ***und*** $\boldsymbol{Y}$.
5. Gilt $\mathrm{Cov}(X, Y) = 0$, so heißen die Zufallsvariablen X und Y unkorreliert.

Bemerkung 8.3 *Sind* $X, Y : \Omega \to \mathbb{C}$ *komplexwertige Zufallsvariable, so definiert man analog.*

1. $\mathrm{Var}(X) := E(|X - E(X)|^2)$ *falls* $E(X^2)$ *existiert.*

2. $\sigma(X) := \sqrt{\mathrm{Var}(X)}$

3. $\mathrm{Cov}(X, Y) = E((X - E(X))\overline{(Y - E(Y))})$ *usw.*

Beispiel 8.4 (Monte-Carlo-Roulette)

Drei Spieler verfolgen beim Roulette die folgenden Strategien:

1. Spieler A setzt stets auf die Zahl 13.
2. Spieler B setzt stets auf das erste Drittel: $\{1, \ldots, 12\}$.
3. Spieler C setzt stets auf Ungerade: $\{1, 3, 5, 7, \ldots, 35\}$.

Die Zufallsvariablen $A, B, C : \Omega \to \mathbb{R}$ sollen den Reingewinn des jeweiligen Spielers bei einem Spiel beschreiben.

1. Es gilt $A(\Omega) = \{-1, 35\}$ mit $P(A = -1) = \frac{36}{37}$ und $P(A = 35) = \frac{1}{37}$.
   $\Rightarrow E(A) = (-1) \cdot \frac{36}{37} + 35 \cdot \frac{1}{37} = -\frac{1}{37}$, sowie
   $(A - E(A))^2(\Omega) = \left\{ \left(\frac{36}{37}\right)^2, \left(35 + \frac{1}{37}\right) \right\}$
   $\Rightarrow \mathrm{Var}(A) = \frac{36}{37}\left(\frac{36}{37}\right)^2 + \frac{1}{37}\left(35 + \frac{1}{37}\right)^2 \approx 34,08$ und $\sigma(A) \approx 5,84$.
2. Es gilt $B(\Omega) = \{-1, 2\}$ mit $P(B = -1) = \frac{25}{37}$ und $P(B = 2) = \frac{12}{37}$.
   $\Rightarrow E(B) = \frac{12}{37}2 + \frac{25}{37}(-1) = -\frac{1}{37}$

Weiter gilt:
$\operatorname{Var}(B) = \frac{25}{37}\left(\frac{36}{37}\right)^2 + \frac{12}{37}\left(2+\frac{1}{37}\right)^2 \approx 1,97$, sowie $\sigma(B) \approx 1,40$.

3. Es gilt $C(\Omega) = \{-1, -\frac{1}{2}, 1\}$ mit $P(C=-1) = \frac{18}{37} = P(C=1)$, $P(C=-\frac{1}{2}) = \frac{1}{37}$
   $\Rightarrow E(C) = \frac{18}{37}(-1) + \frac{18}{37}1 + \frac{1}{37}(-\frac{1}{2}) = -\frac{1}{74}$, sowie
   $\operatorname{Var}(C) \approx 0,98$ und $\sigma(C) \approx 0,99$

Satz 8.5 *(Rechenregeln für die Varianz und Kovarianz)*
Seien $X, Y, X_1, \ldots, X_n : \Omega \to \mathbb{R}$ reellwertige Zufallsvariable, deren Varianz existiere.

1. $\operatorname{Var}(X) = \operatorname{Cov}(X, X)$

2. $\operatorname{Var}(X) = E(X^2) - (E(X))^2$

3. $\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y)$
   Sind X und Y unabhängig, so sind sie insbesondere unkorreliert.

4. *Für $a, b \in \mathbb{R}$ gilt* $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$.

5. *Für $a, b, c, d \in \mathbb{R}$ gilt* $\operatorname{Cov}(aX + b, cY + d) = ac \operatorname{Cov}(X, Y)$.

6. *Es gilt* $\operatorname{Var}(X_1 + \ldots + X_n) = \operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n) + \sum_{i \neq j} \operatorname{Cov}(X_i, X_j)$.
   Für unabhängige Zufallsvariablen $X_1, \ldots, X_n$ gilt also insbesondere
   $\operatorname{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \operatorname{Var}(X_i)$

Beweis:

1. $\operatorname{Cov}(X, X) = E((X - E(X))^2) = \operatorname{Var}(X)$
2. folgt wegen 1. aus 3.
3. $\operatorname{Cov}(X, Y) = E(XY) - E(X \cdot E(Y)) - E(E(X) \cdot Y) + E(E(X)E(Y)) =$
   $E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) =$
   $E(XY) - E(X)E(Y)$
4. folgt aus 5. mit $c = a$, $d = b$, $Y = X$
5. $\operatorname{Cov}(aX + b, cY + d) = E((aX + b - aE(X) - b)(cY + d - cE(Y) - d)) =$
   $E(a(X - E(X))c(Y - E(Y))) = ac \operatorname{Cov}(X, Y)$
6. Wir können o.E. $E(X_i) = 0$ für $i = 1, \ldots, n$ annehmen (vgl. 4. und 5.)
   $\Rightarrow \operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n) = E((X_1 + \ldots + X_n)^2) - (E(X_1 + \ldots + X_n))^2 =$
   $E\left(\sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j\right) - \underbrace{(E(X_1) + \ldots + E(X_n))}_{=0 \text{ nach Annahme}}{}^2 =$
   $\sum_{i=1}^{n} \sum_{j=1}^{n} E(X_i X_j) = \sum_{i=1}^{n} E(X_i^2) + \sum_{i \neq j} E(X_i X_j) =$

$\sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$

$\square$

Beispiel 8.6 (Varianzen klassischer Wahrscheinlichkeitsverteilungen)

1. Die Zufallsvariable $X : \Omega \to \mathbb{R}$ sei binomialverteilt mit Verteilung $b_{n,p}$, d.h. $X(\Omega) = \{0, \dots, n\}$ und $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.
   Dann gilt $E(X) = np$ (Beispiel 7.9) und
   $\mathrm{Var}(X) = E((X - E(X))^2) = \sum_{k=0}^{n} (k - np)^2 \binom{n}{k} p^k (1-p)^{n-k} \overset{\text{Beweis von 5.8}}{=} np(1-p)$.
2. Die Zufallsvariable $X : \Omega \to \mathbb{R}$ sei geometrisch verteilt mit Parameter $p \neq 0$, d.h. $X(\Omega) = \mathbb{N}_+$ mit $P(X = k) = p(1-p)^{k-1}$
   $\Rightarrow E(X) = \frac{1}{p}$ (vgl. Beispiel 7.9) und
   $\mathrm{Var}(X) = E((X - E(X))^2) = \sum_{k=1}^{\infty} (k - \frac{1}{p})^2 p (1-p)^{k-1} =$
   $p \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} - 2 \sum_{k=1}^{\infty} k (1-p)^{k-1} + \frac{1}{p} \sum_{k=1}^{\infty} (1-p)^{k-1} =$
   $$p \underbrace{\sum_{k=1}^{\infty} k(k+1)(1-p)^{k-1}}_{\frac{2}{p^3}} + (-p-2) \underbrace{\sum_{k=1}^{\infty} k(1-p)^{k-1}}_{\frac{1}{p^2} \text{ vgl. Beispiel 7.9}} + \frac{1}{p^2} = (*), \text{ denn}$$
   $\frac{1}{1-x} = 1 + x + x^2 + \dots$
   $\Rightarrow \frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \dots$
   $\Rightarrow 2\frac{1}{(1-x)^3} = 2 + 3 \cdot 2x + 4 \cdot 3x^2 + \dots$ (folgt jeweils durch Ableiten beider Seiten)
   $(*) = \frac{1}{p^2} - \frac{1}{p}$
3. Die Zufallsgröße $X : \Omega \to \mathbb{R}$ sei Poisson-verteilt mit Parameter λ, d.h. es gelte $X(\Omega) = \mathbb{N}$ und $P(X = k) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$.
   Dann gilt $E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \mathrm{e}^{-\lambda} = \lambda$ und
   $\mathrm{Var}(X) = \sum_{k=0}^{\infty} (k - \lambda)^2 \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \left(k^2 \frac{\lambda^k}{k!} - 2\lambda k \frac{\lambda^k}{k!} + \lambda^2 \frac{\lambda^k}{k!} \right) =$
   $\mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \left(k(k-1) \frac{\lambda^k}{k!} \right) + \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} - 2\lambda^2 + \lambda^2 =$
   $$\underbrace{\lambda^2 \, \mathrm{e}^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}}_{=1} + \lambda - \lambda^2 = \lambda$$

Beispiel 8.7 (Standardisierung von Zufallsvariablen)

Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable für die $E(X^2)$ existiere und für die $\sigma(X) \neq 0$

gelte (d.h. X sei nicht konstant).

Dann heißt die Zufallsvariable $X' = \frac{X - E(X)}{\sigma(X)} : \Omega \to \mathbb{R}$ die ***Standardisierung von X***.

Für die Standardisierung gilt:

$E(X') = \frac{1}{\sigma(X)}(E(X) - E(X)) = 0$ und $\operatorname{Var}(X') = \frac{1}{\sigma(X)^2} \operatorname{Var}(X) = 1$

Satz 8.8 *(Die Cauchy-Schwarzsche-Ungleichung)*

Seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariable für die $E(X^2)$ und $E(Y^2)$ existieren.

Dann gilt $(E(XY))^2 \leq E(X^2)E(Y^2)$.

Hierbei gilt Gleichheit genau dann, wenn es ein $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ gibt mit $P(aX + bY \neq 0) = 0$.

Beweis:

Sei $\alpha = E(Y^2)$ und $\beta = -E(XY)$.

1. Fall: $\alpha = 0$

$\Rightarrow \sum\limits_{\omega \in \Omega} Y(\omega)^2 P(\omega) = 0$

$\Rightarrow$ für alle $\omega \in \Omega$ gilt $Y(\omega) = 0$ oder $P(\omega) = 0$

$\Rightarrow E(XY) = 0$ Behauptung klar!

2. Fall: $\alpha > 0$

Es gilt $0 \leq E((\alpha X + \beta Y)^2) = \alpha^2 E(X^2) + 2\alpha\beta E(XY) + \beta^2 E(Y^2) =$
$\alpha(E(X^2)E(Y^2) - 2E(XY)^2 + E(XY)^2)$

$\Rightarrow E(X^2)E(Y^2) \geq E(XY)^2$

Hierbei gilt Gleichheit genau dann, wenn $E(Y^2) = 0$ (nicht möglich weil $\alpha > 0$) oder $E((\alpha X + \beta Y)^2) = 0$ ist. In diesem Fall gibt es ein $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ mit $E((aX + bY)^2) = 0$.

Es folgt, dass für alle $\omega \in \Omega$ gilt:

$P(\omega) = 0$ oder $aX(\omega) + bY(\omega) = 0$

$\Rightarrow P(aX + bY = 0) = 1$

$\Rightarrow P(aX + bY \neq 0) = 0$

Umgekehrt zeigt $P(aX + bY = 0) = 1$, dass für alle $\omega \in \Omega$ gilt $P(\omega) = 0$ oder $aX(\omega) + bY(\omega) = 0$.

O.E. sei dazu $a \neq 0$.

$$\Rightarrow E(XY)^2 = \left(\sum_{\{\omega \in \Omega | P(\omega) \neq 0\}} X(\omega)Y(\omega)P(\omega) \right)^2 =$$
$$\left(\sum_{\{\omega \in \Omega | P(\omega) \neq 0\}} -\frac{b}{a} Y(\omega)^2 P(\omega) \right)^2 = \left(-\frac{b}{a} E(Y^2)\right)^2 \text{ und}$$
$$E(X^2) = \sum_{\{\omega \in \Omega | P(\omega) \neq 0\}} X^2(\omega)P(\omega) = \frac{b^2}{a^2} \sum_{\{\omega \in \Omega | P(\omega) \neq 0\}} Y^2(\omega)P(\omega) = \frac{b^2}{a^2} E(Y^2)$$

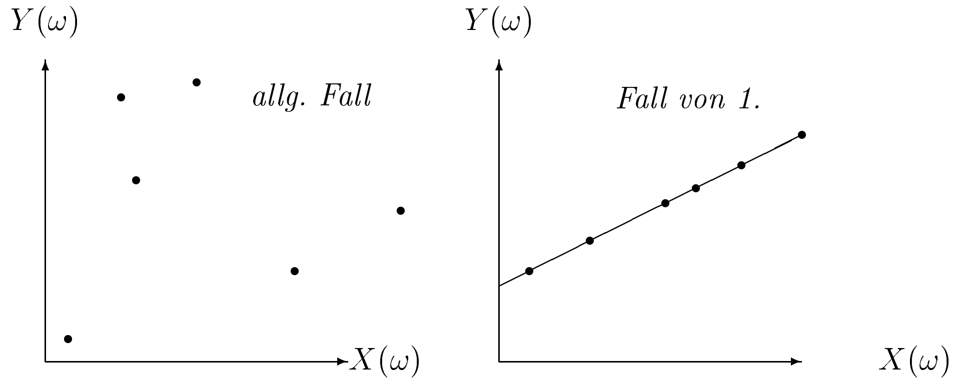
$\Rightarrow E(XY)^2 = E(X^2)E(Y^2)$ $\square$

Korollar 8.9 *Seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariable für die $E(X^2)$ und $E(Y^2)$ existieren.*
Dann gilt $0 \leq |\operatorname{Cov}(XY)| \leq \sigma(X)\sigma(Y)$.
Im Fall $\sigma(X)\sigma(Y) \neq 0$ folgt insbesondere $-1 \leq \rho(X, Y) \leq 1$.

Beweis:
Wende 8.8 an mit $X - E(X)$ und $Y - E(Y)$.
$\Rightarrow \operatorname{Cov}(X, Y)^2 \leq \operatorname{Var}(X) \cdot \operatorname{Var}(Y)$ □

Bemerkung 8.10 *(Anschauliche Bedeutung der Kovarianz)*

1. *Gilt $|\operatorname{Cov}(X, Y)| = \sigma(X)\sigma(Y)$, also $\rho(X, Y) \in \{\pm 1\}$, so gibt es $(a, b) \in \mathbb{R}^2 \setminus \{0\}$ und $c \in \mathbb{R}$ mit $P(aX + bY + c = 0) = 1$.*



2. *Aus $\rho(X, Y) = 0$ folgt nicht, dass X und Y unabhängig sind.*
   Sei z.B. $\Omega = \{1, 2, 3, 4\}$ mit $P(1) = P(2) = \frac{2}{5}$, $P(3) = P(4) = \frac{1}{10}$
   und $X : \Omega \to \mathbb{R}$ mit $1 \mapsto 1,\ 2 \mapsto -1,\ 3 \mapsto 2,\ 4 \mapsto -2$ und $Y : \Omega \to \mathbb{R}$ mit $1 \mapsto -1,\ 2 \mapsto 1,\ 3 \mapsto 2,\ 4 \mapsto -2$
   $\Rightarrow E(X) = E(Y) = 0$
   $\operatorname{Cov}(XY) = E(XY) = (-1)\frac{2}{5} + (-1)\frac{2}{5} + 4\frac{1}{10} + 4\frac{1}{10} = 0$, *aber X, Y sind nicht unabhängig:*
   $P(X = 1, Y = -1) = \frac{2}{5} \neq P(X = 1)P(Y = -1) = \frac{2}{5} \cdot \frac{2}{5} = \frac{4}{25}$

Korollar 8.11 *Ist $X : \Omega \to \mathbb{R}$ eine Zufallsvariable für die $E(X^2)$ existiert, so gilt $E(|X - E(X)|) \leq \sigma(X)$.*

Beweis:
Wende 8.8 an mit $|X - E(X)|$ und $Y = 1$
$\Rightarrow E(|X - E(X)|)^2 \leq E((X - E(X))^2) = \operatorname{Var}(X)$
$\Rightarrow E(|X - E(X)|) \leq \sigma(X)$ □

Satz 8.12 *(Die Tschebyschew-Ungleichung)*
Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable für die $E(X^2)$ existiert und sei $\varepsilon > 0$.
$P(|X - E(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$.

Beweis:
Definiere $Y : \Omega \to \mathbb{R}$ durch $\omega \to \begin{cases} 0 & \text{falls } |X(\omega) - E(X)| < \varepsilon \\ \varepsilon^2 & \text{falls } |X(\omega) - E(X)| \geq \varepsilon \end{cases}$
$\Rightarrow$ Für alle $\omega \in \Omega$ gilt $Y(\omega) \leq (X - E(X))^2(\omega)$.
$\Rightarrow \mathrm{Var}(X) = E((X - E(X))^2) \geq E(Y) = \varepsilon^2 P(Y = \varepsilon^2) = \varepsilon^2 P(|X - E(X)| \geq \varepsilon)$
□

Satz 8.13 *(Das schwache Gesetz der großen Zahlen)*
Seien $X_1, \dots, X_n : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen mit $E := E(X_1) = E(X_2) = \dots = E(X_n)$.
Für $i = 1, \dots, n$ existiere $E(X_i^2)$ und sei $\varepsilon > 0$ vorgegeben.
Dann gilt für die Zufallsvariable $H_n = \frac{X_1 + \ldots + X_n}{n}$ die Ungleichung
$P(|H_n - E| \geq \varepsilon) \leq \frac{\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)}{n^2 \varepsilon^2}$.
Insbesondere folgt $P(|H_n - E| \geq \varepsilon) \to 0$ für $n \to \infty$, wenn es ein $M > 0$ gibt mit $\mathrm{Var}(X_i) \leq M$ für $i = 1, \dots, n$.

Beweis:
Nach 8.5 Nr.4 und 6 gilt $\mathrm{Var}(H_n) = \frac{1}{n^2}(\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n))$, so dass sich die Behauptung aus 8.12 ergibt. □

Beispiel 8.14 Eine zeitlich begrenzte Spekulation werde durch den folgenden Wahrscheinlichkeitsraum beschrieben:
$\Omega = \{0, 1\}$, $P =$Gleichverteilung, $k_0 = 1$ Grundkapital, $X_n : \Omega^n \to \mathbb{R}$ Kapital nach n Spekulationen.
Einzelne Spekulation: $Y_n : \Omega \to \mathbb{R}$ sei def. durch $0 \mapsto \frac{1}{2}$ und $1 \mapsto \frac{5}{3}$.
$\Rightarrow X_n = Y_n \cdot Y_{n-1} \cdot \ldots \cdot Y_1$
$\Rightarrow E(Y_n) = \frac{1}{2}\left(\frac{1}{2}\right) + \frac{1}{2}\left(\frac{5}{3}\right) = \frac{13}{12}$
Die Variablen $Y_1, \dots, Y_n$ sind unabhängig.
$\Rightarrow E(X_n) = E(Y_n)^n$
$\Rightarrow E(Y_n) = \left(\frac{13}{12}\right)^n \to \infty$ für $n \to \infty$.
JEDOCH:
Sei $\mu := E(\ln(Y_n)) = \frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{5}{3} \approx -0,1 < 0$
Für $n \geq 1$ gilt auch $\mathrm{Var}(\ln(Y_n)) = \frac{1}{2}(\ln\frac{1}{2} - \mu)^2 + \frac{1}{2}(\ln\frac{5}{3} - \mu)^2 \approx 0,362$

Setze $\varepsilon := -\frac{\mu}{2} \approx 0,046 > 0$
Nach dem schwachen Gesetz der großen Zahlen gilt:
$P\left(\left|\frac{1}{n}\left(\ln(Y_1) + \ldots + \ln(Y_n)\right) - \mu\right| < \varepsilon\right) \geq 1 - \frac{\mathrm{Var}(\ln(Y_1))}{n\varepsilon^2} \approx 1 - \frac{171,1}{n}$
$\Rightarrow \left(\left|\frac{1}{n}\ln(X_n) - \mu\right| < -\frac{\mu}{2}\right) \to 1$ für $n \to \infty$
$\Rightarrow P\left(\ln X_n < \frac{n\mu}{2}\right) \geq P\left(\frac{3n\mu}{2} < \ln(X_n) < \frac{n\mu}{2}\right) \to 1$
$\Rightarrow P(X_n < \underbrace{\mathrm{e}^{\frac{n\mu}{2}}}_{\to 0 \text{ für } n \to \infty, \text{ denn } \mu < 0} \to 1$
$\Rightarrow$„fast sicher“ist bei $n \gg 0$ das Kapital weg.

9 Erzeugende Funktionen

In diesem Abschnitt sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ sei eine reellwertige Zufallsvariable mit $X(\Omega) \subset \mathbb{N}$.

Definition 9.1 Die Potenzreihe $g_X(t) = \sum\limits_{k=0}^{\infty} P(X = k)t^k \in \mathbb{R}[[t]]$ heißt die ***erzeugende Funktion von X***. (Die Doppelklammer bedeutet Potenzreihe.)

Bemerkung 9.2

1. *Sei für $t > 0$ eine Zufallsvariable $Y_t : \Omega \to \mathbb{R}$ def. durch $Y_t(\omega) = t^{X(\omega)}$ für $\omega \in \Omega$.*
   Dann gilt $g_X(t) = E(Y_t)$.
2. *Wegen $P(X = k) \leq 1$ für $k \in \mathbb{N}$ konvergiert die Reihe $g_X(t)$ für $|t| < 1$.*
3. *Für alle $k \in \mathbb{N}$ gilt $P(X = k) = \frac{1}{k!}g_X^{(k)}(0)$.*
   Also ist der Übergang von P_X zu g_X eineindeutig.

Beispiel 9.3 (Erzeugendenfunktion der klassischen Wahrscheinlichkeitsfunktionen)

1. Die erzeugende Funktion einer binomial verteilten Zufallsvariable $X : \Omega \to \mathbb{R}$ mit $P_X = b_{n,p}$ ist
   $g_X(t) = \sum\limits_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} t^k =$
   $\sum\limits_{k=0}^{n} \binom{n}{k} (pt)^k (1-p)^{n-k} = (pt + 1 - p)^n$.
2. Ist $X : \Omega \to \mathbb{R}$ geometrisch verteilt mit Parameter p, so gilt
   $g_X(t) = \sum\limits_{k=1}^{\infty} p(1-p)^{k-1} t^k = \frac{pt}{1-(1-p)t}$.

3. Ist $X : \Omega \to \mathbb{R}$ Poisson-verteilt mit Parameter λ, so gilt
   $g_X(t) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} t^k = \mathrm{e}^{\lambda(t-1)}$

Satz 9.4 *(Berechnung von Erwartungswert und Varianz mit Hilfe erzeugender Funktionen)*

1. *Die Funktion* $g_X : [0,1] \to \mathbb{R}$, *def. durch* $t \mapsto \sum_{k=0}^{\infty} P(X = k)t^k$ *ist stetig, monoton wachsend, konvex und erfüllt* $g_X(0) = P(X = 0)$, *sowie* $g_X(1) = 1$.
2. *Für alle* $k \in \mathbb{N}_+$ *gilt* $E(X(X-1) \cdot \ldots \cdot (X-k+1)) = \lim_{t \lessdot 1} g_X^{(k)}(t)$.
   Hierbei konvergiert die linke Seite genau dann, wenn der Limes auf der rechten Seite existiert und beide Bedingungen sind äquivalent damit, dass $E(X^k)$ *existiert.*
3. *Es gilt* $E(X) = \lim_{t \lessdot 1} g'_X(t)$ *sowie* $\mathrm{Var}(X) = \lim_{t \lessdot 1} g''_X(t) + E(X) - E(X)^2$.

Beweis:

1. klar, wegen $P(X = k) \geq 0$
2. Für $0 \leq t < 1$ gilt $g_X^{(k)}(t) = \sum_{n=k}^{\infty} n \cdot (n-1) \cdot \ldots \cdot (n-k+1) P(X = n) t^{n-k} \leq \sum_{n=0}^{\infty} n^k P(X = n) = E(X^k)$.
   Also folgt aus $E(X^k) < \infty$, dass die Reihe $g_X^{(k)}(t)$ gleichmäßig konvergiert und die Behauptung folgt für $t \lessdot 1$.
   Gilt $g_X^{(k)}(t) < \infty$, so folgt aus $n \cdot (n-1) \cdot \ldots \cdot (n-k+1) \geq \frac{n^k}{2^k}$ für $n \geq 2k$ auch $\sum_{n=2k}^{\infty} n^k P(X = n) \leq 2^k \sum_{n=2k}^{\infty} n \cdot (n-1) \cdot \ldots \cdot (n-k+1) P(X = n)$ und somit $E(X^k) < \infty$.
   Schließlich liefert $E(X \cdot (X-1) \cdot \ldots \cdot (X-k+1)) = \sum_{n=0}^{\infty} n \cdot (n-1) \cdot \ldots \cdot (n-k+1) P(X = n) < \infty$ auch $g_X^{(k)} \leq E(X \cdot (X-1) \cdot \ldots \cdot (X-k+1))$ für $0 \leq t < 1$ und somit $\lim_{t \lessdot 1} g_X^{(k)}(t) < \infty$.
3. Die erste Behauptung folgt aus 2. mit $k = 1$ und die zweite aus
   $\mathrm{Var}(X) = E(X^2) - E(X)^2 = E(X(X-1)) + E(X) - E(X)^2 \overset{2. \text{ mit } k=2}{=} \lim_{t \lessdot 1} g''_X(t) + E(X) - E(X)^2$.

$\square$

Beispiel 9.5 (Wartezeit auf einen „Run“)
Für jedes $n \in \mathbb{N}$ sei $X_n : \Omega \to \mathbb{R}$ Bernoulli-verteilt zum Parameter $p \in [0,1]$.

Ferner seien $X_1, X_2, \ldots$ unabhängig.

Gilt für ein $\omega \in \Omega$ ein $k \in \mathbb{N}$, die Gleichungskette $X_k(\omega) = X_{k+1}(\omega) = \ldots = X_{k+r-1}(\omega) = 1$, so heißt die Folge $X_k(\omega), \ldots, X_{k+r-1}(\omega)$ ein ***Run der Länge r***.

Für jedes $\omega \in \Omega$ sei $T(\omega) = \inf\{k + r - 1 \mid X_k(\omega) = \ldots = X_{k+r-1}(\omega)\}$ der erste Index, für den die Folge $X_1(\omega), \ldots, X_{T(\omega)}(\omega)$ einen Run der Länge r enthält.

Es ist der Erwartungswert der Zufallsvariablen $T : \Omega \to \mathbb{R}$ zu bestimmen.

Es gilt $P(T = k) = 0$ für $k < r$ und $P(T = r) = p^r$.

Im Fall $k > r$ gilt $P(T = k) = \sum\limits_{j=1}^{r} P(X_1 = 1, \ldots, X_{j-1} = 1, X_j = 0, T = k)$.

Gilt $X_1(\omega) = \ldots = X_{j-1}(\omega) = 1$, $X_j(\omega) = 0$, $X_{k+r-1}(\omega) = \ldots = X_k(\omega) = 1$, so ist $X_{j+1}(\omega), \ldots, X_k(\omega)$ eine Folge mit $T(\omega) = k - j$.

Also folgt $P(X_1 = 1, \ldots, X_{j-1} = 1, X_j = 0, T = k) = p^{j-1}(1-p)P(T = k - j)$ und es ergibt sich die Rekursionsformel

$P(T = k) = \sum\limits_{j=1}^{r} p^{j-1}(1-p)P(T = k - j)$.

Für $g_T(t)$ ergibt sich

$$g_T(t) = \sum_{k=0}^{\infty} P(T = k)t^k = p^r t^r + \sum_{k>r} \sum_{j=1}^{r} p^{j-1}(1-p)P(T = k - j)t^k =$$

$$p^r t^r + (1-p)t \sum_{j=1}^{r} t^{j-1} p^{j-1} \underbrace{\sum_{k>j} t^{k-j} P(T = k - j)}_{g_T(t)} =$$

$$p^r t^r + (1-p)t g_T(t) \tfrac{1-t^r p^r}{1-tp}$$

$$\Rightarrow g_T(t) = \tfrac{t^r p^r (1-tp)}{1-t+t^{r+1}p^r(1-p)}$$

$$\Rightarrow E(T) = \lim_{t \nearrow 1} g'_T(t) = \tfrac{1-p^r}{p^r(1-p)}$$

Tabelle für $E(T)$ bei 1 Spiel pro Sekunde:

r	$p = \frac{1}{2}$	$p = \frac{1}{6}$
5	1 min.	2,6 Std.
10	34 min.	28 Monate
15	18 Std.	$18 \cdot 10^3$ Jahre
20	24 Tage	$141 \cdot 10^6$ Jahre

Satz 9.6 *(Rechenregeln für erzeugende Funktionen)*

1. *Seien $X, Y : \Omega \to \mathbb{N}$ unabhängige Zufallsvariablen.*
   Dann gilt $g_{X+Y}(t) = g_X(t)g_Y(t)$.
2. *Seien $T, X_1, X_2, \ldots : \Omega \to \mathbb{N}$ unabhängige Zufallsvariablen für die $X_1, X_2, \ldots$ die gleiche Verteilung haben und sei $S : \Omega \to \mathbb{N}$ def. durch $\omega \mapsto X_1(\omega) + \ldots +$*

$X_{T(\omega)}(\omega)$.

Dann gilt $g_S(t) = g_T(g_{X_i}(t))$.

3. *In der Situation von 2. gilt $E(S) = E(T) \cdot E(X_i)$ „Wald'sche Identität"*
   Existieren $E(T^2)$ und $E(X_i^2)$, so gilt auch
   $\mathrm{Var}(S) = \mathrm{Var}(T) \cdot E(X_i)^2 + E(T) \cdot \mathrm{Var}(X_i)$.

Beweis:

1. Es gilt $g_{X+Y}(t) = \sum\limits_{k=0}^{\infty} P(X+Y=k)t^k =$
   $\sum\limits_{k=0}^{\infty}\sum\limits_{i=0}^{k} P(X=i, Y=k-i)t^k =$
   $\sum\limits_{k=0}^{\infty}\sum\limits_{i=0}^{k}(P(X=i)t^i)(P(Y=k-i)t^{k-i}) =$
   $\left(\sum\limits_{i=0}^{\infty} P(X=i)t^i\right)\left(\sum\limits_{j=0}^{\infty} P(Y=j)t^j\right) =$
   $g_X(t) \cdot g_Y(t)$
2. Für $n \geq 1$ und die Zufallsvariable $Y_n = X_1 + \ldots + X_n$ gilt $g_{Y_n}(t) = g_{X_i}(t)^n$ nach 1. .
   Da T und Y_n unabhängig sind, folgt für $k \geq 0$ die Gleichung
   $P(S=k) = \sum\limits_{n=0}^{\infty} P(T=n, Y_n=k) = \sum\limits_{n=0}^{\infty} P(T=n)P(Y_n=k)$
   Hieraus ergibt sich
   $g_S(t) = \sum\limits_{k=0}^{\infty} P(S=k)t^k = \sum\limits_{n=0}^{\infty} P(T=n)\sum\limits_{k=0}^{\infty} P(Y_n=k)t^k =$
   $\sum\limits_{n=0}^{\infty} P(T=n)g_{X_i}(t)^n = g_T(g_{X_i}(t))$.
3. Aus 2. und der Kettenregel folgt
   $g'_S(t) = g'_T(g_{X_i}(t)) \cdot g'_{X_i}(t)$ und die Behauptung für den Erwartungswert ergibt sich aus $\lim\limits_{t \lesssim 1} g_{X_i}(t) = 1$ sowie Satz 9.4 3. .
   Nochmaliges Differenzieren liefert
   $g''_S(t) = g'_T(g_{X_i}(t)) \cdot g''_{X_i}(t) + g''_T(g_{X_i}(t)) \cdot g'_{X_i}(t)^2$
   Also folgt:
   $\lim\limits_{t \lesssim 1} g''_S(t) = E(T) \cdot \lim\limits_{t \lesssim 1} g''_{X_i}(t) + \lim\limits_{t \lesssim 1} g''_T(t) \cdot E(X_i)^2$
   Mittels Satz 9.4 3. erhalten wir
   $\mathrm{Var}(S) = \lim\limits_{t \lesssim 1} g''_S(t) + E(S) - E(S)^2 =$
   $E(T)E(X_i) - E(T)^2E(X_i)^2 + E(T)(\mathrm{Var}(X_i) - E(X_i) + E(X_i)^2) +$
   $E(X_i)^2(\mathrm{Var}(T) - E(T) + E(T)^2) =$
   $\mathrm{Var}(T)E(X_i)^2 + E(T)\,\mathrm{Var}(X_i)$

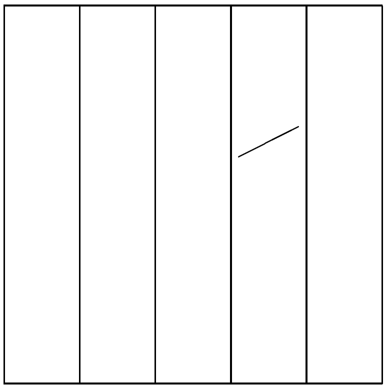
□

Beispiel 9.7 Sind $X, Y : \Omega \to \mathbb{R}$ binomialverteilt mit $P_X = b_{m,p}$ und $P_Y = b_{n,p}$ für $m, n \in \mathbb{N}$ und $p \in [0,1]$, so ist die Zufallsvariable $X + Y : \Omega \to \mathbb{R}$ binomialverteilt mit $P_{X+Y} = b_{m+n,p}$, denn $g_{X+Y} = g_X(t) \cdot g_Y(t) = (pt + 1 - p)^m (pt + 1 - p)^n = (pt + 1 - p)^{m+n}$.

Beispiel 9.8 Ein radioaktives Präparat sendet in der Minute T Teilchen aus. Jedes davon wird mit der Wahrscheinlichkeit p von einem Messgerät registriert.
Ω =Menge der Minuten
Sei $T : \Omega \to \mathbb{N}$, def. durch $\omega \mapsto$Anzahl der in Minute ω ausgestrahlten Teilchen.
Sei $X_i : \Omega \to \{0, 1\}$, def. durch $\omega \mapsto$das i-te in Minute ω ausgestrahlte Teilchen wird registriert(1) oder nicht(0).
$\Rightarrow$ T ist Poisson-verteilt zu einem Parameter λ und X_i ist Bernoulli-verteilt zum Parameter p.
$\Rightarrow$ für die Anzahl S der registrierten Teilchen gilt:
$g_S(t) = \mathrm{e}^{\lambda(g_{X_i}(t)-1)} = \mathrm{e}^{\lambda(-p+pt)} = \mathrm{e}^{p\lambda(t-1)}$
$\Rightarrow$ $S : \Omega \to \mathbb{R}$ ist wieder Poisson-verteilt zum Parameter $p\lambda$.

10 Allgemeine Wahrscheinlichkeitsräume

Beispiel 10.1 (Buffons Nadelproblem (1778))
Auf einen rechteckigen Teil der Ebene werden Linien im Abstand d gezeichnet. Dann wird eine Nadel der Länge $l < d$ zufällig geworfen, so dass der Mittelpunkt auf dem rechteckigen Teil landet. Mit welcher Wahrscheinlichkeit schneidet die Nadel eine der Linien ?



Beispiel 10.2 Eine Kreisscheibe werde in Sektoren unterteilt und die Sektoren mit verschiedenen Zahlen versehen. Diese Scheibe wird nun als „Glücksrad“ benutzt. Mit welcher Wahrscheinlichkeit zeigt der Zeiger beim Stillstand auf eine bestimmte Zahl ?

Bemerkung 10.3 *Die Menge der Versuchsergebnisse Ω soll auch überabzählbar viele Elemente haben dürfen.*
Eine Wahrscheinlichkeitsverteilung P soll wieder die folgenden Eigenschaften haben:
1. $P(\Omega) = 1$ *(Normiertheit)*
2. *Für $A_1, A_2, \ldots \subset \Omega$ paarweise disjunkt, gilt* $P\left(\bigcup\limits_{i\geq 1} A_i\right) = \sum\limits_{i\geq 1} P(A_i)$ *(Additivität)*

Jedoch stoßen wir auf folgendes Problem, wenn wir P auf ganz $\mathfrak{P}(\Omega)$ definieren wollen:
Satz von Banach-Kuratowski (1929):
Ist $\Omega = \mathbb{R}$ und $P : \mathfrak{P}(\Omega) \to [0,1]$ eine Abbildung mit den Eigenschaften 1. und 2., so gibt es ein Element $\omega \in \Omega$ mit $P(\{\omega\}) \neq 0$.
Insbesondere ist z.B. im Beispiel 10.2 die Abbildung P nicht ganz auf $\mathfrak{P}(\Omega)$ definierbar.

Definition 10.4

1. Sei Ω eine Menge. Eine Menge $\mathcal{A} \subset \mathfrak{P}(\Omega)$ heißt $\boldsymbol{\sigma}$ ***-Algebra in*** $\boldsymbol{\Omega}$, wenn gilt:
   1. $\Omega \in \mathcal{A}$
   2. $A \in \mathcal{A} \Rightarrow \Omega \setminus A \in \mathcal{A}$
   3. $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup\limits_{i \geq 1} A_i \in \mathcal{A}$
2. Ist Ω eine Menge und $\mathcal{A} \subset \mathfrak{P}(\Omega)$ eine σ-Algebra in Ω, so heißt das Paar $(\Omega, \mathcal{A})$ ein ***messbarer Raum*** (der ***Messraum***).
   Die Teilmengen $A \subset \Omega$ mit $A \in \mathcal{A}$ heißen $\boldsymbol{\mathcal{A}}$***-messbar***.
3. Ist $(\Omega, \mathcal{A})$ ein messbarer Raum und $P : \mathcal{A} \to [0, 1]$ eine Abbildung mit
   1. $P(\Omega) = 1$ (Normiertheit)
   2. $P\left(\bigcup\limits_{i \geq 1} A_i\right) = \sum\limits_{i \geq 1} P(A_i)$ für paarweise disjunkte $A_1, A_2, \ldots \in \mathcal{A}$ (Additivität)

   so heißt $(\Omega, \mathcal{A}, P)$ ein ***Wahrscheinlichkeitsraum***.
   Die Abbildung $P : \Omega \to [0, 1]$ heißt ***Wahrscheinlichkeitsverteilung*** oder ***Wahrscheinlichkeitsmaß*** und die Teilmengen $A \subset \Omega$ mit $A \in \mathcal{A}$ heißen ***Ereignisse***.

Satz 10.5 *(Eigenschaften von σ-Algebren)*
Sei Ω eine Menge.
1. *Ist $\mathcal{A} \subset \mathfrak{P}(\Omega)$ eine σ-Algebra in Ω, und sind $A_1, A_2, \ldots \in \mathcal{A}$, so gilt $\bigcap\limits_{i \geq 1} A_i \in \mathcal{A}$.*
2. *Ist I eine nichtleere Menge und sind $\mathcal{A}_i \subset \mathfrak{P}(\Omega)$ für $i \in I$ vorgebene σ-Algebren von Ω, so ist auch $\mathcal{A} := \bigcap\limits_{i \in I} \mathcal{A}_i \subset \mathfrak{P}(\Omega)$ eine σ-Algebra.*
3. *Ist I eine Menge und $M = \{A_i\}_{i \in I} \subset \mathfrak{P}(\Omega)$ eine Menge von Teilmengen von Ω, so gibt es eine eindeutige kleinste σ-Algebra $\mathcal{A} \subset \mathfrak{P}(\Omega)$ mit $M \subset \mathcal{A}$.*
   Wir schreiben $\mathcal{A} = A(M)$ und nennen $A(M)$ die **von M erzeugte $\boldsymbol{\sigma}$-Algebra**.

Beweis:

1. Es gilt $\bigcap\limits_{i \geq 1} A_i = \Omega \setminus \bigcup\limits_{i \geq 1} (\Omega \setminus A_i) \in \mathcal{A}$.
2. Wegen $\Omega \in \mathcal{A}_i$ für jedes $i \in I$ gilt $\Omega \in \bigcap\limits_{i \in I} \mathcal{A}_i$.
   Ist $A \in \bigcap\limits_{i \geq 1} \mathcal{A}_i$, so gilt $\Omega \setminus A \in \mathcal{A}_i$ für alle $i \in I$, also auch $\Omega \setminus A \in \bigcap\limits_{i \in I} \mathcal{A}_i$.
   Schließlich seinen $A_1, A_2, \ldots \in \bigcap\limits_{i \in I} \mathcal{A}_i$.

Dann folgt $\bigcup_{j\geq 1} A_j \in \mathcal{A}_i$ für alle $i \in I$, also auch $\bigcup_{j\geq 1} A_j \in \bigcap_{i\in I} \mathcal{A}_i$.

3. Nach 2. ist $\bigcap_{\substack{\mathcal{A}\subset\mathfrak{P}(\Omega)\ \sigma\text{-Algebra}\\ \mathcal{A}\supset M}} \mathcal{A}$ eine σ-Algebra in Ω, die M enthält und die in jeder M enthaltenen σ-Algebra in Ω enthalten ist.

□

Beispiel 10.6 (verschiedene σ-Algebren)

1. Ist Ω eine Menge, so ist $\mathfrak{P}(\Omega)$ eine σ-Algebra in Ω.
2. Ist $n \geq 1$ und $\Omega = \mathbb{R}^n$, so sei
   $M = \{]a_1, b_1] \times \ldots \times]a_n, b_n] \mid a_i, b_i \in \mathbb{R}, a_i < b_i\} \subset \mathfrak{P}(\Omega)$.
   Dann heißt $\mathrm{B}(\mathbb{R}^n) := \mathrm{A}(M)$ die ***borelsche $\boldsymbol{\sigma}$-Algebra in $\mathbb{R}^n$*** und die Mengen $A \in \mathrm{B}(\mathbb{R}^n)$ heißen ***borelsche Mengen*** (oder ***Borel-messbare Mengen***).
   Für jede offene oder abgeschlossene Teilmenge $A \subset \mathbb{R}^n$ gilt $A \in \mathrm{B}(\mathbb{R}^n)$ (aus Analysis bekannt).

Satz 10.7 *(Das Lebesgue-Maß)*

Sei $n \in \mathbb{N}_+$. Es gibt genau eine Abbildung $\lambda : B(\mathbb{R}^n) \to \mathbb{R}$ mit den folgenden Eigenschaften:

1. *Für jedes $A \in B(\mathbb{R}^n)$ ist $\lambda(A) \geq 0$.*
2. *Sind $A_1, A_2, \ldots \in B(\mathbb{R}^n)$ paarweise disjunkt, so gilt $\lambda\left(\dot{\bigcup}_{i\geq 1} A_i\right) = \sum_{i\geq 1} \lambda(A_i)$.*
3. *Sind $a_i, b_i \in \mathbb{R}$ mit $a_i < b_i$ für $i = 1, \ldots, n$, so gilt $\lambda(]a_1, b_1] \times \ldots \times]a_n, b_n]) = (b_1 - a_1) \cdot \ldots \cdot (b_n - a_n)$.*

Dieser Satz ist aus der Analysis bekannt.

Beispiel 10.8 (Fortsetzung von Beispiel 10.2)

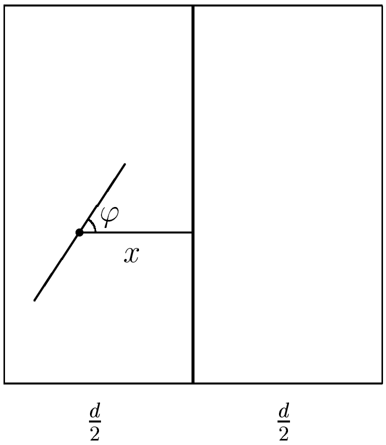
Sei $\Omega =]0, 2\pi]$ und $\mathcal{A} = \mathrm{B}(\mathbb{R}) \cap \mathfrak{P}(\Omega)$.

Durch $P = \frac{\lambda}{2\pi} : \mathcal{A} \to [0, 1]$ mit λ=Lebesguemaß auf $\mathbb{R}$ wird ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ definiert.

Entspricht eine Zahl auf dem Glücksrad z.B. einem Wertebereich $]\varphi_i, \varphi_{i+1}]$ mit $\varphi_i, \varphi_{i+1} \in]0, 2\pi]$ und $\varphi_i \leq \varphi_{i+1}$, so gilt $P(]\varphi_i, \varphi_{i+1}]) = \frac{1}{2\pi}(\varphi_{i+1} - \varphi_i)$.

Beispiel 10.9 (Fortsetzung von Beispiel 10.1)

Wir betrachten den Fall, dass die rechteckige Fläche folgendermaßen aussieht:



$\Rightarrow \Omega =]-\frac{d}{2}, \frac{d}{2}] \times]0, 2\pi] \subset \mathbb{R}^2$ und $\mathcal{A} = \mathrm{B}(\mathbb{R}^2) \cap \mathfrak{P}(\Omega)$
Ferner sei $P := \frac{\lambda}{2\pi d} : \mathcal{A} \to [0,1]$ mit $\lambda =$Lebesgue-Maß auf $\mathbb{R}^2$.
Die Nadel schneidet die Linie genau dann, wenn gilt:
$-\frac{l}{2}\cos\varphi \leq x \leq \frac{l}{2}\cos\varphi$, falls $\varphi \in]0, \frac{\pi}{2}] \cup]\frac{3\pi}{2}, 2\pi]$ oder
$\frac{l}{2}\cos\varphi \leq x \leq -\frac{l}{2}\cos\varphi$ falls $\varphi \in]\frac{\pi}{2}, \frac{3\pi}{2}]$
Es ergibt sich für das Ereignis
$A :=$„die Nadel schneidet die Linie“
$P(A) = \frac{1}{2\pi d} 8 \int\limits_0^{\frac{\pi}{2}} \frac{l}{2} \cos\varphi \, \mathrm{d}\varphi = \frac{2l}{\pi d}[\sin\varphi]_0^{\frac{\pi}{2}} = \frac{2l}{\pi d}$

Satz 10.10 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \ldots \in \mathcal{A}$ mit $A_1 \supset A_2 \supset \ldots$, sowie $B_1, B_2, \ldots \in \mathcal{A}$ mit $B_1 \subset B_2 \subset \ldots$.*
Dann gilt:

1. $P\left(\bigcap\limits_{i \geq 1} A_i\right) = \lim\limits_{i \to \infty} P(A_i)$

2. $P\left(\bigcup\limits_{i \geq 1} B_i\right) = \lim\limits_{i \to \infty} P(B_i)$

Beweis:

1. Für $n \geq 1$ gilt $A_n = \left(\bigcap\limits_{i \geq 1} A_i\right) \cup \left(\dot{\bigcup\limits_{i \geq n}} A_i \setminus A_{i+1}\right)$
   $\Rightarrow P(A_n) = P\left(\bigcap\limits_{i \geq 1} A_i\right) + \sum\limits_{i \geq n} P(A_i \setminus A_{i+1})$
   Da die Reihe $\sum\limits_{i \geq 1} P(A_i \setminus A_{i+1})$ konvergiert, gilt $\lim\limits_{n \to \infty} \sum\limits_{i \geq n} P(A_i \setminus A_{i+1}) = 0$.

Somit folgt $\lim\limits_{n\to\infty} P(A_n) = P\left(\bigcap\limits_{i\geq 1} A_i\right)$

2. Setze $A_i = \Omega \setminus B_i$ für $i \geq 1$
$\Rightarrow A_1 \supset A_2 \supset \dots$ absteigende Kette in $\mathcal{A}$ und $\bigcup\limits_{i\geq 1} B_i = \Omega \setminus \left(\bigcap\limits_{i\geq 1} A_i\right)$
Also folgt die Behauptung aus 1. .

□

11 Stetige Wahrscheinlichkeitsverteilungen

Definition 11.1

1. Eine Funktion $F : \mathbb{R} \to [0,1]$ heißt ***Verteilungsfunktion auf*** $\mathbb{R}$, wenn die folgenden Eigenschaften erfüllt sind:
   1. F ist rechtsstetig, d.h. für alle $x \in \mathbb{R}$ gilt $F(x) = \lim\limits_{t \overset{\geq}{\to} x} F(t)$.
   2. F ist monoton steigend, d.h. für $x_1 \leq x_2$ gilt $F(x_1) \leq F(x_2)$.
   3. $\lim\limits_{x\to-\infty} F(x) = 0$
   4. $\lim\limits_{x\to+\infty} F(x) = 1$
2. Eine Funktion $f : \mathbb{R} \to [0,\infty[$ heißt eine ***Dichte auf*** $\mathbb{R}$, wenn es eine endliche Menge $M \subset \mathbb{R}$ gibt, so dass $f|_{\mathbb{R}\setminus M}$ stetig ist und es gilt $\int\limits_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1$.

Satz 11.2 *Sei $(\Omega, \mathcal{A})$ der messbare Raum $(\mathbb{R}, B(\mathbb{R}))$.*

1. Ist $P : \mathcal{A} \to [0,1]$ ein Wahrscheinlichkeitsmaß, so wird durch $F : \mathbb{R} \to [0,1]$, def. durch $x \mapsto P(]-\infty, x])$, eine Verteilungsfunktion definiert.

2. Sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Dann gibt es ein eindeutig definiertes Wahrscheinlichkeitsmaß $P_F : \mathcal{A} \to [0,1]$ mit $P_F(]a,b]) = F(b) - F(a)$ für $a, b \in \mathbb{R}$ mit $a \leq b$.

3. Ist $f : \mathbb{R} \to [0,\infty[$ eine Dichte, so wird durch $F : \mathbb{R} \to [0,1]$, def. durch $x \mapsto \int\limits_{-\infty}^{x} f(t)\,\mathrm{d}t$, eine Verteilungsfunktion auf $\mathbb{R}$ definiert.
In diesem Fall heißt f eine ***Dichte zu F****.*

4. Ist $F : \mathbb{R} \to [0,1]$ eine stetige Verteilungsfunktion und gibt es eine endliche Menge $M \subset \mathbb{R}$, so dass $F_{\mathbb{R}\setminus M}$ stetig differenzierbar ist, so wird durch $f : \mathbb{R} \to [0,\infty[$, def. durch $x \mapsto \begin{cases} F'(x) & \text{falls } x \notin M \\ \text{beliebig} & \text{falls } x \in M \end{cases}$ eine Dichte zu F definiert.

Beweis:

1. Zu 1.:
   Sei $(x_n)_{n\in\mathbb{N}_+}$ eine Folge von $\mathbb{R}$ mit $x_1 \geq x_2 \geq \ldots$ und $\lim\limits_{n\to\infty} x_n = x$.
   Für $i \geq 1$ setze $A_i =]-\infty, x_i]$. Dann gilt $\bigcap\limits_{i\geq 1} A_i =]-\infty, x]$ und Satz 10.10 liefert $F(x) = P(]-\infty, x]) = P\left(\bigcap\limits_{i\geq 1} A_i\right) = \lim\limits_{i\to\infty} P(A_i) = \lim\limits_{i\to\infty} F(x_i)$.
   Zu 2.:
   Ist $x_2 \geq x_1$, so gilt $F(x_2) = P(]-\infty, x_2]) = F(x_1) + P(]x_1, x_2]) \geq F(x_1)$.
   Zu 3.:
   Sei $(x_i)_{i\geq 1}$ eine Folge in $\mathbb{R}$ mit $\lim\limits_{i\to\infty} x_i = -\infty$.
   Setze $A_i =]-\infty, x_i]$ und erhalte $\bigcap\limits_{i\geq 1} A_i = \emptyset$.
   Nach 10.10 folgt $\lim\limits_{i\to\infty} F(x_i) = \lim\limits_{i\to\infty} P(A_i) = P\left(\bigcap\limits_{i\geq 1} A_i\right) = 0$.
   Zu 4.:
   Sei $(x_i)_{i\geq 1}$ eine Folge in $\mathbb{R}$ mit $\lim\limits_{i\to\infty} x_i = \infty$.
   Setze $A_i =]-\infty, x_i]$ und erhalte $\bigcup\limits_{i\geq 1} A_i = \mathbb{R}$.
   Nach 10.10 folgt $\lim\limits_{i\to\infty} F(x_i) = \lim\limits_{i\to\infty} P(A_i) = P\left(\bigcup\limits_{i\geq 1} A_i\right) = P(\mathbb{R}) = 1$.
2. bekannt aus Maßtheorie
3. Zu 1.:
   Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}$ mit $\lim\limits_{n\to\infty} x_n = x$ und $x_1 \geq x_2 \geq \ldots$.
   Dann gilt: $F(x_n) = \int\limits_{-\infty}^{x_n} f(t)\,\mathrm{d}t \to \int\limits_{-\infty}^{x} f(t)\,\mathrm{d}t = F(x)$ für $n \to \infty$, denn $\left|\int\limits_{x}^{x_n} f(t)\,\mathrm{d}t\right| \overset{(*)}{\leq} (x_n - x) \max\limits_{x\leq t\leq x_n} |f(t)|$.
   (*): Für $n \gg 0$ sei o.E. $[x, x_n] \cap M = \emptyset$
   Zu 2.:
   Seien $x_1, x_2 \in \mathbb{R}$ mit $x_1 \leq x_2$.
   Aus $f(t) \geq 0$ für $t \in [x_1, x_2]$ folgt $\int\limits_{x_1}^{x_2} f(t)\,\mathrm{d}t \geq 0$ und somit $F(x_1) = \int\limits_{-\infty}^{x_1} f(t)\,\mathrm{d}t \leq \int\limits_{-\infty}^{x_2} f(t)\,\mathrm{d}t = F(x_2)$.
   Zu 3.:
   Sei $(x_n)_{n\geq 1}$ eine Folge in $\mathbb{R}$ mit $\lim\limits_{n\to\infty} x_n = -\infty$.
   Dann gilt $\lim\limits_{n\to\infty} \int\limits_{x_n}^{\infty} f(t)\,\mathrm{d}t = \int\limits_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1$.

Also folgt $\lim\limits_{n\to\infty} F(x_n) = \lim\limits_{n\to\infty} \int\limits_{-\infty}^{x_n} f(t)\,\mathrm{d}t = 1 - \lim\limits_{n\to\infty} \int\limits_{x_n}^{\infty} f(t)\,\mathrm{d}t = 0.$

Zu 4.:

Ist $(x_n)_{n\geq 1}$ eine Folge in $\mathbb{R}$ mit $\lim\limits_{n\to\infty} x_n = \infty$.

Dann gilt $\lim\limits_{n\to\infty} F(x_n) = \lim\limits_{n\to\infty} \int\limits_{-\infty}^{x_n} f(t)\,\mathrm{d}t = \int\limits_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1.$

4. Da $F|_{\mathbb{R}\setminus M}$ stetig differenzierbar ist, stellt $f|_{\mathbb{R}\setminus M} = F'|_{\mathbb{R}\setminus M}$ eine stetige Funktion dar.

   Ferner gilt

   $\int\limits_{-\infty}^{\infty} f(t)\,\mathrm{d}t = \lim\limits_{x\to\infty} \int\limits_{-\infty}^{x} F'(t)\,\mathrm{d}t = \lim\limits_{x\to\infty}[F(x) - \lim\limits_{y\to-\infty} F(y)] = \lim\limits_{x\to\infty} F(x) = 1$

□

Bemerkung 11.3

1. *Ist $F : \mathbb{R} \to [0,\infty[$ eine Dichte zu einer Verteilungsfunktion F, bzw. zu einem Wahrscheinlichkeitsmaß P_F, so gilt $P_F(\{a\}) = 0$ für alle $a \in \mathbb{R}$, denn* $P_F(\{a\}) \overset{\text{Satz 10.10}}{=} \lim\limits_{x \overset{<}{\to} a} P_F(]x,a]) = \lim\limits_{x \overset{<}{\to} a}(F(a) - F(x)) = \lim\limits_{x \overset{<}{\to} a} \int\limits_{x}^{a} f(t)\,\mathrm{d}t = 0.$
2. *Nicht jede Verteilungsfunktion besitzt eine Dichte.*

Beispiel 11.4 Die Funktion $F : \mathbb{R} \to [0,1]$ sei folgendermaßen definiert:

$x \mapsto 0$ für $x < 0$

$x \mapsto 1$ für $x \geq 1$

$x \mapsto \frac{1}{2}$ für $\frac{1}{3} \leq x < \frac{2}{3}$

$x \mapsto \frac{1}{4}$ für $\frac{1}{9} \leq x < \frac{2}{9}$

$x \mapsto \frac{3}{4}$ für $\frac{7}{9} \leq x < \frac{8}{9}$

usw., d.h. falls F auf einem Teilintervall $[a,b[$ von $[0,1[$ noch nicht definiert ist, so setze $F(x) = \frac{1}{2}(\lim\limits_{x \overset{<}{\to} a} F(x) + F(b))$ für $x \in [a + \frac{1}{3}(b-a), b - \frac{1}{3}(b-a)[$.

Dann ist F offenbar eine Verteilungsfunktion.

Die Funktion F hat abzählbar unendlich viele Unstetigkeitsstellen M.

Auf $\mathbb{R} \setminus M$ ist F stetig differenzierbar mit $F'(x) = 0$ für $x \in \mathbb{R} \setminus M$.

Besitzt F eine Dichte f, so muss also $f(x) = 0$ gelten für $x \in \mathbb{R} \setminus M$.

Dann wäre $F = 0$. Widerspruch !

Bei dem zu F gehörigen Wahrscheinlichkeitsmaß P_F gilt jedoch $P_F(\{a\}) = 0$ für alle $a \in \mathbb{R}$.

Beispiel 11.5 (Die Gleichverteilung)

Seien $a, b \in \mathbb{R}$ mit $a \leq b$ und sei $f : \mathbb{R} \to [0,\infty[$, def. durch

$$x \mapsto \begin{cases} \frac{1}{b-a} & \text{für} \quad a < x < b \\ 0 & \text{für} \quad x < a \text{ oder } x > b \\ \text{beliebig} & \text{für} \quad x \in \{a, b\} \end{cases}.$$

Dann ist f eine Dichte zu $F : \mathbb{R} \to [0,1]$, def. durch $x \mapsto \int\limits_{-\infty}^{x} f(t)\,\mathrm{d}t$. und die zugehörige Wahrscheinlichkeitsverteilung P_F heißt die ***Gleichverteilung auf*** $\boldsymbol{[a, b]}$.

Offenbar gilt $P_F(A) = \frac{1}{b-a}\lambda(A \cap [a,b])$ für $A \in \mathrm{B}(\mathbb{R})$.

Beispiel 11.6 Das Glücksrad wurde in Beispiel 10.8 beschrieben durch die Gleichverteilung auf $]0, 2\pi]$.

Beispiel 11.7 (Die Normalverteilung)
Sei $\mu \in \mathbb{R}$ und $\sigma \in [0, \infty[$.
Dann heißt die durch die Dichte $f : \mathbb{R} \to [0, \infty[$, def. durch $x \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, definierte Verteilungsfunktion $F : \mathbb{R} \to [0,1]$, bzw. die Wahrscheinlichkeitsverteilung $P_F : \mathrm{B}(\mathbb{R}) \to [0,1]$, die ***(eindimensionale) Normalverteilung mit Erwartungswert*** $\boldsymbol{\mu}$ ***und Standardabweichung*** $\boldsymbol{\sigma}$. Sie wird auch als $\boldsymbol{N(\mu, \sigma^2)}$***-Verteilung*** bezeichnet.
Die Verteilung $N(0,1)$ heißt die ***Standardnormalverteilung***.
Die Tatsache, dass f eine Dichte ist, folgt aus Satz 11.9.

Definition 11.8

1. Der Graph der Funktion $\varphi : \mathbb{R} \to \mathbb{R}$, def. durch $x \mapsto \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$, heißt die ***Gauß'sche Glockenkurve***.
2. Wir definieren auch die Funktion $\Phi : \mathbb{R} \to \mathbb{R}$ durch $x \mapsto \int\limits_{-\infty}^{x} \varphi(t)\,\mathrm{d}t$.

Satz 11.9

1. *Es gilt* $\int\limits_{-\infty}^{\infty} \varphi(t)\,\mathrm{d}t = 1$.
2. *Die Funktion Φ ist die Verteilungsfunktion der Standardnormalverteilung.*
3. *Die Dichte f der $N(\mu, \sigma^2)$-Verteilung erfüllt* $f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$.
4. *Für die Verteilungsfunktion F der $N(\mu, \sigma^2)$-Verteilung gilt*
   $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ *für alle* $x \in \mathbb{R}$.
   Insbesondere ergibt sich
   $P_F(]a,b]) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$ *für alle* $a, b \in \mathbb{R}$ *mit* $a \leq b$.

Beweis:

1. $\left(\int\limits_{-\infty}^{\infty} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t\right)^2 = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2+y^2}{2}}\,\mathrm{d}x\,\mathrm{d}y \overset{(*)}{=}$
   $\int\limits_{0}^{2\pi}\int\limits_{0}^{\infty} \mathrm{e}^{-\frac{r^2}{2}} r\,\mathrm{d}r\,\mathrm{d}\varphi = 2\pi[-\mathrm{e}^{-\frac{r^2}{2}}]_0^\infty = 2\pi$
   (*): Übergang zu Kugelkoordinaten !
2. Offenbar ist φ die Dichte von $N(0,1)$.
   Also ist Φ die Verteilungsfunktion von $N(0,1)$.
3. Es gilt $F(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$
4. Es gilt $F(x) = \int\limits_{-\infty}^{x} f(t)\,\mathrm{d}t = \int\limits_{-\infty}^{x} \frac{1}{\sigma}\varphi\left(\frac{t-\mu}{\sigma}\right)\,\mathrm{d}t \overset{(*)}{=}$
   $\int\limits_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(u)\,\mathrm{d}u = \Phi\left(\frac{x-\mu}{\sigma}\right)$
   (*): Substitution $u := \frac{t-\mu}{\sigma}$

□

Notation: $\mathbb{R}_+ :=]0,\infty[$

Satz 11.10 *(Stirlingsche Formel)*
Es gibt eine Zahl $\tau \in \mathbb{R}_+$ *mit* $\lim\limits_{n\to\infty} \frac{n!}{\tau n^{(n+\frac{1}{2})}\mathrm{e}^{-n}} = 1$
(Später wird $\tau = \sqrt{2\pi}$ *gezeigt.)*
Insbesondere gilt für $n \gg 0$ *die Approximation*
$n! \approx \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$.

Beweis:
Für $k = 1,\dots,n$ gilt $\int\limits_{k-\frac{1}{2}}^{k+\frac{1}{2}} \ln x\,\mathrm{d}x \overset{\ln x \text{ konkav}}{\leq} \ln k \overset{\ln x \text{ mon. steigend}}{\leq} \int\limits_{k}^{k+1} \ln x\,\mathrm{d}x$

Durch Summation ergibt sich $\int\limits_{\frac{1}{2}}^{n+\frac{1}{2}} \ln x\,\mathrm{d}x \leq \ln n! \leq \int\limits_{1}^{n+1} \ln x\,\mathrm{d}x$.
Nun ist $x\ln x - x$ eine Stammfunktion von $\ln x$, deshalb folgt weiter:
$(n+\frac{1}{2})\ln(n+\frac{1}{2}) - n - \frac{1}{2}\ln\frac{1}{2} \leq \ln n! \leq (n+1)\ln(n+1) - n$
Setze $d_n := \ln n! - (n+\frac{1}{2})\ln n + n$
$\Rightarrow d_n \geq -\frac{1}{2}\ln\frac{1}{2}$ und
$d_n - d_{n+1} = -\ln(n+1) + (n+\frac{1}{2})\ln n + n + (n+\frac{3}{2})\ln(n+1) - (n+1) =$
$-1 - (n+\frac{1}{2})\ln n + (n+\frac{1}{2})\ln(n+1) = \left(\ln\frac{n+1}{n}\right) - 1 =$
$\underbrace{\int\limits_{n}^{n+1} \ln x\,\mathrm{d}x}_{[x\ln x - x]_n^{n+1}} - \frac{\ln n + \ln(n+1)}{2} \overset{\ln x \text{ konkav}}{\geq} 0$

$\Rightarrow (d_n)_{n\in\mathbb{N}_+}$ ist monoton fallend und besitzt untere Schranke $-\frac{1}{2}\ln\frac{1}{2}$
$\Rightarrow \lim\limits_{n\to\infty} d_n$ existiert und wir können $\tau := \exp(\lim\limits_{n\to\infty} d_n)$ betrachten.
$\tau = \lim\limits_{n\to\infty} \frac{n!}{n^{n+\frac{1}{2}}\mathrm{e}^{-n}}$ □

Satz 11.11 *(Lokaler Grenzwert für die Binomialverteilung)*
Sei $0 < p < 1$. Für $n \geq 1$ sei $\sigma_n = \sqrt{np(1-p)}$ die Standardabweichung der $b_{n,p}$-Verteilung und $x_n(k) = \frac{k-np}{\sigma_n}$ für $k \in \mathbb{R}$.
Gilt $\lim\limits_{n\to\infty} \frac{x_n(k_n)^3}{\sqrt{n}} = 0$ für eine Folge $(k_n)_{n\geq 1}$ in $\mathbb{R}_+$, so folgt
$\lim\limits_{n\to\infty} \frac{b_{n,p}(k_n)}{\varphi(x_n(k_n))/\sigma_n} = \frac{\sqrt{2\pi}}{\tau}(=1)$

Beweis:
Für $n \geq 1$ setze $t_n := \frac{k_n}{n}$.
Dann gilt $\frac{x_n(k_n)^3}{\sqrt{n}} = \frac{(k_n-np)^3}{\sqrt{np(1-p)}^3\sqrt{n}} = \frac{n(t_n-p)^3}{\sqrt{p(1-p)}^3}$ und die Voraussetzung liefert
$\lim\limits_{n\to\infty} n(t_n-p)^3 = 0$.
Insbesondere folgt $\lim\limits_{n\to\infty} t_n = p$.
Aus der Stirlingschen Formel ergibt sich:
$b_{n,p}(k) = \frac{n!}{(k_n)!(n-k_n)!}p^{k_n}(1-p)^{n-k_n} \to$
$\frac{1}{\tau}\sqrt{\frac{n}{k_n(n-k_n)}}\frac{\mathrm{e}^{-n}}{\mathrm{e}^{-k_n}\mathrm{e}^{-n+k_n}}\frac{n^n}{k_n^{k_n}(n-k_n)^{n-k_n}}p^{k_n}(1-p)^{n-k_n} =$
$\frac{1}{\tau}\sqrt{\frac{n}{k_n(n-k_n)}}\left(\frac{np}{k_n}\right)^{k_n}\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}$ für $n \to \infty$.
Nun gilt $\sqrt{\frac{n}{k_n(n-k_n)}} = \frac{1}{\sqrt{nt_n(1-t_n)}} \to \frac{1}{\sqrt{np(1-p)}} = \frac{1}{\sigma_n}$ für $n \to \infty$.
Es ist nur noch Folgendes zu zeigen:
Für $\chi_n(k_n) := \left(\frac{np}{k_n}\right)^{k_n}\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}$ gilt $\lim\limits_{n\to\infty} \frac{\chi_n(k_n)}{\exp(-\frac{1}{2}x_n(k_n)^2)} = 1$
Es gilt $\ln\chi_n(k_n) = k_n \ln\left(\frac{p}{t_n}\right) + (n-k_n)\ln\left(\frac{1-p}{1-t_n}\right)$
$\Rightarrow -\ln\chi_n(k_n) = n\left[t_n\ln\left(\frac{t_n}{p}\right) + (1-t_n)\ln\left(\frac{1-t_n}{1-p}\right)\right]$
Betrachte die Funktion $g :]0,1[\to \mathbb{R}$, def. durch $t \mapsto t\ln\left(\frac{t}{p}\right) + (1-t)\ln\left(\frac{1-t}{1-p}\right)$
Es gilt: $g(p) = 0$, $g'(p) = 0$, $g''(p) = \frac{1}{p(1-p)}$
Taylorentwicklung von g bei $t = p$ liefert:
$g(t) = \frac{(t-p)^2}{2p(1-p)} + h(t-p)$ mit $|h(t-p)| \leq c|t-p|^3$ für ein $c > 0$ bei $t = p$.
Das heißt: Zu jedem $c > 0$ gibt es eine Umgebung U von p mit $R(t) \leq c|t-p|^3$ für alle $t \in U$.
In unserer Situation war $\lim\limits_{n\to\infty} n(t_n-p)^3 = 0$. Deshalb gilt
$-\ln\chi_n(k_n) \to \frac{n(t_n-p)^2}{2p(1-p)} = \frac{(k_n-np)^2}{2np(1-p)} = \frac{x_n(k_n)^2}{2}$ für $n \to \infty$
$\Rightarrow \lim\limits_{n\to\infty} \frac{\chi_n(k_n)}{\exp(-\frac{1}{2}\chi_n(k_n)^2)}$ □

Satz 11.12 *(Satz von Moivre-Laplace)*

Sei (Ω, P) ein (diskreter) Wahrscheinlichkeitsraum, sei $0 < p < 1$ und sei für jedes $n \geq 1$ eine $b_{n,p}$-verteilte Zufallsvariable $X_n : \Omega \to \mathbb{R}$ gegeben.
Ferner sei $X_n^* = \frac{X_n - E(X_n)}{\sigma(X_n)} = \frac{X_n - np}{\sqrt{np(1-p)}}$ für jedes $n \geq 1$ die Standardisierung von X_n.
Dann gilt für alle $a, b \in \mathbb{R}$ mit $a \leq b$ die Gleichung
$$\lim_{n\to\infty} P(a \leq X_n^* \leq b) = \underbrace{\frac{\sqrt{2\pi}}{\tau}}_{=1 \text{ später}} (\Phi(b) - \Phi(a)) = \underbrace{\frac{\sqrt{2\pi}}{\tau}}_{=1 \text{ später}} P_\Phi(]a, b])$$

Beweis:
Sei $\sigma_n := \sigma(X_n)$.
Für ein $\omega \in \Omega$ gilt $a \leq X_n^* \leq b$ genau dann, wenn $a\sigma_n + np \leq X_n(\omega) \leq b\sigma_n + np$ gilt.
Sei $\alpha_n = \min\{z \in \mathbb{Z} \mid z \geq a\sigma_n + np\}$ und $\beta_n = \max\{z \in \mathbb{Z} \mid z \leq b\sigma_n + np\}$.
Dann gilt also $\{\omega \in \Omega \mid a \leq X_n^*(\omega) \leq b\} = \{\omega \in \Omega \mid \alpha_n \leq X_n(\omega) \leq \beta_n\}$.
Betrachte nun $x_n(\alpha_n) = \frac{\alpha_n - np}{\sigma_n}$ und $x_n(\beta_n) = \frac{\beta_n - np}{\sigma_n}$.
Es gilt $|x_n(\alpha_n) - a| \leq \frac{1}{\sigma_n}$ und $|x_n(\beta_n) - b| \leq \frac{1}{\sigma_n}$
$\Rightarrow$ die Folgen $(x_n(\alpha_n))_{n\geq 1}$ und $(x_n(\beta_n))_{n\geq 1}$ sind beschränkt.
$\Rightarrow \lim\limits_{n\to\infty} \frac{x_n(\alpha_n)^3}{\sqrt{n}} = 0$, $\lim\limits_{n\to\infty} \frac{x_n(\beta_n)^3}{\sqrt{n}} = 0$
$\Rightarrow$ nach Satz 11.11 gilt
$\lim\limits_{n\to\infty} \frac{b_{n,p}(k_n)}{\varphi(x_n(k_n))/\sigma_n} = \frac{\sqrt{2\pi}}{\tau}$ für jede Folge $(k_n)_{n\geq 1}$ mit $\alpha_n \leq k_n \leq \beta_n$.
$\Rightarrow$ es gibt eine Folge $(\varepsilon)_{n\geq 1}$ in $\mathbb{R}_+$ mit $\lim\limits_{n\to\infty} \varepsilon_n = 0$, so dass gilt
(*) $1 - \varepsilon_n < \frac{b_{n,p}(k_n)}{\varphi(x_n(k_n))/\sigma_n} \frac{\tau}{\sqrt{2\pi}} < 1 + \varepsilon_n$
Betrachte nun die eingeschränkte Funktion $\varphi_{[x_n(\alpha_n - \frac{1}{2}), x_n(\alpha_n + \frac{1}{2})]} \to \mathbb{R}$.
Für $k = \alpha_n, \dots, \beta_n$ sind die Punkte $x_n(k)$ die Mittelpunkte von Teilintervallen der Länge $\frac{1}{\sigma_n}$ in die dieses Intervall aufgeteilt ist.
Die zugehörige Riemann-Summe von φ lautet $R_n = \sum\limits_{k=\alpha_n}^{\beta_n} \varphi(x_n(k)) \frac{1}{\sigma_n}$
Für $n \to \infty$ gilt $\left| \int\limits_{x_n(\alpha_n - \frac{1}{2})}^{x_n(\beta_n + \frac{1}{2})} \varphi(x)\,\mathrm{d}x - R_n \right| =$
$= |\Phi(x_n(\beta_n + \frac{1}{2})) - \Phi(x_n(\alpha_n - \frac{1}{2})) - R_n| \to 0$.
Wegen $\lim\limits_{n\to\infty} x_n(\alpha_n - \frac{1}{2}) = a$ und $\lim\limits_{n\to\infty} x_n(\beta_n + \frac{1}{2}) = b$ zeigt dies $\lim\limits_{n\to\infty} R_n = \Phi(b) - \Phi(a)$.
Summiert man die Gleichung (*) über $k = \alpha_n, \dots, \beta_n$, so folgt
$(1 - \varepsilon_n) \sum\limits_{k=\alpha_n}^{\beta_n} \varphi(x_n(k)) \frac{1}{\sigma_n} < \frac{\tau}{\sqrt{2\pi}} \sum\limits_{k=\alpha_n}^{\beta_n} P(X_n = k) < (1 + \varepsilon_n) \sum\limits_{k=\alpha_n}^{\beta_n} \varphi(x_n(k)) \frac{1}{\sigma_n}$
$\Rightarrow (1 - \varepsilon_n) R_n < \frac{\tau}{\sqrt{2\pi}} P(\alpha_n \leq X_n \leq \beta_n) < (1 + \varepsilon_n) R_n$
$\Rightarrow \lim\limits_{n\to\infty} P(a \leq X_n^* \leq b) = \lim\limits_{n\to\infty} P(\alpha_n \leq X_n \leq \beta_n) = \frac{\sqrt{2\pi}}{\tau} \lim\limits_{n\to\infty} R_n =$

$\frac{\sqrt{2\pi}}{\tau}(\Phi(b) - \Phi(a))$ $\square$

Korollar 11.13 *Es gilt* $\tau = \sqrt{2\pi}$.

Beweis:
Wähle $a = -b$ und $b \gg 0$ in Satz 11.12.
Ferner wähle α_n, β_n wie in Beweis von 11.12 mit
$\{\omega \in \Omega \mid -b \leq X_n^* \leq b\} = \{\omega \in \Omega \mid \alpha_n \leq X_n \leq \beta_n\}$
$\Rightarrow \lim\limits_{n\to\infty} x_n(\alpha_n) = -b$ und $\lim\limits_{n\to\infty} x_n(\beta_n) = b$
$\lim\limits_{n\to\infty} R_n = \Phi(b) - \Phi(-b)$
Nach der Tschebyschew-Ungleichung gilt
$1 - P(-b \leq X_n^* \leq b) \geq 1 - \frac{1}{b^2}$
$\overset{\text{Satz 11.12}}{\Rightarrow} \Phi(b) - \Phi(-b) = \frac{\tau}{2\pi} \lim\limits_{n\to\infty} P(-b \leq X_n^* \leq b) \geq \frac{\tau}{\sqrt{2\pi}}(1 - \frac{1}{b^2})$
$\Rightarrow 1 = \lim\limits_{b\to\infty}(\Phi(b) - \Phi(-b)) \geq \frac{\tau}{\sqrt{2\pi}}$
$\Rightarrow \tau \leq \sqrt{2\pi}$
Andererseits gilt $\Phi(b) - \Phi(-b) = \frac{\tau}{\sqrt{2\pi}} \lim\limits_{n\to\infty} \underbrace{P(-b \leq X_n^* \leq b)}_{\leq 1} \leq \frac{\tau}{\sqrt{2\pi}}$ und somit
$1 = \lim\limits_{n\to\infty}(\Phi(b) - \Phi(-b)) \leq \frac{\tau}{\sqrt{2\pi}}$
$\Rightarrow \tau \geq \sqrt{2\pi}$ $\square$

Beispiel 11.14 (Wahlvorhersage)
Es soll geschätzt werden, wieviele Prozent der Wähler eine bestimmte Partei wählen werden.
Wieviele Wähler müssen befragt werden, damit die Vorhersage mit 95% Wahrscheinlichkeit auf ±1% genau stimmt ?
Sei n die Anzahl der befragten Wähler.
$\Omega = \{0,1\}^n$ mit 1 entspricht: Wähler stimmt für die Partei.
$X_n : \Omega \to \mathbb{R}$ sei def. durch $(\omega_1, \ldots, \omega_n) \mapsto \omega_1 + \ldots + \omega_n$
$\Rightarrow X_n$ ist $b_{n,p}$-verteilt mit $p =$gesuchter Prozentsatz.
Wir brauchen $P\left(-0,01 \leq \frac{X_n}{n} - p \leq 0,01\right) = P(-0,01n \leq X_n - np \leq 0,01n) = P\left(\frac{-0,01n}{\sigma_n} \leq X_n^* \leq \frac{0,01n}{\sigma_n}\right) \geq 0,95$ mit $\sigma_n = \sqrt{np(1-p)}$.
$P\left(\frac{-0,01}{\sigma_n} \leq X_n^* \leq \frac{0,01}{\sigma_n}\right) \approx \Phi\left(\frac{0,01n}{\sigma_n}\right) - \Phi\left(\frac{-0,01n}{\sigma_n}\right) \overset{\varphi \text{ symmetrisch}}{=} 2\Phi\left(\frac{0,01n}{\sigma_n}\right) - 1 \geq 0,95$
$\Rightarrow \Phi\left(\frac{0,01n}{\sigma_n}\right) \geq 0,975$
$\Rightarrow \frac{0,01n}{\sigma_n} \geq 1,96$
$\Rightarrow n \geq p(1-p) \cdot 10000 \cdot (1,96)^2$.
Wegen $p(1-p) \leq \frac{1}{4}$ genügt es ungefähr 9600 Wähler zu befragen.

Beispiel 11.15 Ein Flugzeug habe 200 Passagiere.
Wieviele Reservierungen kann die Fluglinie akzeptieren, wenn einzelne Reservierungen mit 20% Wahrscheinlichkeit storniert werden und die Wahrscheinlichkeit einer Überbuchung $< 2,5\%$ sein soll ?
Sei n die Zahl der Reservierungen.
$\Omega = \{0,1\}^n$ mit 1 entspricht: Reservierung wurde wahrgenommen.
(Ω, P) ist n-ter Bernoulli-Raum mit Parameter $p = 0,8$.
$X_n : \Omega \to \mathbb{R}$ sei def. durch $(\omega_1, \dots, \omega_n) \mapsto \omega_1 + \ldots + \omega_n$ Zahl der wahrgenommenen Reservierungen.
$\Rightarrow X_n$ ist $b_{n,p}$-verteilt.
Es soll $P(X_n > 200) \leq 0,025$.
Äquivalent ist $P\left(X_n^* > \frac{200-np}{\sigma_n}\right) \leq 0,025$.
Es gilt $0,025 \geq P\left(X_n^* > \frac{200-np}{\sigma_n}\right) = 1 - P\left(X_n^* \leq \frac{200-np}{\sigma_n}\right) \approx 1 - \Phi\left(\frac{200-np}{\sigma_n}\right)$.
Also $\Phi\left(\frac{200-np}{\sigma_n}\right) \geq 0,975$.
$\Rightarrow 200 - np \geq \sigma_n \Phi^{-1}(0,975) \approx 1,96\sigma_n$
$\Rightarrow 0,8n + 0,784\sqrt{n} \leq 200$
$\Rightarrow \sqrt{n} \leq 15,33 \Rightarrow n \leq 235$
Probe:
Bei $n = 235$ gilt $P(X_n > 200) \approx 0,0208$
Bei $n = 236$ gilt $P(X_n > 200) \approx 0,285$
Also sind 235 Reservierungen möglich.

Definition 11.16 Sei $\lambda \in \mathbb{R}_+$.
Dann heißt die durch die Dichte $f : \mathbb{R} \to [0, \infty[$, def. durch
$x \mapsto \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{falls } x < 0 \end{cases}$, def. Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ bzw. die zugehörige Wahrscheinlichkeitsverteilung $P_\lambda : \mathrm{B}(\mathbb{R}) \to [0,1]$ die ***Exponentialverteilung zum Parameter $\boldsymbol{\lambda}$***.

Bemerkung 11.17

1. *Die Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ der Exponentialverteilung zum Parameter λ ist gegeben durch $F(x) = 0$ für $x < 0$ und $F(x) = \int_0^x f(t)\,\mathrm{d}t = \int_0^x \lambda e^{-\lambda t}\,\mathrm{d}t = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x}$ für $x \geq 0$.*

2. *Sei* $X : \mathbb{R} \to \mathbb{R}$ *geometrisch verteilt zum Parameter* $p \in]0,1[$. *Dann gilt für alle* $k \geq 1$ *die Gleichung*
$P(X \leq k) = \sum\limits_{i=1}^{k} P(X = i) = \sum\limits_{i=1}^{k} p(1-p)^{i-1} = p\frac{1-(1-p)^k}{1-(1-p)} = 1-(1-p)^k = 1 - \mathrm{e}^{k\ln(1-p)} = 1 - \mathrm{e}^{-\lambda k}$ *mit* $\lambda = -\ln(1-p) > 0$.
Also ist die Exponentialverteilung das Analagon der geometrischen Verteilung im Fall stetiger „Zeit".
Sie repräsentiert die Wartezeit bis zum Eintreten eines Ereignisses, das mit zeitlich konstanter Rate auftritt.
(z.B. radioaktiver Zerfall, Meteoriteneinschläge, ...)

Satz 11.18 *(Die Gedächtnislosigkeit der Exponentialverteilung)*
Sei $P : B(\mathbb{R}) \to [0,1]$ *die Exponentialverteilung zum Parameter* $\lambda > 0$.
Dann gelten für alle $x, y \geq 0$ *die Gleichungen*
1. $P(]x+y,\infty[|]x,\infty[) = P(]y,\infty[)$
2. $P(]-\infty,x+y] |]x,\infty[) = P(]-\infty,y])$

Interpretation:
Ist das Ereignis, dessen Wartezeit exponentiell verteilt ist, zum Zeitpunkt $x \geq 0$ *noch nicht eingetreten, so ist die Wahrscheinlichkeit, dass man noch mindestens* y *Zeiteinheiten warten muss, genauso groß wie zur Zeit 0.*

Beweis:

1. $P(]x+y,\infty[|]x,\infty) = \frac{P(]x+y,\infty[)}{P(]x,\infty[)} = \frac{1-P(]-\infty,x+y])}{1-P(]-\infty,x])} = \frac{1-(1-\mathrm{e}^{\lambda(x+y)})}{1-(1-\mathrm{e}^{\lambda x})} = \mathrm{e}^{-\lambda y} = 1 - P(]-\infty,y]) = P(]y,\infty[)$
2. $P(]-\infty,x+y] \, |]x,\infty[) = 1 - P(]x+y,\infty[|]x,\infty[) \overset{1.}{=} 1 - P(]y,\infty[) = P(]-\infty,y])$

$\square$

Beispiel 11.19 Die Zeit, die ein Verkäufer braucht um einen Kunden zu bedienen, sei exponential verteilt zum Parameter $\lambda = 0,8$.
Dann ist die Wahrscheinlichkeit, dass ein Verkaufsgespräch länger als 2 Minuten dauert, gegeben durch $P(]2,\infty[) = 1 - P(]-\infty,2]) = 1 - (1 - \mathrm{e}^{-0,8\cdot 2}) \approx 0,202$.

Satz 11.20 *Sei* $P : B(\mathbb{R}) \to [0,1]$ *eine Wahrscheinlichkeitsverteilung mit* $P(]-\infty,0]) = 0$.
Ferner sei P *gedächtnislos, d.h. für* $x, y \geq 0$ *mit* $P(]x,\infty[) > 0$ *gilt*

$P(]x+y,\infty[\,|\,]x,\infty[) = P(]y,\infty[)$.
Dann ist P eine Exponentialverteilung.

Beweis:
Für jedes $t \in \mathbb{R}$ sei $G(t) = 1 - F_P(t) = P(]t,\infty[)$.
Für $x, y \geq 0$ ist also $G(x+y) = G(x) \cdot G(y)$ gegeben. (*)
Für $t \leq 0$ gilt nach Voraussetzung $G(t) = 1$.
Für $n \in \mathbb{N}_+$ folgt aus (*) mit Induktion $G(n) = G(1)^n$.
Nach Satz 10.10 gilt $\lim\limits_{n\to\infty} G(\frac{1}{n}) = \lim\limits_{n\to\infty} P(]\frac{1}{n},\infty[) = P(]0,\infty[) = 1$.
Insbesondere gilt $G(\frac{1}{n}) > 0$ für genügend großes n und aus (*) folgt $G(\frac{1}{n}) = G(1)^{\frac{1}{n}}$.
Induktiv ergibt sich $G(q) = G(1)^q$ für alle $q \in \mathbb{Q}_+$.
Nun ist $G(t) = 1 - F_P(t)$ monoton fallend und rechtsstetig.
Also folgt $G(t) = G(1)^t$ für alle $t \in \mathbb{R}_+$.
Hierbei muss $0 < G(1)^t \leq 1$ gelten.
Wäre $G(1) = 1$, so wäre $F_P(t) = 0$ für alle $t > 0$, im Widerspruch zu $\lim\limits_{t\to\infty} F_P(t) = 1$.
Also gilt $\lambda := -\ln(G(1)) > 0$ und $G(t) = G(1)^t = \exp(\ln(G(1)^t)) = \exp(t\ln(G(1))) = \exp(t(-\lambda)) = \mathrm{e}^{-\lambda t}$ für $t \in \mathbb{R}_+$.
Dies liefert $F_P(t) = 1 - G(t) = 1 - \mathrm{e}^{-\lambda t}$. □

Definition 11.21 Sei $n \in \mathbb{N}_+$ und $(\Omega, \mathcal{A}) = (\mathbb{R}^n, \mathrm{B}(\mathbb{R}^n))$.

1. Ist $P : \mathcal{A} \to [0,1]$ eine Wahrscheinlichkeitsverteilung, so heißt $F_P : \mathbb{R}^n \to [0,1]$, def. durch $(x_1, \ldots, x_n) \mapsto P(\{(y_1, \ldots, y_n) \mid y_i \leq x_i \text{ für } i = 1, \ldots, n\}) = P(]-\infty, x_1] \times \ldots \times]-\infty, x_n])$, die ***zu P gehörige Verteilungsfunktion***.
2. Eine (Riemann-)integrierbare Funktion $f : \mathbb{R}^n \to [0,\infty[$ heißt eine ***Dichte***, wenn $\int\limits_{\mathbb{R}^n} f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdot \ldots \cdot \mathrm{d}x_n = 1$ gilt.
3. Eine Wahrscheinlichkeitsverteilung $P : \mathcal{A} \to [0,1]$ besitzt eine Dichte $f : \mathbb{R}^n \to [0,\infty[$, wenn f eine Dichte ist und für alle $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in \mathbb{R}^n$ mit $a_i \leq b_i$ für $i = 1, \ldots, n$ gilt:
$P(]a_1, b_1] \times \ldots \times]a_n, b_n]) = \int\limits_{a_1}^{b_1} \ldots \left(\int\limits_{a_n}^{b_n} f(x_1, \ldots, x_n)\,\mathrm{d}x_n \right) \ldots \mathrm{d}x_1$

Beispiel 11.22 (Gleichverteilung auf einem Gebiet)
Sei $n \in \mathbb{N}_+$ und sei $G \subset \mathbb{R}^n$ eine (lebesgue-)messbare Menge mit $\lambda(G) < \infty$.
Dann heißt die Wahrscheinlichkeitsverteilung $P : \mathrm{B}(\mathbb{R}^n) \to [0,1]$, def. durch $A \mapsto \frac{1}{\lambda(G)}\lambda(A \cap G)$, die ***Gleichverteilung auf G***.

Beispiel 11.23 (Bertrands Paradoxon)
Eine Sehne wird zufällig in einem Kreis mit Radius r und Mittelpunkt 0 gezogen. Mit welcher Wahrscheinlichkeit ist die Sehne (Sehnenlänge: s) länger als die Seite d des einbeschriebenen gleichseitigen Dreiecks ?

<u>Ansatz 1:</u>
o.E. sei ein Endpunkt der Sehne die Spitze des Dreiecks. Genau dann gilt $s > d$, wenn der andere Endpunkt im unteren Kreissegment liegt.
$\Rightarrow p = \frac{1}{3}$

<u>Ansatz 2:</u>
Sei M der Mittelpunkt der Sehne.
Es gilt $s > d$ genau dann, wenn $\overline{OM} < \frac{r}{2}$ ist.
Also $s > d \Leftrightarrow \overline{OM} \in [0, \frac{r}{2}[$ und $p = \frac{1}{2}$.

<u>Ansatz 3:</u>
Der Mittelpunkt der Sehne liegt zufällig im Kreis.
$s > d \Leftrightarrow \overline{OM} < \frac{r}{2} \Leftrightarrow M$ liegt im Kreis mit Radius $\frac{r}{2}$ um 0.
$\Rightarrow p = \frac{\text{Fläche des kleinen Kreises}}{\text{Fläche des großen Kreises}} = \frac{1}{4}$
Versuchsergebnisse:
Kreislinie entspricht $]0, 2\pi]$
Sehne entspricht $]0, 2\pi]^2 \setminus \{(x, x) \mid x \in]0, 2\pi]\}$

12 Stetige Zufallsvariable

Definition 12.1 Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{B})$ zwei messbare Räume.

1. Eine Abbildung $f : \Omega \to \Omega'$ heißt ***messbar***, wenn für alle $B \in \mathcal{B}$ gilt $f^{-1}(B) \in \mathcal{A}$.
2. Ist $(\Omega, \mathcal{A}, P)$ sogar ein Wahrscheinlichkeitsraum, so heißt eine messbare Abbildung $f : \Omega \to \Omega'$ auch eine ***Zufallsvariable***.

Satz 12.2 *(Eigenschaften messbarer Abbildungen)*

1. *Ist $(\Omega, \mathcal{A})$ ein messbarer Raum und $f : \Omega \to \Omega'$ eine Abbildung in eine Menge Ω', so ist $\mathcal{B} = \{B \subset \Omega' \mid f^{-1}(B) \in \mathcal{A}\}$ eine σ-Algebra in Ω' und f ist als Abbildung zwischen $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{B})$ messbar.*
2. *Ist $f : \Omega \to \Omega'$ eine Abbildung zwischen zwei messbaren Räumen $(\Omega, \mathcal{A})$ und*

$(\Omega', \mathcal{B})$ *und ist* $M \subset B$ *ein Erzeugendensystem von* $\mathcal{B}$*, so ist* f *genau dann messbar, wenn* $\{f^{-1}(B) \mid B \in M\} \subset \mathcal{A}$*.*

3. *Sind* $(\Omega, \mathcal{A})$*,* $(\Omega', \mathcal{B})$*,*$(\Omega'', \mathcal{C})$ *drei messbare Räume und* $f : \Omega \to \Omega'$ *sowie* $g : \Omega' \to \Omega''$ *zwei messbare Abbildungen, so ist* $g \circ f : \Omega \to \Omega''$ *messbar.*

Beweis:

1. Wegen $f^{-1}(\Omega') = \Omega \in \mathcal{A}$ gilt $\Omega' \in \mathcal{B}$.
   Ist $B \in \mathcal{B}$, so gilt $f^{-1}(\Omega' \setminus B) = \Omega \setminus f^{-1}(B) \in \mathcal{A}$, also $\Omega' \setminus B \in \mathcal{B}$.
   Sind für $i \in \mathbb{N}$ Mengen $B_i \in \mathcal{B}$ gegeben, so gilt $f^{-1}\left(\bigcup_{i\in\mathbb{N}} B_i\right) = \bigcup_{i\in\mathbb{N}} f^{-1}(B_i) \in \mathcal{A}$, also auch $\bigcup_{i\in\mathbb{N}} B_i \in \mathcal{B}$.
2. „$\Rightarrow$“ klar nach Definition
   „$\Leftarrow$“ Das Mengensystem $\mathcal{C} = \{C \subset \Omega' \mid f^{-1}(C) \in \mathcal{A}\}$ ist nach 1. eine σ-Algebra.
   Nach Voraussetzung gilt $M \subset \mathcal{C}$, also $\mathcal{B} = \mathrm{A}(M) \subset \mathcal{C}$.
   Das bedeutet nach Definition, dass f messbar ist.
3. Für $C \in \mathcal{C}$ gilt $(g \circ f)^{-1}(C) = (f^{-1} \circ g^{-1})(C) = f^{-1}(g^{-1}(C)) \in \mathcal{A}$.
   Also ist $g \circ f$ messbar.

$\square$

Korollar 12.3 *(Eigenschaften reellwertiger messbarer Abbildungen)*
Sei $(\Omega, \mathcal{A})$ *ein messbarer Raum und sei* $n \in \mathbb{N}_+$*.*

1. *Eine Abbildung* $f : \Omega \to \mathbb{R}^n$ *ist genau dann messbar, wenn für alle* $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ *gilt:* $f^{-1}(]-\infty, x_1] \times \ldots \times]-\infty, x_n]) \in \mathcal{A}$
2. *Ist* $m \in \mathbb{N}_+$*, so ist jede stetige Abbildung* $f : \mathbb{R}^m \to \mathbb{R}^n$ *messbar.*
3. *Ist* $f : \mathbb{R} \to \mathbb{R}$ *eine monoton wachsende oder eine monoton fallende Funktion, so ist* f *messbar.*
4. *Sind Funktionen* $f_1, \ldots, f_n : \Omega \to \mathbb{R}$ *gegeben, so ist die Funktion* $f := (f_1, \ldots, f_n) : \Omega \to \mathbb{R}^n$*, def. durch* $x \mapsto (f_1(x), \ldots, f_n(x))$ *genau dann messbar, wenn* $f_1, \ldots, f_n$ *messbar sind.*
5. *Sind* $f_i : \Omega \to \mathbb{R}$ *für* $i \in \mathbb{N}_+$ *messbare Funktionen, so sind auch die folgenden Funktionen messbar:*
   1. $c_1 f_1 + \ldots + c_n f_n$ *für* $c_1, \ldots, c_n \in \mathbb{R}$
   2. $f_1 \cdot \ldots \cdot f_n$
   3. $\sup\{f_i \mid i \in \mathbb{N}_+\}, \inf\{f_i \mid i \in \mathbb{N}_+\} : \Omega \to [-\infty, \infty]$*, wobei Teilmengen* $A \subset [-\infty, \infty]$ *genau dann messbar sein sollen, wenn* $A \cap \mathbb{R}$ *messbar ist.*

Beweis:

1. folgt aus Satz 12.2 2.
2. Für jede offene Teilmenge $B \subset \mathbb{R}^n$ gilt $f^{-1}(B) \subset \mathbb{R}^n$ offen, also borelsch. Jetzt folgt die Behauptung aus 12.2 2., weil die offenen Mengen ein Erzeugendensystem der Borelmengen sind. (vgl. Analysis)
3. Für alle $x \in \mathbb{R}$ ist $f^{-1}(]-\infty, x])$ ein Intervall und somit Borel-messbar. Die Behauptung folgt wieder aus 12.2 2. .
4. „$\Rightarrow$“
   Sei f messbar und seien $a, b \in \mathbb{R}$ mit $a \leq b$.
   Für $i \in \{1, \ldots, n\}$ sei $D_i = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a < x_i \leq b\} \in \mathrm{B}(\mathbb{R}^n)$.
   Nun gilt $f_i^{-1}(]a, b]) = f^{-1}(D_i) \in \mathcal{A}$, d.h. f_i ist messbar
   „$\Leftarrow$“
   Seien $f_1, \ldots, f_n$ messbar.
   Für $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in \mathbb{R}^n$ mit $a_i \leq b_i$ für $i = 1, \ldots, n$ gilt
   $f^{-1}(]a_1, b_1] \times \ldots \times]a_n, b_n]) = \bigcap\limits_{i=1}^{n} f_i^{-1}(]a_i, b_i]) \in \mathcal{A}$, d.h. f ist messbar.
5. Zu 1.:
   Sei $f = (f_1, \ldots, f_n) : \Omega \to \mathbb{R}^n$ und $g : \mathbb{R}^n \to \mathbb{R}$ sei def. durch $(x_1, \ldots, x_n) \mapsto c_1 x_1 + \ldots + c_n x_n$.
   Dann ist f messbar nach 4. und g messbar nach 2. und somit $g \circ f$ messbar nach Satz 12.2 3. .
   Zu 2.:
   Sei $f = (f_1, \ldots, f_n) : \Omega \to \mathbb{R}^n$ und sei $g : \mathbb{R}^n \to \mathbb{R}$ def. durch $(x_1, \ldots, x_n) \mapsto x_1 \cdot \ldots \cdot x_n$.
   Wegen $f_1 \cdot \ldots \cdot f_n = g \circ f$ folgt die Behauptung wie oben.
   Zu 3.:
   Die Behauptung folgt aus $\sup\{f_i \mid i \in \mathbb{N}_+\}^{-1}(]-\infty, x]) = \bigcap\limits_{i \geq 1} f_i^{-1}(]-\infty, x]) \in \mathcal{A}$ bzw. aus $\inf\{f_i \mid i \in \mathbb{N}_+\}^{-1}(]-\infty, x]) = \bigcup\limits_{i \geq 1} f_i^{-1}(]-\infty, x]) \in \mathcal{A}$ für $x \in \mathbb{R}$.

$\square$

Bemerkung 12.4 *Ist Ω abzählbar und $\mathcal{A} = \mathfrak{P}(\Omega)$, so ist jede Abbildung $f : \Omega \to \Omega'$ in einen messbaren Raum $(\Omega', \mathcal{B})$ messbar nach 12.1.*
Insbesondere ist jede Abbildung $f : \Omega \to \Omega'$ in den messbaren Raum $(\Omega', \mathfrak{P}(\Omega'))$ messbar, d.h. die früher definierten Zufallsvariablen sind auch Zufallsvariable im Sinne von Definition 12.1.

Satz und Definition 12.5 *(Verteilung und Dichte einer Zufallsvariablen)*
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $(\Omega', \mathcal{B})$ ein messbarer Raum und sei $X : \Omega \to \Omega'$ eine Zufallsvariable.

1. *Durch $P_X : \mathcal{B} \to [0,1]$, def. durch $B \mapsto P(X \in B) = P(X^{-1}(B))$, wird eine Wahrscheinlichkeitsverteilung auf Ω' definiert. Sie heißt die* ***Verteilung von $\boldsymbol{X}$***.
2. *Ist $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable, so heißt die Funktion $F_X : \mathbb{R} \to [0,1]$, def. durch $x \mapsto P(X \leq x) = P(X^{-1}(]-\infty, x]))$, die* ***Verteilungsfunktion von $\boldsymbol{X}$***.
   Sie stellt eine Verteilungsfunktion im Sinne von Definition 11.1 dar.
3. *Ist $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable und besitzt die Verteilungsfunktion $F_X : \mathbb{R} \to [0,1]$ eine Dichte $f : \mathbb{R} \to [0,\infty[$, so sagt man auch, X habe die* ***Dichte $\boldsymbol{f}$*** *bzw. f sei die* ***Dichte von $\boldsymbol{X}$***.

Beweis:

1. $P_X(\Omega') = P(X^{-1}(\Omega')) = P(\Omega) = 1$
   Sei $(B_i)_{i\in\mathbb{N}}$ eine Familie paarweise disjunkter Elemente von B.
   $P_X\left(\dot{\bigcup\limits_{i\in\mathbb{N}}} B_i\right) = P\left(X^{-1}\left(\dot{\bigcup\limits_{i\in\mathbb{N}}} B_i\right)\right) =$
   $P\left(\dot{\bigcup\limits_{i\in\mathbb{N}}} X^{-1}(B_i)\right) = \sum\limits_{i\in\mathbb{N}} P(X^{-1}(B_i)) = \sum\limits_{i\in\mathbb{N}} P_X(B_i)$
2. Die Verteilungsfunktion F_X ist wegen $F_X(x) = P(X^{-1}(]-\infty, x])) = P_X(]-\infty, x])$ die Verteilungsfunktion der Wahrscheinlichkeitsverteilung $P_X : \mathrm{B}(\mathbb{R}) \to [0,1]$, so dass die Behauptung aus Satz 11.2 1. folgt.

$\square$

Satz 12.6 *(Eigenschaften der Dichte)*
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable mit Dichte $f : \mathbb{R} \to [0,\infty[$.

1. *Ist $g : \mathbb{R} \to \mathbb{R}$ bijektive Abbildung und gibt es eine endliche Menge $M \subset \mathbb{R}$, so dass $g|_{\mathbb{R}\setminus M} : \mathbb{R} \setminus M \to \mathbb{R}$ stetig differenzierbar ist mit $g'(x) > 0$ für alle $x \in (\mathbb{R} \setminus M) \cap X(\Omega)$ und $\mathbb{R} \setminus g(\mathbb{R} \setminus M)$ endlich, so besitzt die Zufallsvariable $Y = g \circ X : \Omega \to \mathbb{R}$ eine Dichte h, nämlich $h : \mathbb{R} \to [0,\infty[$, def. durch*
   $$x \mapsto \begin{cases} \frac{f(g^{-1}(x))}{g'(g^{-1}(x))} & \text{für } x \in g(\mathbb{R} \setminus M) \\ \text{beliebig} & \text{für } x \notin g(\mathbb{R} \setminus M) \end{cases}.$$
2. *Gilt stattdessen in der Situation von 1., $g'(x) < 0$ für alle $x \in (\mathbb{R}\setminus M) \cap X(\Omega)$, so besitzt die Zufallsvariable $Y = g \circ X$ die Dichte $h : \mathbb{R} \to [0,\infty[$, def. durch*
   $$x \mapsto \begin{cases} \frac{-f(g^{-1}(x))}{g'(g^{-1}(x))} & \text{für } x \in g(\mathbb{R} \setminus M) \\ \text{beliebig} & \text{für } x \notin g(\mathbb{R} \setminus M) \end{cases}.$$

3. *Für $a \in \mathbb{R}$ besitzt die Zufallsvariable $X + a$ die Dichte $h : \mathbb{R} \to [0, \infty[$, def. durch $x \mapsto f(x - a)$.*
4. *Für $a \in \mathbb{R} \setminus \{0\}$ besitzt die Zufallsvariable $a \cdot X$ die Dichte $h : \mathbb{R} \to [0, \infty[$, def. durch $x \mapsto \frac{1}{|a|} f\left(\frac{x}{a}\right)$.*

Beweis:

1. Es gilt für alle $x \in g(\mathbb{R} \setminus M)$ die Gleichung $F_Y(x) = P(Y^{-1}(]-\infty, x])) = P(X^{-1}(g^{-1}(]-\infty, x])) =$
   $P_X(g^{-1}(]-\infty, x])) \overset{\text{Monotonie}}{=} \int\limits_{-\infty}^{g^{-1}(x)} f(t)\,\mathrm{d}t \overset{\text{Subst.:}u:=g(t)}{=} \int\limits_{-\infty}^{x} f(g^{-1}(u)) \frac{1}{g'(g^{-1}(u))}\,\mathrm{d}u =$
   $\int\limits_{-\infty}^{x} h(t)\,\mathrm{d}t$
   $\Rightarrow F_Y$ besitzt die Dichte h.
2. folgt analog 1.
3. folgt sofort aus 1. und 2.
4. folgt sofort aus 1. und 2.

$\square$

Satz 12.7 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit Dichte $f : \mathbb{R} \to [0, \infty[$.*
Dann hat die Zufallsvariable $Y = X^2 : \Omega \to \mathbb{R}$ die Dichte $g : \mathbb{R} \to [0, \infty[$, def. durch $y \mapsto \begin{cases} \frac{f(\sqrt{y})}{2\sqrt{y}} + \frac{f(-\sqrt{y})}{2\sqrt{y}} & \text{für } y > 0 \\ 0 & \text{sonst} \end{cases}$.

Beweis:
Für $y < 0$ gilt $P(Y \leq y) = 0$ und für $y = 0$ gilt $P(Y \leq y) = P(X = 0) = 0$, da X eine Dichte hat.
Für $y > 0$ gilt
$P(Y \leq y) = P(Y \leq y, X < 0) + \underbrace{P(Y \leq y, X = 0)}_{=0} + P(Y \leq y, X > 0) =$
$P(-\sqrt{y} \leq X < 0) + P(0 < X \leq \sqrt{y}) =$
$\int\limits_{-\sqrt{y}}^{0} f(t)\,\mathrm{d}t + \int\limits_{0}^{\sqrt{y}} f(t)\,\mathrm{d}t =$
$\int\limits_{-\sqrt{y}}^{0} f(t)\,\mathrm{d}t - \int\limits_{-\infty}^{0} f(t)\,\mathrm{d}t + \int\limits_{-\infty}^{\sqrt{y}} f(t)\,\mathrm{d}t =$
$-\int\limits_{-\infty}^{-\sqrt{y}} f(t)\,\mathrm{d}t + \int\limits_{-\infty}^{\sqrt{y}} f(t)\,\mathrm{d}t \overset{(*)}{=}$

$- \int\limits_{-\infty}^{y} \frac{f(-\sqrt{u})}{-2\sqrt{u}}\,\mathrm{d}u + \int\limits_{-\infty}^{y} \frac{f(\sqrt{u})}{2\sqrt{u}}\,\mathrm{d}u = \int\limits_{-\infty}^{y} g(t)\,\mathrm{d}t$

(*): 1. Integral: Substitution $t := -\sqrt{u}$; 2. Integral: Substitution $t := \sqrt{u}$ $\square$

Im Folgenden sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable.

Definition 12.8 Hat X eine abzählbare Bildmenge $X(\Omega) = \{x_1, x_2, \dots\} \subset \mathbb{R}$ und existiert $\sum\limits_{i=1}^{\infty} |x_i| P(X = x_i)$, so heißt $E(X) := \sum\limits_{i=1}^{\infty} x_i P(X = x_i)$ der ***Erwartungswert von X*** .

Lemma 12.9 *Für $k \in \mathbb{Z}$ und $n \in \mathbb{N}_+$ definieren wir*
$A_{nk} = \{\omega \in \Omega \mid \frac{k}{n} \leq X(\omega) < \frac{k+1}{n}\} \in \mathcal{A}$ *und* $X_n : \Omega \to \mathbb{R}$ *durch*
$\omega \mapsto \sum\limits_{k=-\infty}^{\infty} \frac{k}{n} \chi_{A_{nk}}(\omega)$.
(=„größte rationale Zahl der Form $\frac{k}{n}$ mit $k \in \mathbb{Z}$ und $\frac{k}{n} \leq X(\omega)$ (vgl 1.)“)

1. *Für $\omega \in \Omega$ gilt $X_n(\omega) \leq X(\omega) < X_n(\omega) + \frac{1}{n}$.*
2. *Für alle $m, n \in \mathbb{N}$ und alle $\omega \in \Omega$ gilt $|X_n(\omega) - X_m(\omega)| = \frac{1}{n} + \frac{1}{m}$.*
3. *Existiert $E(X_n)$ für ein $n \in \mathbb{N}_+$, so gilt $E(X_n) = \sum\limits_{k=-\infty}^{\infty} \frac{k}{n} P(A_{nk})$ und $E(X_m)$ existiert für alle $m \in \mathbb{N}_+$ und es gilt $|E(X_n) - E(X_m)| \leq \frac{1}{n} + \frac{1}{m}$.*

Beweis:

1. Es gilt $X_n(\omega) = \sum\limits_{\{k \in \mathbb{Z} \mid \omega \in A_{nk}\}} \frac{k}{n} \leq \underbrace{\sum\limits_{\{k \in \mathbb{Z} \mid \omega \in A_{nk}\}}}_{\text{es gibt genau ein solches } k} X(\omega) = X(\omega) <$
   $\sum\limits_{\{k \in \mathbb{Z} \mid \omega \in A_{nk}\}} \frac{k+1}{n} = X_n(\omega) + \frac{1}{n}$
2. Aus 1. folgt $|X_n(\omega) - X(\omega)| \leq \frac{1}{n}$.
   Nach der Dreiecksungleichung folgt:
   $|X_n(\omega) - X_m(\omega)| \leq |X_n(\omega) - X(\omega)| + |X_m(\omega) - X(\omega)| \leq \frac{1}{n} + \frac{1}{m}$
3. Existiert $E(X_n)$, so gilt $E(X_n) = \sum\limits_{k \in \mathbb{Z}} \frac{k}{n} P(X_n = \frac{k}{n}) = \sum\limits_{k \in \mathbb{Z}} \frac{k}{n} P(A_{nk}) =: (*)$

   Es gilt $A_{nk} = \{\frac{k}{n} \leq X < \frac{k+1}{n}\} = \dot{\bigcup}\limits_{i=0}^{m-1} \{\frac{kn+i}{nm} \leq X < \frac{km+i+1}{nm}\} =$
   $\dot{\bigcup}\limits_{i=0}^{m-1} A_{nm,km+i}$
   Somit folgt:
   $E(X_n)$ existiert$\Leftrightarrow \sum\limits_{k=-\infty}^{\infty} \left|\frac{k}{n}\right| P(A_{nk}) < \infty \Leftrightarrow$
   $\sum\limits_{k=-\infty}^{\infty} \left|\frac{km}{nm}\right| P\left(\dot{\bigcup}\limits_{i=0}^{m-1} A_{nm,km+i}\right) < \infty \Leftrightarrow$

$\sum\limits_{k=-\infty}^{\infty} \sum\limits_{i=0}^{m-1} \left|\frac{km}{nm}\right| P\left(A_{nm,km+i}\right) < \infty$

$E(X_{nm})$ existiert$\Leftrightarrow \sum\limits_{k=-\infty}^{\infty} \sum\limits_{i=0}^{m-1} \underbrace{\left|\frac{km+i}{nm}\right|}_{\leq\left|\frac{km}{nm}\right|+\frac{i}{nm}} P(A_{nm,km+i}) < \infty$

Beachte, dass $\sum\limits_{k=-\infty}^{\infty} \sum\limits_{i=0}^{m-1} \frac{i}{nm} P(A_{nm,km+i}) \leq \frac{m-1}{nm} \sum\limits_{l=-\infty}^{\infty} P(A_{nm,l}) = \frac{m-1}{nm}$ gilt.

Es ergibt sich $E(X_n)$ existiert$\Leftrightarrow E(X_{nm})$ existiert $\stackrel{\text{analog}}{\Leftrightarrow} E(X_m)$ existiert.
Schließlich bleibt $|E(X_n) - E(X_m)| \leq \frac{1}{n} + \frac{1}{m}$ zu zeigen.
Wie oben errechnet man $|E(X_n) - E(X_{mn})| \leq \frac{m-1}{mn} < \frac{1}{n}$.
Analog folgt $|E(X_m) - E(X_{nm})| < \frac{1}{m}$ und die Behauptung ist eine Konsequenz der Dreiecksungleichung

□

Definition 12.10 Ist $\sum\limits_{k\in\mathbb{Z}} |k| P(k \leq X < k+1) < \infty$, so existiert $E(X) := \lim\limits_{n\to\infty} \sum\limits_{k\in\mathbb{Z}} \frac{k}{n} P(\frac{k}{n} \leq X < \frac{k+1}{n})$ und heißt der ***Erwartungswert von X***.

Satz 12.11 *(Eigenschaften des Erwartungswerts)*
Seien $X, Y : \Omega \to \mathbb{R}$ zwei reellwertige Zufallsvariablen für die $E(X)$ und $E(Y)$ existieren.

1. *(Linearität)*
   Sind $a, b \in \mathbb{R}$, so gilt $E(aX + bY) = aE(X) + bE(Y)$.
2. *(Monotonie)*
   Gilt $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$, so folgt $E(X) \leq E(Y)$.
3. *(Transformationsformel)*
   Besitzt X eine Dichte $f : \mathbb{R} \to [0, \infty[$ und ist $g : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion, so existert $E(g(x))$ genau dann, wenn $\int\limits_{-\infty}^{\infty} |g(x)| f(x)\,\mathrm{d}x < \infty$ ist und in diesem Fall gilt $E(g(X)) = \int\limits_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x$.
4. *Besitzt X eine Dichte $f : \mathbb{R} \to [0, \infty[$, so gilt $E(X) = \int\limits_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$*

Beweis:

1. Definiere X_n, Y_n, $(X+Y)_n$ für $n \in \mathbb{N}_+$ wie in Lemma 12.9.
   Dann gilt für alle $\omega \in \Omega$ die Ungleichungskette (vgl. 12.9 1.):
   $(X+Y)_n(\omega) \leq (X+Y)(\omega) = X(\omega) + Y(\omega) \leq X_n(\omega) + \frac{1}{n} + Y_n(\omega) + \frac{1}{n} \leq$

$X(\omega) + Y(\omega) + \frac{2}{n} \leq (X+Y)_n(\omega) + \frac{3}{n}$
Hieraus folgt $|(X+Y)_n(\omega) - (X_n(\omega) + Y_n(\omega))| \leq \frac{2}{n}$
und somit $|E((X+Y)_n) - (E(X_n) + E(Y_n))| \leq \frac{2}{n}$.
Wir erhalten
$E(X+Y) = \lim\limits_{n\to\infty} E((X+Y)_n) = \lim\limits_{n\to\infty} E(X_n) + \lim\limits_{n\to\infty} E(Y_n) = E(X) + E(Y)$.
Die Gleichung $E(aX) = aE(X)$ folgt unmittelbar aus der Definition.

2. Für alle $n \in \mathbb{N}_+$ und $\omega \in \Omega$ gilt
$X_n(\omega)$ =größte Zahl der Form $\frac{k}{n}$ mit $k \in \mathbb{Z}$ und $\frac{k}{n} \leq X(\omega)$.
Somit folgt aus $X(\omega) \leq Y(\omega)$ sofort $X_n(\omega) \leq Y_n(\omega)$ für alle $n \in \mathbb{N}_+$.
Nun gilt $E(X) = \lim\limits_{n\to\infty} E(X_n) \leq \lim\limits_{n\to\infty} E(Y_n) = E(Y)$.
3. Da g stetig ist, ist $g|_{[k,k+1]}$ für jedes $k \in \mathbb{Z}$ gleichmäßig stetig.
Zu jedem $\delta > 0$ gibt es damit eine streng monoton steigende Familie $(x_n)_{n\in\mathbb{Z}}$ in $\mathbb{R}$ mit $\lim\limits_{n\to-\infty} x_n = -\infty$ und $\lim\limits_{n\to\infty} x_n = \infty$ und mit $|g(x) - g(x_n)| < \delta$ für $x_n \leq x < x_{n+1}$.
Definiere eine Funktion $h_g : \mathbb{R} \to \mathbb{R}$ durch $x \mapsto g(x_n)$ falls $x_n \leq x < x_{n+1}$.
Es gilt $|h_g(x) - g(x)| < \delta$ für alle $x \in \mathbb{R}$.
Ferner gilt
$E(h_g(X)) = \sum\limits_{n=-\infty}^{\infty} g(x_n)P(X \in [x_n, x_{n+1}[) = \sum\limits_{n=-\infty}^{\infty} g(x_n)P(X^{-1}([x_n, x_{n+1}[)) = \sum\limits_{n=-\infty}^{\infty} g(x_n) \int\limits_{x_n}^{x_{n+1}} f(t)\,\mathrm{d}t.$
Wegen $|g(x) - g(x_n)| < \delta$ für $x \in [x_n, x_{n+1}[$ und $\sum\limits_{n=-\infty}^{\infty} \delta \int\limits_{x_n}^{x_{n+1}} f(t)\,\mathrm{d}t = \delta < \infty$ gilt:
$\sum\limits_{n=-\infty}^{\infty} |g(x_n)| \int\limits_{x_n}^{x_{n+1}} f(t)\,\mathrm{d}t = \sum\limits_{n=-\infty}^{\infty} \int\limits_{x_n}^{x_{n+1}} |h_g(t)| f(t)\,\mathrm{d}t < \infty \Leftrightarrow$
$\int\limits_{-\infty}^{\infty} |g(t)| f(t)\,\mathrm{d}t = \sum\limits_{n=-\infty}^{\infty} \int\limits_{x_n}^{x_{n+1}} |g(t)| f(t)\,\mathrm{d}t < \infty.$
Ist dies erfüllt, so gilt $\left| E(h_g(X)) - \int\limits_{-\infty}^{\infty} g(t)f(t)\,\mathrm{d}t \right| < \delta$.
Wegen $|h_g(x) - g(x)| < \delta$ für $x \in \mathbb{R}$ gilt auch $|E(h_g(X)) - E(g(X))| \overset{1.}{=} |E(h_g(X) - g(X))| \overset{2.}{\leq} |E(\delta)| = \delta$.
Nach der Dreiecksungleichung folgt $\left| E(g(X)) - \int\limits_{-\infty}^{\infty} g(t)f(t)\,\mathrm{d}t \right| < 2\delta$.
Da $\delta > 0$ beliebig war, folgt die Behauptung.
4. folgt aus 3. mit $g(x) = x$

□

Bemerkung 12.12 *Besitzt $X : \Omega \to \mathbb{R}$ eine Dichte $f : \mathbb{R} \to [0, \infty[$ und existiert $E(X^2) = \sum\limits_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$, so existiert wegen $|x| \leq \frac{1}{2}(1 + x^2)$ auch $E(X) =$*

$\int\limits_{-\infty}^{\infty} xf(x)\,\mathrm{d}x.$

Definition 12.13 Sei $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable für die $E(X)$ und $E(X^2)$ existieren.

1. Die Zahl $\mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - E(X)^2 \geq 0$ heißt die ***Varianz von X***.
2. Die Zahl $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ heißt die ***Standardabweichung von X***.

Beispiel 12.14 (Erwartungswert und Varianz der Normalverteilung)
Eine reellwertige Zufallsvariable $X : \Omega \to \mathbb{R}$ sei $N(\mu, \sigma^2)$-verteilt mit $\mu \in \mathbb{R}$ und $\sigma^2 \in [0, \infty[$, d.h. $P_X = N(\mu, \sigma^2)$.
Dann gilt $E(X) = \mu$ und $\mathrm{Var}(X) = \sigma^2$.

Beweis:
Die Zufallsvariable X hat die Dichte $f : \mathbb{R} \to [0, \infty[$, def. durch
$x \mapsto \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.
Nach Satz 12.6 hat die Zufallsvariable $Y := \frac{X-\mu}{\sigma}$ dann die Dichte $\varphi : \mathbb{R} \to [0, \infty[$, def. durch $x \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.
Wegen $\varphi(-x) = \varphi(x)$ für $x \in \mathbb{R}$ gilt $\int\limits_{-\infty}^{0} x\varphi(x)\,\mathrm{d}x = \int\limits_{0}^{\infty} x\varphi(x)\,\mathrm{d}x$, also $E(Y) = 0$.
Es folgt $E(X) = E(\sigma Y + \mu) = \sigma E(Y) + \mu = \mu$.
Wegen $E(Y) = 0$ gilt auch $\mathrm{Var}(Y) = E(Y^2)$.

$$E(Y^2) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x =$$

$$-\frac{1}{\sqrt{2\pi}} \lim_{n\to\infty} \int\limits_{-n}^{n} x \cdot \left(-x \exp\left(-\frac{x^2}{2}\right)\right)\,\mathrm{d}x \overset{\text{partielle Integration}}{=}$$

$$-\frac{1}{\sqrt{2\pi}} \lim_{n\to\infty} \left(\left[x \exp\left(-\frac{x^2}{2}\right)\right]_{-n}^{n} - \int\limits_{-n}^{n} \exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x\right) =$$

$$-\frac{1}{\sqrt{2\pi}} \left(\underbrace{\lim_{n\to\infty} \left(2n \exp\left(-\frac{n^2}{2}\right)\right)}_{=0} - \lim_{n\to\infty} \int\limits_{-n}^{n} \exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x\right) =$$

$$\lim_{n\to\infty} \left(\int\limits_{-n}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x\right) = \int\limits_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x = 1$$

Also ergibt sich $\mathrm{Var}(X) = \mathrm{Var}(X - \mu) = \sigma^2 \,\mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \sigma^2 \,\mathrm{Var}(Y) = \sigma^2$ □

Beispiel 12.15 (Erwartungswert und Varianz der Exponentialverteilung)
Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ sei exponential verteilt zum Parameter $\lambda > 0$.

Dann gilt $E(X) = \int\limits_0^\infty x\lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \lambda \lim\limits_{n\to\infty}\left(\left[-\frac{1}{\lambda}x\mathrm{e}^{-\lambda x}\right]_0^n - \int\limits_0^n \left(-\frac{1}{\lambda}\right)\mathrm{e}^{-\lambda x}\,\mathrm{d}x\right) =$
$\lim\limits_{n\to\infty}\left(n\mathrm{e}^{-\lambda n}\right) + \lim\limits_{n\to\infty}\int\limits_0^n \mathrm{e}^{-\lambda x}\,\mathrm{d}x = 0 + \lim\limits_{n\to\infty}\left[-\frac{1}{\lambda}\mathrm{e}^{-\lambda x}\right]_0^n = \frac{1}{\lambda}$, sowie
$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \int\limits_0^\infty \lambda x^2 \mathrm{e}^{-\lambda x}\,\mathrm{d}x - \frac{1}{\lambda^2} =$
$\lambda \lim\limits_{n\to\infty}\int\limits_0^n x^2\mathrm{e}^{-\lambda x}\,\mathrm{d}x - \frac{1}{\lambda^2} \overset{\text{partiell}}{=}$
$\lambda \lim\limits_{n\to\infty}\left(\left[x^2\left(-\frac{1}{\lambda}\right)\mathrm{e}^{-\lambda x}\right]_0^n - \int\limits_0^n 2x\left(-\frac{1}{\lambda}\right)\mathrm{e}^{-\lambda x}\,\mathrm{d}x\right) - \frac{1}{\lambda^2} =$
$\lim\limits_{n\to\infty}\left(-n^2\mathrm{e}^{-\lambda x}\right) + \frac{2}{\lambda}\int\limits_0^\infty x\lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x - \frac{1}{\lambda^2} =$
$0 + \frac{2}{\lambda}E(X) - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Definition 12.16

1. Sei I eine Menge und sei $\{A_i\}_{i\in I} \subset \mathcal{A}$ eine Familie von Ereignissen in Ω.
   Die Familie $\{A_i\}_{i\in I}$ heißt ***unabhängig***, wenn für jede endliche Teilfamilie $\{A_{i_1}, \dots, A_{i_n}\}$ die Produktformel $P(A_{i_1} \cap \ldots \cap A_{i_n}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_n})$ gilt.
2. Sei I eine Menge und für jedes $i \in I$ sei $X_i : \Omega \to \Omega_i$ eine Zufallsvariable mit Werten in einem messbaren Raum $(\Omega_i, \mathcal{B}_i)$.
   Die Familie von Zufallsvariablen $\{X_i\}_{i\in I}$ heißt ***unabhängig***, wenn für jede Wahl von Mengen $B_i \in \mathcal{B}_i$ für $i \in I$ die Ereignisse $\{X_i \in B_i\}_{i\in I}$ unabhängig sind.

Satz 12.17 *(Charakterisierung unabhängiger Zufallsvariablen)*

1. *Sei I eine Menge und $\{X_i\}_{i\in I}$ eine Familie von Zufallsvariablen $X_i : \Omega \to \mathbb{R}$.*
   Genau dann ist $\{X_i\}_{i\in I}$ unabhängig, wenn für alle $\{x_i\}_{i\in I} \subset \mathbb{R}$ die Ereignisse $\{X_i \leq x_i\}$ unabhängig sind.
2. *Sei $n \in \mathbb{N}_+$ und seien $X_1, \dots, X_n : \Omega \to \mathbb{R}$ reellwertige Zufallsvariable, wobei X_i für $i = 1, \dots, n$ eine Dichte $f_i : \mathbb{R} \to [0, \infty[$ besitze.*
   Genau dann ist $\{X_1, \dots, X_n\}$ unabhängig, wenn die Abbildung $f : \mathbb{R}^n \to [0, \infty[$, def. durch $(x_1, \dots, x_n) \mapsto f(x_1) \cdot \ldots \cdot f(x_n)$, eine Dichte der Zufallsvariablen $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$, def. durch $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$, darstellt.

Beweis:

1. In der Maßtheorie wird gezeigt:
   Ist $\mathcal{A}$ eine σ-Algebra und $\mathcal{B}$ ein Erzeugendensystem mit $A_1, A_2 \in \mathcal{B} \Rightarrow A_1 \cap A_2 \in \mathcal{B}$, so folgt aus der Unabhängigkeit von $\mathcal{B}$ bereits die von $\mathcal{A}$. Hieraus ergibt sich die Behauptung.
2. „$\Rightarrow$"
   Sei $Q : \mathrm{B}(\mathbb{R}^n) \to [0,1]$ das Wahrscheinlichkeitsmaß mit Dichte f.
   Für $(a_1, \dots, a_n), (b_1, \dots, b_n) \in \mathbb{R}^n$ mit $a_i \leq b_i$ für $i = 1, \dots, n$ gilt dann
   $Q(]a_1, b_1] \times \dots \times]a_n, b_n]) = \int\limits_{a_1}^{b_1} \dots \int\limits_{a_n}^{b_n} f_1(x_1) \cdot \ldots \cdot f_n(x_n) \, \mathrm{d}x_n \cdot \ldots \cdot \mathrm{d}x_1 =$
   $\left(\int\limits_{a_1}^{b_1} f_1(x_1) \, \mathrm{d}x_1\right) \cdot \ldots \cdot \left(\int\limits_{a_n}^{b_n} f_n(x_n) \, \mathrm{d}x_n\right) =$
   $\prod\limits_{i=1}^{n} P(X_i \in]a_i, b_i]) = P(X_1 \in]a_1, b_1], \dots, X_n \in]a_n, b_n]) =$
   $P_X(]a_1, b_1] \times \dots \times]a_n, b_n])$
   Da $\mathrm{B}(\mathbb{R}^n)$ durch die Mengen $]a_1, b_1] \times \dots \times]a_n, b_n]$ erzeugt wird, folgt $Q = P_X$, d.h. X hat die Dichte f.
   „$\Leftarrow$"
   Sei $Q : \mathrm{B}(\mathbb{R}^n) \to [0,1]$ wie eben definiert.
   Nach Voraussetzung gilt $Q = P_X$, also $Q(\mathbb{R} \times \dots \times]a_i, b_i] \times \mathbb{R} \times \dots \times \mathbb{R}) =$
   $P_X(X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i \in]a_i, b_i], X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) =$
   $P(X_i \in]a_i, b_i])$ für $i \in \{1, \dots, n\}$ und $a_i, b_i \in \mathbb{R}$ mit $a_i \leq b_i$.
   Andererseits gilt
   $\int\limits_{a_i}^{b_i} f(x_i) \, \mathrm{d}x_i = \int\limits_{\mathbb{R}} \dots \int\limits_{\mathbb{R}} \int\limits_{a_i}^{b_i} \int\limits_{\mathbb{R}} \dots \int\limits_{\mathbb{R}} f_1(x_1) \cdot \dots \cdot f_n(x_n) \, \mathrm{d}x_n \cdot \ldots \cdot \mathrm{d}x_1 =$
   $Q(\mathbb{R} \times \dots \times \mathbb{R} \times]a_i, b_i] \times \mathbb{R} \times \dots \times \mathbb{R})$
   $\prod\limits_{i=1}^{n} P(X_i \in]a_i, b_i]) = \prod\limits_{i=1}^{n} \int\limits_{a_i}^{b_i} f_i(x_i) \, \mathrm{d}x_i = \ldots = P_X(]a_1, b_1] \times \dots \times]a_n, b_n]) =$
   $P\left(\bigcap\limits_{i=1}^{n} \{X_i \in]a_i, b_i]\}\right)$
   Nach 1. sind die Zufallsvariablen $\{X_1, \dots, X_n\}$ damit unabhängig.

□

Korollar 12.18 *Seien $X_1, X_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariable mit Dichten $f_i : \mathbb{R} \to [0, \infty[$ für $i = 1, 2$.*
Dann besitzt $X_1 + X_2 : \Omega \to \mathbb{R}$ die Dichte $g : \mathbb{R} \to [0, \infty[$, def. durch $x \mapsto \int\limits_{-\infty}^{\infty} f_1(x - t) f_2(t) \, \mathrm{d}t$.

Beweis:
Sei $Y = X_1 + X_2 : \Omega \to \mathbb{R}$. Für $y \in \mathbb{R}$ gilt
$P(Y \leq y) = \iint_{\{(x_1, x_2) \in \mathbb{R}^2 | x_1 + x_2 \leq y\}} f_1(x_1) \cdot f_2(x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \overset{\text{Substitution:} u := x_1 + x_2, v := x_2}{=}$

$\int\limits_{-\infty}^{y} \left(\int\limits_{-\infty}^{\infty} f_1(u-v) \cdot f_2(v)\, \mathrm{d}v \right) \mathrm{d}u = \int\limits_{-\infty}^{u} g(t)\, \mathrm{d}t.$
Hieraus folgt die Behauptung. $\square$

Definition 12.19 Sind $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ zwei Funktionen und existiert die Funktion $f_1 * f_2 : \mathbb{R} \to \mathbb{R}$, def. durch $x \mapsto \int\limits_{-\infty}^{\infty} f_1(x-t) f_2(t)\, \mathrm{d}t$, so heißt $f_1 * f_2$ die ***Faltung von f_1 und f_2***.

Beispiel 12.20 Für $i = 1, 2$ sei $X_i : \Omega \to \mathbb{R}$ eine $N(\mu_i, \sigma_i^2)$-verteilte Zufallsvariable, wobei $\mu_1, \mu_2 \in \mathbb{R}$ und $\sigma_1^2, \sigma_2^2 \in [0, \infty[$ gelte.
Sind X_1, X_2 unabhängig, so ist $X_1 + X_2 : \Omega \to \mathbb{R}$ eine $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$-verteilte Zufallsvariable.

Beweis:
Aus Satz 12.6 folgt, dass X_i genau dann $N(\mu_i, \sigma_i^2)$-verteilt ist, wenn $X_i - \mu_i$ eine $N(0, \sigma_i^2)$-verteilte Zufallsvariable ist.
Also können wir o.E. $\mu_1 = \mu_2 = 0$ annehmen.
Nun gilt (mit φ_i =Dichte von X_i) für alle $u \in \mathbb{R}$:
$(\varphi_1 * \varphi_2)(u) = \frac{1}{2\pi} \frac{1}{\sigma_1 \sigma_2} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \left(\frac{(u-v)^2}{\sigma_1^2} + \frac{v^2}{\sigma_2^2} \right) \right) \mathrm{d}v \overset{\text{Substitution } (*)}{=}$
$\frac{1}{2\pi\sigma_1\sigma_2} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \left(z^2 + \frac{u^2}{\sigma_1^2 + \sigma_2^2} \right) \right) \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\, \mathrm{d}z =$
$\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{1}{2} \left(\frac{u^2}{\sigma_1^2 + \sigma_2^2} \right) \right) = \varphi_3(u)$ mit
φ_3 =Dichte von $N(0, \sigma_1^2 + \sigma_2^2)$
(*): $z := \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} v - \frac{\sigma_2}{\sigma_1 \sqrt{\sigma_1^2 + \sigma_2^2}} u$ $\square$

Beispiel 12.21
Unabhängig voneinander werden zwei Punkte x, y in $[0, 1]$ gewählt, wobei jeweils Gleichverteilung gelte.
Die Dichten x bzw. y sind $f_1 = f_2$:
$\mathbb{R} \to [0, \infty[$, def. durch $x \mapsto \begin{cases} 1 & \text{falls } 0 \leq x < 1 \\ 0 & \text{sonst} \end{cases}$
Dann hat die Differenz $\triangle := x - y$ die Dichte $g := f_1 * f_2 : \mathbb{R} \to [0, \infty[$, wobei

$$g(x) = \int\limits_{-\infty}^{\infty} f_1(x-t) f_2(-t)\, \mathrm{d}t = \begin{cases} 0 & \text{falls } x < -1 \text{ oder } x \geq 1 \\ 1 + x & \text{falls } -1 \leq x < 0 \\ 1 - x & \text{falls } 0 \leq x < 1 \end{cases}.$$

Für den Abstand $d : [0, 1]^2 \to \mathbb{R}$ gilt:

$$d(x,y) = \begin{cases} \triangle(x,y) & \text{falls} \quad x \geq y \\ -\triangle(x,y) & \text{falls} \quad x < y \end{cases}$$

Nach Satz 12.6 folgt, dass d die folgende Dichte besitzt:

$h : \mathbb{R} \to [0,\infty[$, def. durch $x \mapsto \begin{cases} g(x) + g(-x) & \text{für } 0 \leq x < 1 \\ 0 & \text{sonst} \end{cases} =$

$\begin{cases} 2 - 2x & \text{für } 0 \leq x < 1 \\ 0 & \text{sonst} \end{cases}$

Es folgt:

$E(d) = \int\limits_0^1 x(2-2x)\,\mathrm{d}x = \left[x^2 - \frac{2}{3}x^3\right]_0^1 = \frac{1}{3}$, sowie

$\operatorname{Var}(d) = \int\limits_0^1 x^2(2-2x)\,\mathrm{d}x - E(d)^2 = \left[\frac{2}{3}x^3 - \frac{1}{2}x^4\right]_0^1 - \frac{1}{9} = \frac{1}{18}$

Satz 12.22 *Sind $X, Y : \Omega \to \mathbb{R}$ zwei unabhängige Zufallsvariablen und existieren $E(X), E(Y)$, so existiert auch $E(XY)$ und es gilt $E(XY) = E(X)E(Y)$.*

13 Grenzwertsätze

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und seien $X, X_1, X_2 : \Omega \to \mathbb{R}$ reellwertige Zufallsvariablen.

Definition 13.1

1. Die Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen $X_n : \Omega \to \mathbb{R}$ ***konvergiert stochastisch*** gegen die Zufallsvariable $X : \Omega \to \mathbb{R}$, wenn für alle $\varepsilon > 0$ gilt: $P(|X_n - X| \geq \varepsilon) \to 0$ für $n \to \infty$.
2. Die Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen $X_n : \Omega \to \mathbb{R}$ ***konvergiert fast sicher*** gegen $X : \Omega \to \mathbb{R}$, wenn $P\left(\left\{\omega \in \Omega \mid \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$ gilt.

Satz 13.2 *(Schwaches Gesetz der großen Zahlen)*

Ist $(X_n)_{n\in\mathbb{N}_+}$ eine Folge unabhängiger Zufallsvariablen $X_n : \Omega \to \mathbb{R}$ mit gleichem Erwartungswert $E := E(X_1) = E(X_2) = \ldots$ und mit beschränkten Varianzen $\operatorname{Var}(X_i) \leq V$ für ein $V \geq 0$ für $i \in \mathbb{N}_+$, so konvergiert die Folge $Y_n = \frac{1}{n}(X_1 + \ldots + X_n)$ stochastisch gegen die konstante Zufallsvariable $E : \Omega \to \mathbb{R}$, def. durch $\omega \mapsto E$.

Beweis:
Folgt sofort aus Satz 8.13, denn zum Beweis von 8.13 wurden nur die Rechenregeln für Erwartungswert und Varianz verwendet. $\square$

Satz 13.3 *Konvergiert eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen $X_n : \Omega \to \mathbb{R}$ fast sicher gegen $X : \Omega \to \mathbb{R}$, so konvergiert sie auch stochastisch gegen x.*

Beweis:
Sei $\varepsilon > 0$ beliebig.
Sei $A = \left\{\omega \in \Omega \mid \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\right\}$ und für jedes $n \in \mathbb{N}_+$ sei
$B_n = \{\omega \in \Omega \mid |X_i(\omega) - X(\omega)| < \varepsilon$ für $i \geq n\}$.
Dann gilt $B_1 \subset B_2 \subset \ldots$ und $A \subset \bigcup\limits_{i\geq 1} B_i$.

Nach Satz 10.10 folgt aus $P(A) = 1$, dann $\lim\limits_{n\to\infty} P(B_n) = P\left(\bigcup\limits_{i\geq 1} B_i\right) = 1$.
Also gilt $\lim\limits_{n\to\infty} P(|X_n - X| \geq \varepsilon) = \lim\limits_{n\to\infty}(1 - P(B_n)) = 0$ □

Bemerkung 13.4 *Konvergiert eine Folge $(X_n)_{n\geq 1}$ stochastisch gegen $X : \Omega \to \mathbb{R}$, so braucht sie nicht fast sicher gegen X konvergieren.*
Es braucht nicht einmal ein einziges $\omega \in \Omega$ zu geben mit $\lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$.
Sei z.B. $\Omega = [0, 1[\subset \mathbb{R}$ und $\mathcal{A} = B(\mathbb{R}) \cap P(\Omega)$ mit $P =$Gleichverteilung.
Für $n \in \mathbb{N}_+$ sei $c_n := 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}$ und
$A_n := \{\omega \in \Omega \mid \exists m \in \mathbb{N}\ \omega + m \in [c_n, c_{n+1}[\}$, sowie
$X_n := \chi_{A_n} : \Omega \to \mathbb{R}$, def. durch $\omega \mapsto \begin{cases} 1 & \text{falls } \omega \in A_n \\ 0 & \text{sonst} \end{cases}$.

1. *Die Folge $(X_n)_{n\geq 1}$ konvergiert stochastisch gegen die Nullfunktion, denn für $0 < \varepsilon < 1$ gilt:*
   $P(|X_n - 0| \geq \varepsilon) = P(\{\omega \in \Omega \mid X_n(\omega) \geq \varepsilon\}) = P(A_n) = \frac{1}{n+1} \to 0$ für $n \to \infty$
2. *Für jedes $\omega \in \Omega$ und jedes $n \geq 1$ gibt es eine Zahl $i \geq n$ mit $\omega \in A_i$, denn die Reihe $(c_j)_{j\geq 1}$ divergiert.*
   Also gilt $\lim\limits_{n\to\infty} \sup_{i\geq n} X_n(\omega) = 1$.
   Für jedes $\omega \in \Omega$ und jedes $n \geq 1$ gibt es auch eine Zahl $j \geq n$ mit $\omega \notin A_j$, denn $\omega \in A_i \Rightarrow \omega \notin A_{i+1}$.
   Somit gilt $\lim\limits_{n\to\infty} \inf_{i\geq n} X_n(\omega) = 0$ und für <u>kein</u> $\omega \in \Omega$ gilt $\lim\limits_{n\to\infty} X_n(\omega) = X(\omega) = 0$

Lemma 13.5 *(Lemma von Borel-Cantelli)*
Seien $A_n \in \mathcal{A}$ für $n \in \mathbb{N}_+$ Ereignisse und sei
$B := \{\omega \in \Omega \mid \omega \in A_n$ für unendlich viele $n \in \mathbb{N}_+\}$.

1. *Gilt $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$, so ist $P(B) = 0$.*

2. *Sind die Ereignisse $\{A_n\}_{n\in\mathbb{N}_+}$ unabhängig und es gilt $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$, so folgt $P(B) = 1$.*

Beweis:

1. Es gilt $B = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k\geq n} A_k$.
   Für jedes $n \in \mathbb{N}_+$ ist dabei
   $P(B) \leq P\left(\bigcup\limits_{k\geq n} A_k\right) \leq \sum\limits_{k\geq n} P(A_k) \to 0$ für $n \to \infty$.
2. Für $i \in \mathbb{N}_+$ sei $\alpha_i \in [0,1]$.
   Für $m, n \in \mathbb{N}_+$ mit $m < n$ gilt dann
   $\ln(1-\alpha_i) \leq -\alpha_i \Rightarrow \ln\left(\prod\limits_{i=m}^{n}(1-\alpha_i)\right) \leq -\sum\limits_{i=m}^{n} \alpha_i$
   $\Rightarrow \prod\limits_{i=m}^{n}(1-\alpha_i) \leq \exp\left(-\sum\limits_{i=m}^{n}\alpha_i\right)$
   Da die Ereignisse $\{A_n\}_{n\geq 1}$ unabhängig sind, folgt
   $P\left(\bigcap\limits_{i=m}^{n}(\Omega\setminus A_i)\right) = \prod\limits_{i=m}^{n} P(\Omega\setminus A_i) = \prod\limits_{i=m}^{n}(1-P(A_i)) \leq \exp\left(-\sum\limits_{i=m}^{n} P(A_i)\right)$
   $\Rightarrow P\left(\bigcap\limits_{i=m}^{\infty}(\Omega\setminus A_i)\right) = \lim\limits_{n\to\infty} P\left(\bigcap\limits_{i=m}^{n}(\Omega\setminus A_i)\right) = 0$, da nach Voraussetzung $\lim\limits_{n\to\infty}\sum\limits_{i=m}^{n} P(A_i) = \infty$ gilt.
   $\Rightarrow P(\Omega\setminus B) = P\left(\bigcup\limits_{m=1}^{\infty}\bigcap\limits_{i=m}^{\infty}(\Omega\setminus A_i)\right) \leq \sum\limits_{m=1}^{\infty} P\left(\bigcap\limits_{i=m}^{\infty}(\Omega\setminus A_i)\right) = 0$
   $\Rightarrow P(B) = 1$

□

Definition 13.6 Eine Aussage gilt für ***fast alle $\boldsymbol{\omega \in \Omega}$*** (bzw. ***gilt fast sicher*** bzw. ***fast überall***), wenn die Menge B aller $\omega \in \Omega$, für die die Aussage gilt, die Wahrscheinlichkeit $P(B) = 1$ hat.

Korollar 13.7 *Sind $\{A_n\}_{n\geq 1} \subset \mathcal{A}$ Ereignisse, so gilt:*

1. *Konvergiert $\sum\limits_{n=1}^{\infty} P(A_n)$, so gehört fast jedes $\omega \in \Omega$ nur zu endlich vielen Ereignissen A_n.*
2. *Ist $\{A_n\}_{n\geq 1}$ unabhängig und gilt $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$, so gehört fast jedes $\omega \in \Omega$ zu unendlich vielen Ereignissen A_n.*

Beispiel 13.8 Für jedes $n \in \mathbb{N}_+$ sei eine Urne U_n gegeben, die 1 weiße und $f(n) \geq 1$ schwarze Kugeln enthält.

Man zieht aus jeder Urne eine Kugel.

$A_n =$ „aus U_n wird die weiße Kugel gezogen“, $P(A_n) = \frac{1}{f(n)+1}$

1. Gilt $f(n) = n$, so werden fast sicher unendlich viele weiße Kugeln gezogen, denn $\sum\limits_{n\geq 1} P(A_n) = \sum\limits_{n\geq 1} \frac{1}{n+1} = \infty$.
2. Gilt $f(n) = n^2 - 1$, so werden fast sicher nur endlich viele weiße Kugeln gezogen, denn $\sum\limits_{n\geq 1} P(A_n) = \sum\limits_{n\geq 1} \frac{1}{n^2} = \frac{\pi^2}{6} < \infty$.
3. Egal, wie $f(n)$ beschaffen ist. Stets ist die Wahrscheinlichkeit dafür, dass unendlich viele weiße Kugeln gezogen werden gleich 0 oder 1.

Theorem 13.9 *(Das starke Gesetz der großen Zahlen)*
Sei $N \in \mathbb{R}_+$. Für jedes $n \in \mathbb{N}_+$ sei $X_n : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable, deren Varianz existiert und $\operatorname{Var}(X_n) < V$ für ein $V > 0$ gilt.
Sind Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}_+}$ paarweise unkorreliert (d.h. $\operatorname{Cov}(X_i, X_j) = 0$ für $i \neq j$), so konvergiert die Folge $Z_n = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - E(X_i))$ fast sicher gegen die Nullfunktion.
Im Fall $E := E(X_1) = E(X_2) = \ldots$ konvergiert die Folge $Y_n = \frac{1}{n}(X_1 + \ldots + X_n)$ insbesondere fast sicher gegen die konstante Funktion E.

Beweis:

1. Schritt: Die Folge $\{Z_{n^2}\}_{n\geq 1}$ konvergiert fast sicher gegen 0.
   Sei o.E. $E(X_i) = 0$ für alle $i \in \mathbb{N}_+$.
   Nach Voraussetzung gilt $\operatorname{Var}(Z_{n^2}) = \frac{1}{n^4}(\operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_{n^2})) \leq \frac{V}{n^2}$
   Nach der Tschebyschew-Ungleichung folgt für alle $\varepsilon > 0$ dann
   $P(|Z_{n^2}| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\operatorname{Var}(Z_{n^2}) \leq \frac{V}{\varepsilon^2 n^2}$.
   Für $A_n = \{|Z_{n^2}| \geq \varepsilon\}$ gilt folglich $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$.
   Nach dem Lemma gilt, dass fast jedes $\omega \in \Omega$ nur zu endlich vielen A_n gehört.
   Für $k \in \mathbb{N}_+$ wähle $\varepsilon = \frac{1}{k}$ und betrachte
   $E_k = \left\{\omega \in \Omega \mid |Z_{n^2}(\omega)| \geq \frac{1}{k} \text{ für unendlich viele } n\right\}$
   $\Rightarrow P(E_k) = 0 \Rightarrow P\left(\bigcup\limits_{k\geq 1} E_k\right) = 0 \Rightarrow P\left(\Omega \setminus \bigcup\limits_{k\geq 1} E_k\right) = 1$
   Für $\omega \in \Omega \setminus \bigcup\limits_{k\geq 1} E_k$ gibt es zu jedem $k \geq 1$ nur endlich viele n mit $|Z_{n^2}(\omega)| \geq \frac{1}{k}$,
   d.h. es gilt $\lim\limits_{n\to\infty} Z_{n^2}(\omega) = 0$
   $\Rightarrow P\left(\left\{\omega \in \Omega \mid \lim\limits_{n\to\infty} Z_{n^2}(\omega) = 0\right\}\right) = 1$

2. Schritt: Für jedes $m \in \mathbb{N}_+$ wähle $n = n(m) \in \mathbb{N}_+$ so, dass $n^2 \leq m < (n+1)^2$
Es sollen Z_m und Z_{n^2} verglichen werden.
Für alle $k \in \mathbb{N}_+$ sei $S_k = X_1 + \ldots + X_k$
$\Rightarrow \mathrm{Var}(S_m - S_{n^2}) = \mathrm{Var}(X_{n^2+1} + \ldots + X_m) = \mathrm{Var}(X_{n^2+1}) + \ldots + \mathrm{Var}(X_m) \leq V(m - n^2)$
Nach der Tschebyschew-Ungleichung folgt für $\varepsilon > 0$ die Abschätzung
$P(|S_m - S_{n^2}| \leq \frac{V}{\varepsilon^2 n^4}(m - n^2)$
$\Rightarrow \sum\limits_{m=1}^{\infty} P\left(\frac{1}{n(m)^2}|S_m - S_{n(m)^2}| \geq \varepsilon\right) \leq \frac{V}{\varepsilon^2} \sum\limits_{n=1}^{\infty} \sum\limits_{m=n^2}^{(n+1)^2-1} \frac{m-n^2}{n^4} =$
$\frac{V}{\varepsilon^2} \sum\limits_{n=1}^{\infty} \frac{1}{n^4}(1 + 2 + \ldots + 2n) = \frac{V}{\varepsilon^2} \sum\limits_{n=1}^{\infty} \frac{(2n)(2n+1)}{2n^4} < \infty$
Nach Borel-Cantelli gilt dann für fast alle $\omega \in \Omega$, dass sie nur in endlich vielen Mengen
$\left\{\omega \in \Omega \mid |S_m(\omega) - S_{n(m)^2}(\omega)| \geq \varepsilon n(m)^2\right\}$ enthalten sind.
$\Rightarrow$ für fast alle $\omega \in \Omega$ gibt es ein $m(\omega)$ mit $\frac{1}{n(m)^2}|S_m(\omega) - S_{n(m)^2}(\omega)| < \varepsilon$ für $m \geq m(\omega)$.
Nach 1. Schritt gilt für fast alle $\omega \in \Omega$ und für $m \gg 0$:
$\frac{1}{n(m)^2}|S_{n(m)^2}| < \varepsilon$
$\Rightarrow$ für fast alle $\omega \in \Omega$ und für $m \gg 0$ gilt
$|Z_m(\omega)| = \frac{1}{m}|S_m(\omega)| \leq \frac{1}{n(m)^2}|S_m(\omega)| \overset{\triangle\text{-Ungl.}}{<} 2\varepsilon$
Für $k \in \mathbb{N}_+$ wähle $\varepsilon = \frac{1}{k}$ und
$E_k = \left\{\omega \in \Omega \mid |Z_m(\omega)| > \frac{1}{k} \text{ für unendlich viele } m\right\}$
$$\Rightarrow P(E_k) = 0 \Rightarrow P\left(\Omega \setminus \bigcup_{k \geq 1} E_k\right) = 1$$
Für $\omega \in \Omega \setminus \bigcup\limits_{k \geq 1} E_k$ gilt $\lim\limits_{m \to \infty} Z_m(\omega) = 0$
$$\Rightarrow P\left(\left\{\omega \in \Omega \mid \lim_{m \to \infty} Z_m(\omega) = 0\right\}\right) = 1$$
□

Beispiel 13.10 (6-er Runs beim Würfeln)

Ein idealer Würfel werde unendlich oft geworfen.

$\Omega = \{1, \ldots, 6\}^{\mathbb{N}}$ Ergebnismenge, $\mathcal{A} = \mathfrak{P}(\Omega)$, $P : \mathcal{A} \to [0,1]$,

wobei für $A \in \mathcal{A}$ und $n \in \mathbb{N}_+$ gelte

$A_n = \{(a_1, \ldots, a_n) \in \{1, \ldots, 6\}^n \mid \exists a_{n+1}, a_{n+2}, \ldots \in \{1, \ldots, 6\}$ mit $(a_1, a_2, \ldots) \in A\}$ und $A \mapsto \lim\limits_{n \to \infty} \underbrace{\frac{\#A_n}{6^n}}_{\text{monoton fallende Folge}}$.

Für $n \in \mathbb{N}_+$ sei $X_n : \Omega \to \{0,1\}$ das Ergebnis des n-ten Versuchs $(a_1, a_2, \ldots) \mapsto a_n$ (die Projektion auf die n-te Komponente).

Ein k-Run von 6-ern ist das Ereignis $X_n(\omega) = 1, \ldots, X_{n+k-1}(\omega) = 1$ für ein

$n \in \mathbb{N}_+$.

1. Aus dem Lemma von Borel-Cantelli folgt, dass für jedes $N \in \mathbb{N}_+$ fast jedes $\omega \in \Omega$ eine Folge von N 6-ern enthält, d.h. einen 6-er Run der Länge N.
2. Das schwache Gesetz der großen Zahlen besagt:
   Für jedes $\varepsilon > 0$ gilt
   $P\left(\left|\frac{1}{n}(X_1 + \ldots + X_n) - \frac{1}{6}\right| > \varepsilon\right) \to 0$ für $n \to \infty$.
   Es wäre aber immer noch möglich, dass für fast jedes $\omega \in \Omega$ beliebig große n mit $\frac{1}{n}(X_1(\omega) + \ldots + X_n(\omega)) \approx 1$ existieren.
3. Nach dem starken Gesetz der großen Zahl, gilt jedoch für fast alle $\omega \in \Omega$ die Konvergenz $\frac{1}{n}(X_1(\omega) + \ldots + X_n(\omega)) \to \frac{1}{6}$
   „Fast sicher kommen lange 6-er Runs erst so spät, dass sie die relative Häufigkeit nicht mehr stark beeinflussen.“

Definition 13.11

1. Ist $F : \mathbb{R} \to \mathbb{R}$ eine Funktion, so heißt
   $\|F\| := \sup_{x\in\mathbb{R}} |F(x)| \in \mathbb{R} \cup \{\infty\}$ die ***Supremumsnorm von F***.
2. Sind $F, F_1, F_2, \ldots : \mathbb{R} \to \mathbb{R}$ Funktionen mit $\lim\limits_{n\to\infty} \|F - F_n\| = 0$, so sagen wirm die Folge $(F_n)_{n\geq 1}$ ***konvergiert gleichmäßig*** auf $\mathbb{R}$ gegen F.
3. Sind $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen mit Verteilungsfunktionen $F_X, F_Y : \mathbb{R} \to [0, 1]$, so schreiben wir $d(F_X, F_Y) := \|F_X - F_Y\|$.

Lemma 13.12 *Sei $\Phi : \mathbb{R} \to [0, 1]$ die Verteilungsfunktion der Standardnormalverteilung.*
Sind $X, Y : \Omega \to \mathbb{R}$ zwei Zufallsvariablen, wobei $E(Y^2)$ existiere und $E(Y) = 0$ sei.
Dann gilt:
$d(F_{X+Y}, \phi) \leq d(F_X, \Phi) + 2\sqrt[3]{E(Y^2)}$

Beweis:
Für alle $x \in \mathbb{R}$ gilt $\Phi'(x) = \varphi(x) \leq \varphi(0) = \frac{1}{\sqrt{2\pi}} < 1$
$\Rightarrow$ Für alle $\delta > 0$ und für alle $x \in \mathbb{R}$ gilt $|\Phi(x \pm \delta) - \Phi(x)| \leq \delta$.
Sei $\delta > 0$ und $x \in \mathbb{R}$.
Ist $X \leq x - \delta$, so ist $Y > \delta$ oder $X + Y \leq x$.
Ist $X + Y \leq x$, so ist $X \leq x + \delta$ oder $Y < -\delta$
$\Rightarrow P(X \leq x - \delta) - P(Y > \delta) \leq P(X + Y \leq x) \leq P(X \leq x + \delta) + P(Y < -\delta)$
$\Rightarrow P(X \leq x - \delta) - \Phi(x - \delta) - \delta - P(Y > \delta) \leq P(X + Y \leq x) - \Phi(x) \leq$

$P(X \leq x+\delta) - \Phi(x+\delta) + \delta + P(Y < -\delta)$
$\Rightarrow d(F_{X+Y}, \Phi) \leq d(F_X, \Phi) + \delta + P(|Y| > \delta)$.
Wähle nun $\delta = \sqrt[3]{E(Y^2)}$.
Dann gilt nach der Tschebyschew-Ungleichung
$P(|Y| > \delta) \leq \frac{1}{\delta^2} \operatorname{Var}(Y) = \frac{1}{\delta^2} E(Y^2) = \frac{1}{\delta^2}\delta^3 = \delta$ und wir erhalten
$d(F_{X+Y}, \Phi) \leq d(F_X, \Phi) + 2\sqrt[3]{E(Y^2)}$ □

Satz 13.13 *Für jedes $n \in \mathbb{N}_+$ sei $X_n : \Omega \to \mathbb{R}$ eine Zufallsvariable mit endlich vielen Werten, $E(X_n) = 0$ und $\operatorname{Var} X_n = 1$.*
Ferner sei $\{X_n\}_{n\in\mathbb{N}_+}$ unabhängig und es gebe ein $B \in \mathbb{R}_+$ mit $|X_n(\omega)| < B$ für alle $\omega \in \Omega$ und $n \in \mathbb{N}_+$.
Ist dann $S_n^* = \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)$ die Standardisierung von $S_n = X_1 + \ldots + X_n$, so gilt $\lim\limits_{n\to\infty} d(F_{S_n^*}, \Phi) = 0$.

Beweis:

1. Schritt:
   Sei $X_0 : \Omega \to \mathbb{R}$ eine weitere Zufallsvariable, die $N(0,1)$-verteilt und unabhängig von $\{X_n\}_{n\geq 1}$ ist.
   Sei $N \in \mathbb{N}_+$ vorgegeben.
   Für jedes $n \geq N$ betracxhten wir $Z_n := \sqrt{\frac{N}{n}} X_0 + \sqrt{\frac{1}{n}}(S_n - S_N) : \Omega \to \mathbb{R}$ und bezeichnen die Verteilungsfunktion von Z_n mit $F_n : \mathbb{R} \to [0,1]$.
   Beachte, dass dabei $F_N = \Phi$.
   Wir studieren zunächst die Folge $\{d(F_n, \Phi)\}_{n\geq N}$.
   Aus der Definition von Z_n ergibt sich
   $Z_{n+1} = \sqrt{\frac{n}{n+1}} Z_n + \sqrt{\frac{1}{n+1}} X_{n+1}$.
   Sei $X_{n+1}(\Omega) = \{x_1, \ldots, x_k\}$ und $P(X_{n+1} = x_i) =: p_i$ für $i = 1, \ldots, k$.
   Dann folgt aus der Unabhängigkeit von Z_n und X_{n+1} die Formel
   $F_{n+1}(x) = \sum\limits_{i=1}^{k} P(X_{n+1} = x_i, \sqrt{\frac{n}{n+1}} Z_n + \sqrt{\frac{1}{n+1}} X_{n+1} \leq x) =$
   $\sum\limits_{i=1}^{k} p_i F_n\left(\sqrt{\frac{n+1}{n}} x - \sqrt{\frac{1}{n}} x_i\right) = E(Y_x)$
   mit $Y_x : \Omega \overset{X_{n+1}}{\to} \mathbb{R} \overset{l_x}{\to} \mathbb{R} \overset{F_n}{\to} \mathbb{R}$, wobei l_x, def. ist durch $y \mapsto \sqrt{\frac{n+1}{n}} x - \sqrt{\frac{1}{n}} y$.
   Ebenso betrachte die Zufallsvariable $\tilde{\Phi}_x : \Omega \overset{X_{n+1}}{\to} \mathbb{R} \overset{l_x}{\to} \mathbb{R} \overset{\Phi}{\to} \mathbb{R}$
   Es folgt:
   $|F_{n+1}(x) - \Phi(x)| \leq |E(Y_x) - E(\tilde{\Phi}_x)| + |E(\tilde{\Phi}_x) - \Phi(x)|$
   Für $i = 1, \ldots, k$ gilt:
   $|(Y_x - \tilde{\Phi_x}(\omega)| = \left| F_n\left(\sqrt{\frac{n+1}{n}} x - \sqrt{\frac{1}{n}} x_i\right) - \Phi\left(\sqrt{\frac{n+1}{n}} x - \sqrt{\frac{1}{n}} x_i\right)\right| \leq d(F_n, \Phi)$
   Also folgt:

$|E(Y_x) - E(\tilde{\Phi}_x)| = |E(Y_x - \tilde{\Phi}_x)| \leq \sum_{i=1}^{k} p_i |\dots| \leq$
$\left(\sum_{i=1}^{k} p_i\right) d(F_n, \Phi) = d(F_n, \Phi)$
Damit erhalten wir
$d(F_{n+1}, \Phi) \leq d(F_n, \Phi) + \sup_{x\in\mathbb{R}} |E(\tilde{\Phi}_x) - \Phi(x)|$

2. Abschätzung von $|E(\tilde{\Phi}_x) - \Phi(x)|$
   Schreibe $\alpha := \sqrt{\frac{n+1}{n}}$ und $\beta := \sqrt{\frac{1}{n}}$.
   Dann sind die Werte von $\tilde{\Phi}_x$ gegeben durch $\{\Phi(\alpha x - \beta x_i)\}_{i=1,\dots,k}$ und es gilt
   $E(\tilde{\Phi}_x) = \sum_{i=1}^{k} p_i \Phi(\alpha x - \beta x_i)$.
   Die Taylorentwicklung von Φ liefert
   $\Phi(\alpha x - \beta x_i) = \Phi(\alpha x) - \beta x_i \Phi'(\alpha x) + \frac{(\beta x_i)^2}{2}\Phi''(\alpha x) - \frac{(\beta x_i)^3}{6}\Phi'''(\alpha x - \vartheta_i(x)\beta x_i)$
   mit $|\vartheta_i(x)| \leq 1$ (Lagrange-Restglied)
   Nach Voraussetzung gilt $|x_i| < B$ für $i = 1, \dots, k$.
   Die Funktion $\Phi''' = \varphi''$ ist beschränkt und es gilt $\beta^3 = n^{-\frac{3}{2}}$.
   $\Rightarrow \left|-\frac{(\beta x_i)^3}{6}\Phi'''(\alpha x - \vartheta_i(x)\beta x_i)\right| = O\left(n^{-\frac{3}{2}}\right)$
   Weiter gilt $\alpha = \sqrt{\frac{n+1}{n}} = \sqrt{1 + \frac{1}{n}} = 1 + \frac{2n}{+} O(n^{-2})$ und $\Phi, \Phi' = \varphi, \Phi'', x\Phi', x^2\Phi''$ beschränkt.
   $\Rightarrow$ Taylorentwicklung von $\Phi(\alpha x)$ liefert
   $\Phi(\alpha x) = \Phi(x + (\frac{1}{2n} + O(n^{-2}))x) =$
   $\Phi(x) + \left(\frac{1}{2n} + O(n^{-2})\right) x\Phi'(x) + \frac{1}{2}\left(\frac{1}{2n} + O(n^{-2})\right)^2 x^2 \Phi''\left(x + \vartheta\left(\frac{1}{2n} + O(n^{-2})\right)x\right)$
   mit $0 \leq |\vartheta| \leq 1$
   $\Phi(\alpha x) = \Phi(x) + \frac{x}{2n}\Phi'(x) + O(n^{-2})$
   Analog ergibt sich $\Phi''(\alpha x) = \Phi''(x) + \frac{x}{2n}\Phi''(x) + \frac{x}{2n}\Phi'''(x) + O(n^{-2})$.
   Insgesamt erhalten wir:
   $E(\tilde{\Phi}_x) = \sum_{i=1}^{k} p_i \Phi(\alpha x - \beta x_i) =$
   $\sum_{i=1}^{k} p_i \left[\Phi(\alpha x) - \beta x_i \Phi'(\alpha x) + \frac{(\beta x_i)^2}{2}\Phi''(\alpha x) + O(n^{-\frac{3}{2}})\right] =$
   $\sum_{i=1}^{k} p_i \left[\Phi(x) + \frac{x}{2n}\Phi'(x) - \beta x_i \Phi'(\alpha x) + \frac{(\beta x_i)^2}{2}\left(\Phi''(x) + \frac{x}{2n}\Phi'''(x)\right) + O(n^{-\frac{3}{2}})\right] =$
   $\Phi(x) + \frac{x}{2n}\Phi'(x) - \frac{1}{\sqrt{n}}\Bigg(\underbrace{\sum_{i=1}^{n} p_i x_i}_{=E(X_{n+1})=0}\Bigg)\Phi'(\alpha x) + \frac{1}{2n}\Bigg(\underbrace{\sum_{i=1}^{k} p_i x_i^2}_{=\mathrm{Var}(X_{n+1})=1}\Bigg)\Phi''(x) + O(n^{-\frac{3}{2}}) =$
   $\Phi(x) + \frac{x}{2n}\Phi'(x) + \frac{1}{2n}\Phi''(x) + O(n^{-\frac{3}{2}}) =$
   $\Phi(x) + O(n^{-\frac{3}{2}})$, weil $\Phi''(x) = \varphi'(x) = -x\varphi(x) = -x\Phi'(x)$
   $\Rightarrow |d(F_{n+1}, \Phi) - d(F_n, \Phi)| \leq |E(\tilde{\Phi}_x) - \Phi(x)| \leq kn^{-\frac{3}{2}}$ für ein $k > 0$

3. Vollendung des Beweises
   Es gilt $F_N = \Phi$, also $d(F_n, \Phi) = 0$ und für $n \geq N$ folgt
   $d(F_n, \Phi) \leq \sum\limits_{i=N}^{n-1} |d(F_{i+1}, \Phi) - d(F_i, \Phi)| \leq k \sum\limits_{i=N}^{n-1} i^{-\frac{3}{2}} \leq k \sum\limits_{i=N}^{\infty} i^{-\frac{3}{2}}$
   Wegen $\sum\limits_{i=1}^{\infty} i^{-\frac{3}{2}} < \infty$ gilt $\lim\limits_{N\to\infty} \left(\sum\limits_{i=N}^{\infty} i^{-\frac{3}{2}} \right) = 0.$
   Also gibt es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}_+$ mit $d(F_n, \Phi) < \frac{\varepsilon}{2}$ für $n \geq N$.
   Es gilt $E((S_n^* - Z_n)^2) = E\left(\left(\frac{1}{\sqrt{n}} S_n - \sqrt{\frac{N}{n}} X_0 - \frac{1}{\sqrt{n}} S_n + \frac{1}{\sqrt{n}} S_N \right)^2 \right) =$
   $\frac{1}{n} E((\sqrt{N} X_0 - S_N)^2) \to 0$ für $n \to \infty$ und
   $E(S_n^* - Z_n) = \frac{1}{\sqrt{n}} E(\sqrt{N} X_0 - S_N) = 0$
   Nach Lemma 13.12 folgt
   $d(F_{S_n^*}, \Phi) = d(F_{Z_n + (S_n^* - Z_n)}, \Phi) \leq \underbrace{d(F_{Z_n}, \Phi)}_{< \frac{\varepsilon}{2} \text{ für } n \gg 0} + \underbrace{2 \sqrt[3]{E((S_n^* - Z_n)^2)}}_{< \frac{\varepsilon}{2} \text{ für } n \gg 0} \to 0$
   für $n \to \infty$

$\square$

Theorem 13.14 *(Der zentrale Grenzwertsatz)*
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und seien $X_n : \Omega \to \mathbb{R}$ für $n \in \mathbb{N}_+$ unabhängige Zufallsvariablen mit identischen Verteilungen und endlicher Varianz $\mathrm{Var}(X_n) = \sigma^2 \in \mathbb{R}_+$.
Ist $S_n = X_1 + \ldots + X_n$ und $S_n^* = \frac{1}{\sqrt{n}\sigma}(S_n - nE(X_1))$, so gilt:
$\lim\limits_{n\to\infty} d(F_{S_n^*}, \Phi) = 0$, d.h. die Verteilungsfunktion von S_n^* konvergiert gleichmäßig gegen die Verteilungsfunktion Φ der Standardnormalverteilung.

Beweis:
Sei $\mu := E(X_1) = E(X_2) = \ldots$.
Für $m \in \mathbb{N}_+$ betrachte $f_m : \mathbb{R} \to \mathbb{R}$,
def. durch $x \mapsto \begin{cases} \frac{k}{m} & \text{falls } x \in \left[\frac{k}{m}, \frac{k+1}{m}\right[\text{ und } -m^2 \leq k \leq m^2 \\ 0 & \text{sonst} \end{cases}$
Ferner definiere für $i, m \in \mathbb{N}_+$ Zufallsvariable $Y_{mi} := f_m \circ \mathbb{R}$ und setze $\mu_m = E(Y_{mi})$, sowie $\sigma_m = \sigma(Y_{mi})$.
Hierbei hängen μ_m und σ_m nicht von i ab, weil alle Zufallsvariablen X_i und damit alle $\{Y_{mi}\}_{i\geq 1}$ dieselbe Verteilungsfunktion besitzen.
Die Zufallsvariablen Y_{mi} haben nur endlich viele Werte.
Es gilt $E(Y_{mi}) = \sum\limits_{-m^2 \leq k \leq m^2} \frac{k}{m} P\left(\frac{k}{m} \leq X_i < \frac{k+1}{m}\right)$
$E(X_i) = \lim\limits_{m\to\infty} \sum\limits_{k\in\mathbb{Z}} \frac{k}{m} P\left(\frac{k}{m} \leq X_i < \frac{k+1}{m}\right)$
Hieraus folgt $\lim\limits_{m\to\infty} \underbrace{E(Y_{mi})}_{\mu_m} = E(X_i)$
Analog sieht man $\lim\limits_{m\to\infty} \sigma_m^2 = \sigma^2$.

Ebenso ergibt sich
$\lim\limits_{m\to\infty} E(Y_{mi}\cdot X_i) = \lim\limits_{m\to\infty} \sum\limits_{k\in\mathbb{Z}} \frac{k}{m} P\left(\frac{k}{m} \le Y_{mi}\cdot X_i < \frac{k+1}{m}\right) = \lim\limits_{m\to\infty} E(Y_{mi}^2) = E(X_i^2).$
Hieraus folgt $\lim\limits_{m\to\infty} E((Y_{mi}-X_i)^2) = \lim\limits_{m\to\infty} E(Y_{mi}^2) - 2\lim\limits_{m\to\infty} E(Y_{mi}X_i) + E(X_i^2) = 0$
Für $i,m\in\mathbb{N}_+$ sei nun $X_{mi} := \frac{1}{\sigma_m}(Y_{mi}-\mu_m) : \Omega\to\mathbb{R}$ die Standardisierung von Y_{mi} und für $n\in\mathbb{N}_+$ sei $S_{mn}^* = \frac{1}{\sqrt{n}}(X_{m_1}+\ldots+X_{mn})$.
Die Zufallsvariablen $\{X_{mi}\}_{i\in\mathbb{N}_+}$ sind unabhängig, haben nur endlich viele Werte, sind standardisiert und erfüllen
$|X_{mi}(\omega)| \le \frac{1}{\sigma_m}(m+|\mu_m|)$ für alle $i\in\mathbb{N}_+$ und alle $\omega\in\Omega$.
Daher zeigt Satz 13.13, dass $\lim\limits_{n\to\infty} d(F_{S_{mn}^*},\Phi) = 0$ gilt.
Wir berechnen $S_n^* - S_{mn}^* = \frac{1}{\sqrt{n}\sigma}\sum\limits_{i=1}^n (X_i-\mu) - \frac{1}{\sqrt{n}\sigma_m}\sum\limits_{i=1}^n (Y_{mi}-\mu_m) =$

$$= \underbrace{\left(\frac{\sigma_m}{\sigma}-1\right) S_{mn}^*}_{=:A_{mn}} + \underbrace{\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n [(X_i-\mu)-(Y_{mi}-\mu_m)]}_{=:B_{mn}}$$

Es gilt $E(A_{mn}^2) = \left(\frac{\sigma_m}{\sigma}-1\right)^2 E((S_{mn}^*)^2) = \left(\frac{\sigma_m}{\sigma}-1\right)^2 \mathrm{Var}(S_{mn}^*) = \left(\frac{\sigma_m}{\sigma}-1\right)^2$ (*) und

$$E(B_{mn}^2) = \frac{1}{n\sigma^2} E\left(\Big(\sum_{i=1}^m \underbrace{[(X_i-\mu)-(Y_{mi}-\mu_i)]}_{\substack{\text{unabhängige, identisch verteilte}\\ \text{Zufallsvariablen mit Erwartungswert 0}}}\Big)^2\right) =$$

$\frac{1}{n\sigma^2}\sum\limits_{i=1}^n \mathrm{Var}(X_i - Y_{mi}) = \frac{1}{\sigma^2}\mathrm{Var}(X_1 - Y_{m1}) \le$
$\frac{1}{\sigma^2}E((X_1-Y_{m1})^2) \to 0$ für $m\to\infty$.
Ebenso: (*)$\to 0$ für $m\to\infty$
$\Rightarrow E((S_n^*-S_{mn}^*)^2) = E((A_{mn}+B_{mn})^2) \le$
$E(A_{mn})^2 + 2|E(A_{mn}B_{mn})| + E(B_{mn}^2) \overset{\text{Cauchy-Schwarz}}{\le}$
$E(A_{mn}^2) + 2\sqrt{E(A_{mn}^2)+E(B_{mn}^2)} + E(B_{mn}^2) \to 0$ für $m\to\infty$
Weiter gilt $E(S_n^* - S_{mn}^*) = 0$.
Nach Lemma 13.12 folgt
$d(F_{S_n^*},\Phi) = d(F_{S_{mn}^*+(S_n^*-S_{mn}^*)},\Phi) \le d(F_{S_{mn}^*},\Phi) + 2\sqrt[3]{E((S_n^*-S_{mn}^*)^2)} \to 0$ für $m,n\to\infty$ □

Satz 13.15 *(Qualität der Abschätzung im zentralen Grenzwertsatz)*
Seien $X_n : \Omega\to\mathbb{R}$ für $n\in\mathbb{N}_+$ unabhängige und identisch verteilte Zufallsvariablen und S_n^* sei wie in Theorem 13.14.
Ist $0 < \sigma^2 := \mathrm{Var}(X_1) < \infty$ und $\gamma := E(|X_1 - E(X_1)|^3) < \infty$, so gilt $d(S_n^*,\Phi) \le 0{,}8\frac{\gamma}{\sigma^3\sqrt{n}}$ (mit $\sigma = \sigma(X_1)$).
(ohne Beweis)

Beispiel 13.16 (Relative Häufigkeiten beim Münzenwurf)

Wie oft muss man eine ideale Münze werfen, damit die relative Häufigkeit von „Wappen" von $p = \frac{1}{2}$ mit Wahrscheinlichkeit 95% um höchstens 0,01 abweicht (bzw. 0,001)?

Sei $\Omega = \{0,1\}^N$ mit $N \gg 0$, sei P die N-te Bernoulli-Verteilung zum Parameter $p = \frac{1}{2}$, wobei 0 $\cong$Zahl und 1$\cong$Wappen.

Für $n \geq 1$, sei $S_n : \Omega \to \mathbb{R}$, def. durch $(\omega_1, \ldots, \omega_N) \mapsto \omega_1 + \ldots + \omega_n$ Zahl der Wappen bei den ersten n Versuchen.

$\Rightarrow E(S_n) = \frac{n}{2}$, $\operatorname{Var}(S_n) = n \operatorname{Var}(S_1) = \frac{n}{4}$, $\sigma(S_n) = \frac{\sqrt{n}}{2}$, $\sigma = \sigma(S_1) = \frac{1}{2}$

Die Standardisierung ist

$S_n^* = \frac{1}{\sqrt{n}\sigma}(S_n - nE(S_1)) = \frac{2}{\sqrt{n}}(S_n - \frac{n}{2}) = \frac{2}{\sqrt{n}}(S_n - \frac{n}{2})$

$S_n = \frac{\sqrt{n}}{2} S_n^* + \frac{n}{2}$

$\frac{S_n}{n} = \frac{1}{2\sqrt{n}} S_n^* + \frac{1}{2}$ relative Häufigkeit

Nach dem zentralen Grenzwertsatz gilt

$P(0,5 - \varepsilon \leq \frac{S_n}{n} < 0,5 + \varepsilon) =$

$P\left(0,5 - \varepsilon \leq \frac{1}{2\sqrt{n}} S_n^* + 0,5 < 0,5 + \varepsilon\right) =$

$P(-2\sqrt{n}\varepsilon \leq S_n^* \leq 2\sqrt{n}\varepsilon) =$

$P(S_n^* \leq 2\sqrt{n}\varepsilon) - P(S_n^* < -2\sqrt{n}\varepsilon) \approx$

$\Phi(2\sqrt{n}\varepsilon) - \Phi(-2\sqrt{n}\varepsilon) = 2\Phi(2\sqrt{n}\varepsilon) - 1$

Im Fall $\varepsilon = 0,01$ folgt, dass $2\Phi(0,02\sqrt{n}) - 1 \geq 0,95$ gelten muss, also $0,02\sqrt{n} \geq 1,96$ und $n \geq 9604$

Im Fall $\varepsilon = 0,001$ folgt analog, dass $n \geq 960400$ gelten muss.

Beispiel 13.17 Die Batterie eines Handys hält durchschnittlich 50 Stunden mit einer Standardabweichung von 30 Stunden.

Wieviele Batterien muss man kaufen, damit mit 95% Wahrscheinlichkeit das Handy mindestens 5000 Stunden Betrieb durchhält ?

$\Omega = \mathbb{R}$, $\mathcal{A} = \mathrm{B}(\mathbb{R})$, $P =$Lebesgue-Maß

$X_i : \Omega \to \mathbb{R}$ Lebensdauer der i-ten Batterie.

Die Verteilung von X_i sei unbekannt (meist Exponentialverteilung)

$E(X_i) = 50$, $\sigma(X_i) = 30$, $\operatorname{Var}(X_i) = 900$

Für $n \geq 1$ sei $S_n = X_1 + \ldots + X_n$ die Gesamtbetriebszeit des Handys

$\Rightarrow E(S_n) = 50n$, $\operatorname{Var}(S_n) = 900n$, $\sigma(S_n) = 30\sqrt{n}$

$S_n^* = \frac{1}{\sqrt{n}\sigma(X_i)}(S_n - nE(X_i)) = \frac{1}{30\sqrt{n}}(S_n - 50n)$

Nach dem zentralen Grenzwertsatz gilt

$P(S_n \geq 5000) = 1 - P(S_n < 5000) = 1 - P\left(S_n^* < \frac{5000-50n}{30\sqrt{n}}\right) \approx$
$1 - \Phi\left(\frac{5000-50n}{30\sqrt{n}}\right) = \Phi\left(\frac{50n-5000}{30\sqrt{n}}\right) \geq 0,95 \Leftrightarrow$
$\frac{50n-5000}{30\sqrt{n}} \geq 1,645 \Leftrightarrow n \geq 111$

14 Schätztheorie

Gegeben: Wahrscheinlichkeitsexperiment, Versuchsergebnisse,
zu schätzende Größe
Gesucht: ein geeigneter Wahrscheinlichkeitsraum und / oder eine Menge „denkbarer" Wahrscheinlichkeitsverteilungen, parametrisiert durch $\vartheta \in \theta$.

Definition 14.1

1. Ein ***(diskretes) Schätzproblem*** ist ein Tripel $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, g)$ bestehend aus einer höchstens abzählbaren Menge Ω, einer Menge θ, einer Familie $\{P_\vartheta\}_{\vartheta\in\theta}$ von Wahrscheinlichkeitsverteilungen auf Ω und einer Abbildung $g : \theta \to M$ mit Werten in einer Menge M.
2. Ist $(\Omega, \{P\}_{\vartheta\in\theta}, g)$ ein Schätzproblem mit $g : \theta \to M$, so heißt jede Abbildung $T : \Omega \to M$ ein ***Schätzer von g***.

Beispiel 14.2 (Schätzung von Erfolgswahrscheinlichkeiten)
Ein Bernoulli-Versuch mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ werde n mal unabhängig durchgeführt.
Aus den Versuchsergebnissen soll der Parameter p geschätzt werden.
Sei $\Omega = \{0, 1\}^n$, $\theta = [0, 1]$, $P_\vartheta = n$-te Bernoulli-Verteilung zum Parameter $\vartheta \in \theta$, d.h. $P_\vartheta((\omega_1, \ldots, \omega_n)) = \vartheta^{\omega_1+\ldots+\omega_n}(1-\vartheta)^{n-\omega_1-\omega_2-\ldots-\omega_n}$
Zu schätzende Funktion: $g : \theta \to \mathbb{R}$, def. durch $\vartheta \mapsto \vartheta$
Ein möglicher Schätzer ist die relative Häufigkeit:
$T : \Omega \to \mathbb{R}$, def. durch $(\omega_1, \ldots, \omega_n) \mapsto \frac{1}{n}(\omega_1 + \ldots + \omega_n)$

Beispiel 14.3 (Panzernummern im 2. Weltkrieg)
Aus den Seriennummern erbeuteter Panzer soll die Gesamtzahl N der Panzer des Feindes geschätzt werden.
(Annahme: Die Panzer sind mit $1, \ldots, N$ durchnummeriert.)
Es soll n erbeutete Panzer geben.
Sei $\Omega =$Menge der n-elementigen Teilmengen von $\{1, \ldots, N\} \cong$

$\{(x_1, \ldots, x_n) \in \{1, \ldots, N\}^n \mid 1 \leq x_1 < \ldots < x_n \leq N\}$

$\theta = \mathbb{N}_+$ =Menge der Zahlen N

Für $\vartheta \in \theta$ sei P_θ die Gleichverteilung auf Ω.

Zu schätzende Funktion $g : \theta \to \mathbb{R}$, def. durch $N \mapsto N$.

$P_\theta((x_1, \ldots, x_n)) = \frac{1}{\binom{N}{n}}$

Mögliche Schätzer:

1. $T_1 : \Omega \to \mathbb{R}$ $(x_1, \ldots, x_n) \mapsto x_n$ „max. Seriennummer“
2. $T_2 : \Omega \to \mathbb{R}$ $(x_1, \ldots, x_n) \mapsto x_n + x_1 - 1$ „max. Nummer plus erste Lücke“
3. $T_3 : \Omega \to \mathbb{R}$ $(x_1, \ldots, x_n) \mapsto x_n + \frac{x_n - n}{n}$ „max. Nummer plus durchschnittliche Lücke“

Definition 14.4 Sei $(\Omega, \{P_\vartheta\}_{\vartheta \in \theta}$ ein Schätzproblem mit $g : \theta \to M$.

1. Für jedes $\omega \in \Omega$ heißt $L_\omega : \theta \to [0, 1]$, def. durch $\vartheta \to P_\vartheta(\omega)$ die ***Likelihood-Funktion*** an der Stelle $\omega \in \Omega$.
2. Gibt es für jedes $\omega \in \Omega$ ein $\vartheta_\omega \in \theta$ mit $L_\omega(\vartheta_\omega) = \sup\limits_{\vartheta \in \theta} L_\omega(\vartheta)$ so heißt der Schätzer $L : \Omega \to M$, def. durch $\omega \mapsto g(\vartheta_\omega)$ ein ***Maximum-Likelihood-Schätzer***.

Beispiel 14.5 (Fortsetzung von 14.2 und 14.3)

1. In Beispiel 14.2 gilt:
   $L_\omega : \theta = [0, 1] \to [0, 1]$ ist def. durch $\vartheta \mapsto P_\vartheta(\omega) = \theta(1 - \theta)^{n-i}$ mit $i = \omega_1 + \ldots + \omega_n$ für $\omega = (\omega_1, \ldots, \omega_n)$.
   Wegen der Monotonie von $x \mapsto \ln x$ genügt es, das Maximum von $\ln(L_\omega)(\vartheta) = i \ln \vartheta + (n - i) \ln(1 - \theta)$ auf dem Intervall $\theta = [0, 1]$ zu finden.
   Es gilt:
   $\frac{\partial}{\partial \vartheta}(\ln L_\omega(\vartheta)) = i \frac{1}{\vartheta} - (n - i) \frac{1}{1 - \vartheta} = 0 \Leftrightarrow \vartheta = \frac{i}{n}$
   Also ist $L = T : \Omega \Rightarrow \mathbb{R}$ $\omega \mapsto \frac{\omega_1 + \ldots + \omega_n}{n}$ der Maximum-Likelihood-Schätzer.
2. In Beispiel 14.3 gilt:
   $L_\omega : \theta = \mathbb{N}_+ \to [0, 1]$ ist def. durch $\vartheta \mapsto P_\vartheta(\omega) = \frac{1}{\binom{\vartheta}{n}}$
   $\Rightarrow$ Die Funktion ist maximal für minimales ϑ.
   Für $\omega = (x_1, \ldots, x_n)$ muss stets $\vartheta = N \geq x_n$ sein.
   Also ist $T_1 : \Omega \to \mathbb{R}$ $(x_1, \ldots, x_n) \mapsto x_n$ der Maximum-Likelihood-Schätzer.
   Allerdings ist T_1 kein sehr guter Schätzer, da er stets unterschätzt.

Definition 14.6 Sei $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, g)$ ein Schätzproblem mit $g : \theta \to \mathbb{R}$.

1. Ist $T : \Omega \to \mathbb{R}$ ein Schätzer für g und existiert $E_\vartheta(T) = \sum\limits_{\omega\in\Omega} T(\omega)P_\vartheta(\omega)$, so heißt $E_\vartheta(T)$ der ***Erwartungswert von T bei Parameter ϑ***.
2. Der Schätzer T heißt ***erwartungstreu***, wenn $E_\vartheta(T)$ für alle $\vartheta \in \theta$ existiert und $E_\vartheta(T) = g(\vartheta)$ erfüllt.
3. In der Situation von 2. heißt die Abbildung $b_T : \theta \to \mathbb{R}$, def. durch $\vartheta \mapsto E_\vartheta(T) - g(\vartheta)$, der ***Bias von T***.

   Der Schätzer ist also genau dann erwartungstreu, wenn $b_T = 0$ gilt.

Beispiel 14.7

1. In Beispiel 14.2 gilt für $T : \Omega \to \mathbb{R}$ $(\omega_1, \dots, \omega_n) \mapsto \frac{\omega_1+\dots+\omega_n}{n}$, dass

   $E_\vartheta(T) = \sum\limits_{\omega\in\Omega} T(\omega)P_\vartheta(\omega) = \sum\limits_{\omega\in\Omega} \frac{\omega_1+\dots+\omega_n}{n}\vartheta^{\omega_1+\dots+\omega_n}(1-\vartheta)^{n-\omega_1-\dots-\omega_n} =$
   $\sum\limits_{i=0}^{n} \binom{n}{i}\frac{i}{n}\vartheta^i(1-\vartheta)^{n-i} = \sum\limits_{i=1}^{n} \binom{n}{i}\frac{i}{n}\vartheta^i(1-\vartheta)^{n-i} =$
   $\sum\limits_{i=1}^{n} \binom{n-1}{i-1}\vartheta^i(1-\vartheta)^{n-i} = \sum\limits_{i=0}^{n} \binom{n-1}{i}\vartheta^{i+1}(1-\vartheta)^{n-1-i} =$
   $\vartheta\sum\limits_{i=0}^{n} \binom{n-1}{i}\vartheta^i(1-\vartheta)^{n-1-i} = \vartheta(\vartheta+(1-\vartheta))^{n-1} = \vartheta\cdot 1^{n-1} = \vartheta$ ist.

   Also gilt $E_\vartheta(T) = \vartheta = g(\vartheta)$ für alle $\vartheta \in \theta$, d.h. T ist erwartungstreu.
2. In Beispiel 14.3 gilt für $T_1 : \Omega \to \mathbb{R}$ $(x_1, \dots, x_n) \mapsto x_n$, dass $E_\vartheta(T_1) = \sum\limits_{1\le x_1<\dots<x_n\le N} x_n\frac{1}{\binom{N}{n}}$ ist.

   Wie man z.B. im Fall $N = n+1$ sofort sieht, ist T_1 nicht erwartungstreu.

   Es gilt $E_\vartheta(T_1) \neq N = g(\vartheta)$

   Für $T_2 : \Omega \to \mathbb{R}$ $(x_1, \dots, x_n) \mapsto x_n + x_1 - 1$ gilt

   $E_\vartheta(T_2) = \sum\limits_{1\le x_1<\dots<x_n\le N} (x_1+x_n-1)\frac{1}{\binom{N}{n}} =$
   $\sum\limits_{i=1}^{N-n+1} i\binom{N-i}{n-1}\frac{1}{\binom{N}{n}} + \sum\limits_{i=n}^{N} i\binom{i-1}{n-1}\frac{1}{\binom{N}{n}} - 1 =$
   $\frac{N+1}{n+1} + \frac{n(N+1)}{n+1} - 1 = N$

   $\Rightarrow T_2$ erwartungstreu

   Für $T_3 : \Omega \to \mathbb{R}$ $(x_1, \dots, x_n) \mapsto x_n + \frac{x_n-n}{n}$ gilt

   $E_\vartheta(T_3) = \sum\limits_{1\le x_1<\dots<x_n\le N} \left(\frac{n+1}{n} - 1\right)\frac{1}{\binom{N}{n}} = \frac{n+1}{n}\frac{n(N+1)}{n+1} - 1 = N$

   $\Rightarrow T_3$ erwartungstreu

Bemerkung 14.8 *Es braucht beim Schätzproblem keinen erwartungstreuen Schätzer zu geben.*

Z.B. sei $\Omega = \{0, \ldots, n\}$ *für ein* $n \in \mathbb{N}$,
$P_\vartheta = P_{n,\vartheta}$ =*Binomialverteilung zum Parameter* $\vartheta \in \theta = [0, 1]$.
Dann gibt es für die Abbildung $g : \theta = [0, 1] \to \mathbb{R}$, *def. durch* $\vartheta \mapsto \sqrt{n\vartheta(1 - \vartheta)}$ (*„Standardabweichung" von* $X : \Omega \to \mathbb{R}$, *def. durch* $i \mapsto P_\vartheta(i)$, *keinen erwartungstreuen Schätzer* $T : \Omega \to \mathbb{R}$, *denn* $E_\vartheta(T) = \sum_{i=0}^{n} T(i)\binom{n}{i}\vartheta(1 - \vartheta)^{n-i}$ *ist eine Polynomfunktion in* ϑ *und* $g : \theta \to \mathbb{R}$ *ist keine Polynomfunktion (z.B. gilt* $\frac{\partial}{\partial\vartheta}E_\vartheta(T)(0) < \infty$ *und* $\frac{\partial}{\partial\vartheta}g(0) = \infty$*)*

Beispiel 14.9 (Schätzung von Erwartungswerten und Varianzen)
In einem Wahrscheinlichkeitsexperiment (mit abzählbar vielen reellen möglichen Versuchsergebnissen) werden n unabhängige Wiederholungen durchgeführt. Es sind der Erwartungswert und die Varianz zu schätzen.
Sei $(\tilde{\Omega}, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und für $i = 1, \ldots, n$ beschreibe die Zufallsvariable $X_i : \tilde{\Omega} \to \mathbb{R}$ die i-te Wiederholung des Versuchs.
Sei $\Omega = X_1(\tilde{\Omega}) \times \ldots \times X_n(\tilde{\Omega})$.
Nach Voraussetzung gilt $P_{X_1} = \ldots = P_{X_n}$. Wir nehmen an, dass $\mu = E(X_1) = \ldots = E(X_n)$ und $\sigma^2 = \mathrm{Var}(X_1) = \ldots = \mathrm{Var}(X_n)$ existieren.
Für $\vartheta \in \theta$ sei $P_\vartheta : \mathfrak{P}(\Omega) \to [0, 1]$ die Wahrscheinlichkeitsverteilung $P_{X_1} \times \ldots \times P_{X_n}$.

1. Für den Erwartungswert $g_1 : \theta = \mathbb{R} \to \mathbb{R}$, def. durch $\mu \mapsto \mu$ ist der Mittelwert $T_1 = \frac{1}{n}(X_1 + \ldots + X_n) : \Omega \to \mathbb{R}$ $(X_1(\tilde{\omega_1}), \ldots, X_n(\tilde{\omega_n})) \mapsto \frac{1}{n}(X_1(\tilde{\omega_1}) + \ldots + X_n(\tilde{\omega_n}))$ ein erwartungstreuer Schätzer, denn $E_\mu(T_1) = \frac{1}{n}(E(X_1) + \ldots + E(X_n)) = \frac{1}{n}n\mu = \mu = g(\mu)$
2. Für die Varianz $g_2 : \theta = \mathbb{R} \to \mathbb{R}$, def. durch $\vartheta = \sigma^2 \mapsto \sigma^2$, ist $T_2 : \Omega \to \mathbb{R}$, def. durch $\omega = (X_1(\tilde{\omega_1}), \ldots, X_n(\tilde{\omega_n})) \mapsto \frac{1}{n-1}\sum_{i=1}^{n}(X_i(\tilde{\omega_i}) - T_1(\omega))^2$, ein erwartungstreuer Schätzer (nachrechnen!)

Definition 14.10 Sei $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, g)$ ein Schätzproblem mit $g : \theta \to \mathbb{R}$ und $T : \Omega \to \mathbb{R}$ ein Schätzer.

1. Existiert $\mathrm{Var}_\vartheta(T) := E_\vartheta((T - E_\vartheta(T))^2)$, so heißt $\mathrm{Var}_\vartheta(T)$ die ***Varianz des Schätzers T***.
2. Existiert $E_\vartheta((T - g(\vartheta))^2)$ für jedes $\vartheta \in \theta$, so heißt $R_T : \theta \to \mathbb{R}$, def. durch $\vartheta \mapsto E_\vartheta((T - g(\vartheta))^2)$ der ***mittlere quadratische Fehler des Schätzers T***.

Satz 14.11 *Sei $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, g)$ ein Schätzproblem mit $g : \theta \to \mathbb{R}$ und sei $T : \Omega \to \mathbb{R}$ ein Schätzer für den $\mathrm{Var}_\vartheta(T)$ und R_T existieren (für alle $\vartheta \in \theta$).*
Dann gilt $R_T(\vartheta) = \mathrm{Var}_\vartheta(T) + b_T(\vartheta)^2$.

Beweis:
Es gilt $R_T(\vartheta) = E_\vartheta(((T - E_\vartheta(T)) - g(\vartheta) - E_\vartheta(T))^2) =$
$E_\vartheta((T - E_\vartheta(T))^2) - 2(g(\vartheta) - E_\vartheta(T))\underbrace{E_\vartheta(T - E_\vartheta(T))}_{=0} + (g(\vartheta) - E_\vartheta(T))^2 =$
$\mathrm{Var}_\vartheta(T) + b_T(\vartheta)^2$ □

Beispiel 14.12

1. In Beispiel 14.2 war $\Omega = \{0,1\}^n$, $\theta = [0,1]$, $P_\vartheta((\omega_1, \dots, \omega_n)) = \vartheta^{\omega_1+\dots+\omega_n}(1-\vartheta)^{n-\omega_1-\dots-\omega_n}$, $g : \theta \to \mathbb{R}\ \vartheta \mapsto \vartheta$, $T : \Omega \to \mathbb{R}\ (\omega_1, \dots, \omega_n) \mapsto \frac{1}{n}(\omega_1 + \dots + \omega_n)$ Schätzer.
   $\Rightarrow \mathrm{Var}_\vartheta(T) = E_\vartheta((T - E_\vartheta(T))^2) = E_\vartheta((T - \vartheta)^2) =$
   $\sum\limits_{\omega\in\Omega} (T(\omega) - \vartheta)^2 P_\vartheta(\omega) = \sum\limits_{i=0}^{n} \binom{n}{i} \left(\frac{i}{n} - \vartheta\right)^2 \vartheta^i (1-\vartheta)^{n-i} =$
   $\frac{1}{n^2}\mathrm{Var}(b_{n,\vartheta}) = \frac{1}{n}\vartheta(1-\vartheta)$ und
   $R_T(\vartheta) = \mathrm{Var}_\vartheta(T) + b_T(\vartheta)^2 = \mathrm{Var}_\vartheta(T) = \frac{1}{n}\vartheta(1-\vartheta)$
   Sei $\tilde{T} : \Omega \to \mathbb{R}$ ein beliebiger erwartungstreuer Schätzer für $g : \theta = [0,1] \to \mathbb{R}$, d.h. es gelte $E_\vartheta(\tilde{T}) = \vartheta$ für alle $\vartheta \in [0,1]$.
   Dann kann man zeigen, dass $\mathrm{Var}_\vartheta(\tilde{T}) \geq \frac{1}{n}\vartheta(1-\vartheta) = \mathrm{Var}_\vartheta(T)$ gilt (vgl. Krengel §4.5)
   $\Rightarrow \forall \vartheta \in \theta : R_{\tilde{T}}(\vartheta) \geq R_T(\vartheta)$
   $\Rightarrow T$ ist optimal unter den erwartungstreuen Scbätzern.
2. In Beispiel 14.3 waren $T_2 : \Omega \to \mathbb{R}\ (x_1, \dots, x_n) \mapsto x_n + x_1 - 1$ und $T_3 : \Omega \to \mathbb{R}\ (x_1, \dots, x_n) \mapsto x_n + \frac{x_n - n}{n}$ erwartungstreu.
   Es gilt $\mathrm{Var}_\vartheta(T_2) = \sum\limits_{1\leq x_1<\dots<x_n\leq N} (x_1 + x_n - 1 - N)^2 \frac{1}{\binom{N}{n}}$ und
   $\mathrm{Var}_\vartheta(T_3) = \sum\limits_{1\leq x_1<\dots<x_n\leq N} (x_1 + \frac{x_n - n}{n} - N)^2 \frac{1}{\binom{N}{n}}$, sowie $R_{T_2}(\vartheta) = \mathrm{Var}_\vartheta(T_2) \geq \mathrm{Var}_\vartheta(T_3) = R_{T_3}(\vartheta)$ für alle $\vartheta \in \theta$.
   Also hat T_3 einen kleineren mittleren quadratischen Fehler als T_2.

Definition 14.13 Sei M eine Menge und sei $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, g)$ ein Schätzproblem mit $g : \theta \to M$ und sei $\alpha \in \mathbb{R}_+$.

1. Ist für jedes $\omega \in \Omega$ eine Menge $C(\omega) \subset M$ gegeben mit $P_\vartheta(\{\omega \in \Omega \mid g(\vartheta) \in$

$C(\omega)\}) \geq 1-\alpha$, so heißt $C(\omega)$ ein ***Konfidenzbereich zum Konfidenzniveau*** $\boldsymbol{\alpha}$.

2. Ist $g : \theta \to \mathbb{R}$ und $g(\theta) \subset \mathbb{R}$ ein Intervall und $C(\omega) \subset g(\theta)$ ein Teilintervall, so heißt $C(\omega)$ ein ***Konfidenzintervall zum Niveau*** $\boldsymbol{\alpha}$.

Bemerkung 14.14 *(Bedeutung der Konfidenzintervalle)*
Sei $\{C(\omega)\}_{\omega\in\Omega}$ eine Familie von Konfidenzintervallen für $g : \theta \to \mathbb{R}$ zum Niveau $\alpha \in \mathbb{R}_+$.

1. *Für jedes $\omega \in \Omega$ und $g(\vartheta) \in g(\theta)$ gilt entweder $g(\vartheta) \in C(\omega)$ oder nicht.*
   Für die Menge $\{\vartheta \in \theta \mid g(\vartheta) \in C(\omega)\}$ gibt es keine „Wahrscheinlichkeit“.
   Es gibt also keine Wahrscheinlichkeit dafür, dass $C(\omega)$ den Wert $g(\vartheta)$ enthält.
2. *Vielmehr geben wir eine Familie von Konfidenzintervallen an und behaupten mit Wahrscheinlichkeit$\geq 1-\alpha$ liefert das aus der Beobachtung von $\omega \in \Omega$ berechnete Intervall $C(\omega)$ ein Resultat mit $g(\vartheta) \in C(\omega)$.*

Bemerkung 14.15 *(Konstruktion von Konfidenzintervallen)*
Sei $g : \theta \to \mathbb{R}$ der zu schätzende Parameter.
Die Angabe einer Familie $\{C(\omega)\}_{\omega\in\Omega}$ ist gleichbedeutend mit der Angabe der Menge $C = \{(\omega, \vartheta) \in \Omega \times \theta \mid \vartheta \in C(\omega)\}$ und daher mit der Angabe der Familie $\{A(\vartheta)\}_{\vartheta\in\theta}$ mit $A(\vartheta) = \{\omega \in \Omega \mid (\omega, \vartheta) \in C\} = \{\omega \in \Omega \mid \vartheta \in C(\omega)\}$
Offenbar gilt $P_\vartheta(\{\omega \in \Omega \mid \vartheta \in C(\omega)\}) \geq 1-\alpha \Leftrightarrow P_\vartheta(A(\vartheta)) \geq 1-\alpha$.
Also genügt es, für jedes $\vartheta \in \theta$ eine Menge $A(\vartheta) \subset \Omega$ anzugeben mit $P_\vartheta(A(\vartheta)) \geq 1-\alpha$.
Damit die Konfidenzintervalle $C(\omega)$ möglichst klein sind, ist $A(\vartheta)$ möglichst „klein“ zu wählen, z.B. indem wir in $A(\vartheta)$ die Ereignisse $\omega \in \Omega$ wählen, für die $P_\vartheta(\omega)$ maximal ist.
Verfahren:

1. *Für jedes $\vartheta \in \theta$ sortiere $\Omega = \{\omega_1, \omega_2, \ldots\}$, so dass $P_\vartheta(\omega_1) \geq P_\vartheta(\omega_2) \geq \ldots$ gilt.*
2. *Wähle $A(\vartheta) = \{\omega_1, \ldots, \omega_k \mid P_\vartheta(\omega_1) + \ldots + P_\vartheta(\omega_k) \geq 1-\alpha$ mit k minimal $\}$.*
3. *Bestimme die $C(\omega)$ aus den $A(\vartheta)$.*

Definition 14.16

1. Sei θ eine Menge und sei $(\Omega, \mathcal{A}_\vartheta, P_\vartheta)_{\vartheta\in\theta}$ eine Familie von Wahrscheinlichkeitsräumen. Ferner sei $(\Omega', \mathcal{B})$ ein messbarer Raum und eine Abbildung $X : \Omega \to \Omega'$ sei bzgl. allen $(\mathcal{A}_\vartheta, \mathcal{B})$ messbar.
   Dann induziert X für jedes $\vartheta \in \theta$ eine Wahrscheinlichkeitsverteilung P_ϑ von $(\Omega', \mathcal{B})$ (d.h. für $B \in \mathcal{B}$ sei $P_\vartheta(B) = \tilde{P}_\vartheta(X^{-1}(B))$).
   Schließlich sei M eine Menge und $g : \theta \to M$ eine Abbildung. Dann heißt $(\Omega', X, \{P_\vartheta\}_{\vartheta\in\theta}, g)$ ein ***(allgemeines) Schätzproblem*** und jede Abbildung $T : X(\Omega) \to M$ heißt ein ***Schätzer***.

Beispiel 14.17 Die Zufallsvariablen $X_{(\mu,\sigma^2)} : \Omega \to \mathbb{R}$ seien $N(\mu, \sigma^2)$-verteilt mit $\vartheta = (\mu, \sigma^2) \in \theta = \mathbb{R} \times \mathbb{R}_+$.

1. Gesucht sei ein Schätzer für $g : \theta \to \mathbb{R}$ $(\mu, \sigma^2) \mapsto \mu$.
   Dazu benötigen wir für jedes $\vartheta = (\mu, \sigma^2) \in \theta$ das Maximum der Funktion $\mathbb{R} \to [0, \infty[$ $x \mapsto f_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
   Dieses liegt bei $\mu = x$, d.h. der Maximum-Likelihood-Schätzer ist $L : \Omega \to \mathbb{R}$ $\omega \mapsto x = X_{(\mu,\sigma^2)}(\omega)$.
   Der Schätzer L ist erwartungstreu, d.h. $E_\vartheta(L) = E(X_{(\mu,\sigma^2)}) = \mu = g(\vartheta)$.
   Die Zufallsvariable $Y_\mu : \Omega \to \mathbb{R}$ $\omega \mapsto \frac{X_{(\mu,\sigma^2)} - \mu}{\sigma}$, also $Y_\mu = (X^*_{(\mu,\sigma^2)})$, ist $N(0, 1)$ verteilt.
   Sei z.B. $\alpha = 0,005$ vorgegeben.
   Wegen $\Phi(1,96) = 0,975$ gilt dann
   $P_{(\mu,\sigma^2)}(|Y_\mu| \leq 1,96) = \Phi(1,96) - \Phi(-1,96) = 0,95$.
   Sei für jedes $\omega \in \Omega$ der Konfidenzbereich $C(\omega)$ definiert durch $C(\omega) = \{\mu \in \mathbb{R} \mid |X(\omega) - \mu| \leq 1,96\sigma\}$.
   Dann gilt $\mu \in C(\omega) \Leftrightarrow |Y_\mu(\omega)| \leq 1,96$, also $P_{(\mu,\sigma^2)}(\{\mu \in \mathbb{R} \mid \mu \in C(\omega)\}) = P_{(\mu,\sigma^2)}(|Y_\mu| \leq 1,96) = 0,95$,
   d.h. $C(\omega)$ ist ein Konfidenzintervall zum Niveau 0,05.
2. Der Maximum-Likelihood-Schätzer für $g : \theta \to \mathbb{R}$ $(\mu, \sigma^2) \mapsto \sigma^2$ ist analog gegeben durch $L : \Omega \to \mathbb{R}$ $\omega \mapsto (X_{(\mu,\sigma^2)}(\omega) - \mu)^2$.
   Auch er ist erwartungstreu.

Beispiel 14.18 (Die Methode der kleinsten Quadrate)
Aufgabe: Ausgehend von einer Größe $x \in \mathbb{R}$ wird eine Größe $y \in \mathbb{R}$ gemessen und zwar n-mal unabhängig voneinander.

Man erhält n Messpunkte $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^2$.
Es soll eine bestimmte Kurve (z.B. Gerade, Parabel, Exponentialkurve ...) so bestimmt werden, dass sie möglichst nahe an den Punkten vorbei geht.
Mathematischer Ansatz:
Seien $\vartheta_1, \ldots, \vartheta_p \in \mathbb{R}$ Parameter und $\varphi_{(\vartheta_1,\ldots,\vartheta_p)} : \mathbb{R} \to \mathbb{R}$ die zu bestimmende Funktion mit $y_i \approx \varphi_{(\vartheta_1,\ldots,\vartheta_p)}(x)$.
Die Methode der kleinsten Quadrate besagt, dass man $(\vartheta_1, \ldots, \vartheta_p) \in \mathbb{R}^p$ so wählen soll, dass $Q(\vartheta_1, \ldots, \vartheta_p) = \sum\limits_{i=1}^{n}(g_i - \varphi_{(\vartheta_1,\ldots,\vartheta_p)}(x_i))^2$ minimal ist.
Mathematische Begründung:
Für $i = 1, \ldots, n$ sei $\xi_i = \varphi_{(\vartheta_1,\ldots,\vartheta_p)}(x_i)$ der „wirkliche" zu messende Wert und $z_i = y_i - \xi_i$ der „Messfehler".
Sei $\Omega = \mathbb{R}$ und für $i = 1, \ldots, n$ sei $Y_i : \Omega \to \mathbb{R}\ t \mapsto y_i(x_i, t)$ das Ergebnis der Messung von y_i (abhängig von x_i) zur Zeit t.
Ferner sei $Y = (Y_1, \ldots, Y_n) : \Omega \to \mathbb{R}^n\ x \mapsto (y_1(t), \ldots, y_n(t))$.
Die Verteilungen von $Y_1, \ldots, Y_n$ sind unbekannt.
Angenommen die Zufallsvariablen $Y_1, \ldots, Y_n$ sind unabhängig und Y_i ist $N(\xi, \sigma^2)$-verteilt für $i = 1, \ldots, n$.
$\Rightarrow$ nach Satz 12.17 ist die Dichte von Y gegeben durch
$f : \mathbb{R}^n \to [0, \infty[\ (y_1, \ldots, y_n) \mapsto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \mathrm{e}^{-\frac{1}{2\sigma^2}\left(\sum\limits_{i=1}^{n}(y_i - \xi_i)^2\right)}$.
Also ist $f(y_1, \ldots, y_n)$ maximal genau dann, wenn
$Q(\vartheta_1, \ldots, \vartheta_p) = \sum\limits_{i=1}^{n}\left(y_i - \varphi_{(\vartheta_1,\ldots,\vartheta_p)}(x_i)\right)^2 = \sum\limits_{i=1}^{n}(y_i - \xi_i)^2$ minimal ist.
$\Rightarrow$ der Maximum-Likelihood-Schätzer für $g : \theta \to \mathbb{R}^p\ (\vartheta_1, \ldots, \vartheta_p) \mapsto (\vartheta_1, \ldots, \vartheta_p)$ ist $T : \mathbb{R}^n \to \mathbb{R}^p\ (y_1, \ldots, y_n) \mapsto (\vartheta_1, \ldots, \vartheta_p)$ mit $Q(\vartheta_1, \ldots, \vartheta_p)$ minimal.

Beispiel 14.19 (Die Regressionsgerade)
Den Messpunkten $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^2$ soll eine Gerade $\varphi_{(\alpha,\beta)} : \mathbb{R} \to \mathbb{R}$ $x \mapsto \alpha + \beta x$ möglichst gut angepasst werden.
Es ist also $Q(\alpha, \beta) = \sum\limits_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2$ zu minimieren.
$\Rightarrow \frac{\partial Q}{\partial \alpha} = 0$ liefert $\sum\limits_{i=1}^{n}(y_i - \alpha - \beta x_i) = 0$ und $\frac{\partial Q}{\partial \beta} = 0$ liefert $\sum\limits_{i=1}^{n} x_i(y_i - \alpha - \beta x_i) = 0$.
Definiere $\overline{x} = \frac{1}{n}(x_1 + \ldots + x_n)$, $\overline{y} = \frac{1}{n}(y_1 + \ldots + y_n)$, $\overline{xy} = \frac{1}{n}(x_1 y_1 + \ldots + x_n y_n)$, $\overline{x^2} = \frac{1}{n}(x_1^2 + \ldots + x_n^2)$
Die Bedingungen lauten:

$\overline{y} = \alpha + \beta\overline{x}$, $\overline{xy} - \alpha\overline{x} - \beta\overline{x^2} = 0$
$\Rightarrow$ der Maximum-Likelihood-Schätzer für $g : \mathbb{R}^2 \to \mathbb{R}^2$ $(\alpha, \beta) \mapsto (\alpha, \beta)$ lautet $T : \mathbb{R}^n \to \mathbb{R}^2$ $(y_1, \ldots, y_n) \mapsto (\hat{\alpha}, \hat{\beta}) = (\overline{y} - \gamma\overline{x}, \gamma)$ mit $\gamma = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}\,\overline{x}}$
Die Gerade $y = \hat{\alpha} + \hat{\beta}x$ heißt ***Regressionsgerade***. Die Zahl γ heißt der ***Regressionskoeffizient***.
Der Schätzer T ist erwartungstreu.

15 Testtheorie

Beispiel 15.1 Ein Wünschelrutengänger behauptet, er könne Wasseradern orten (mit „an Sicherheit grenzender Wahrscheinlichkeit").
Versuchsaufbau:
Eine Wasserleitung wird in einer Wiese verlegt.
Darüber wird ein Zelt aufgebaut. Der Wünschelrutengänger schreitet n-mal über die Wasserleitung, wobei er nicht weiß, ob gerade Wasser fließt. (Letzteres wird zufällig bestimmt). Er gibt das Ergebnis seiner „Ortung" bekannt. Kann man seine Behauptung widerlegen ?
Mathematisches Modell:
$\Omega = \{0, 1\}^n$ =Menge der Versuchsergebnisse mit „an der i-ten Stelle" entspricht „i-te Ortung war korrekt".
Sei p die wirkliche Wahrscheinlichkeit, mit der der Wünschelrutengänger korrekt ortet.
„Nullhypothese" $p = \frac{1}{2}$
„Alternative" $p > \frac{1}{2}$
Für $\vartheta \in \left[\frac{1}{2}, 1\right] = \theta$ sei $\tilde{P}_\vartheta$ die n-te Bernoulli-Verteilung zum Parameter θ und $X : \Omega \to \mathbb{R}$ $(\omega_1, \ldots, \omega_n) \mapsto \omega_1 + \ldots + \omega_n$
$\Rightarrow P_{X,\vartheta} = b_{n,\theta}$ Binomialverteilung
Es soll nun zu einem gewissen ***Signifikanzniveau*** $\alpha \in \mathbb{R}_+$ ein ***kritischer Wert*** $t \in \mathbb{N}$ so festgelegt werden, dass

1. unter der Nullhypothese gilt $P(X \geq t) \leq \alpha$.
2. im Fall $X(\omega) > t$ wir die Behauptung des Wünschelrutengängers akzeptiert, d.h. die Nullhypothese wird verworfen.
3. Im Fall $X(\omega) < t$ wir die Behauptung des Wünschelrutengängers als nicht

erwiesen betrachtet, d.h. die Nullhypothese wird akzeptiert. (Sie ist nicht „bewiesen“).

Definition 15.2

1. Ein ***Testproblem*** $(\Omega, \{\tilde{P}_\vartheta\}_{\vartheta\in\theta}, X, H)$ besteht aus folgenden Daten:
   1. einer abzählbaren Menge Ω
   2. einer Menge θ
   3. einer Familie von Wahrscheinlichkeitsverteilungen $\{\tilde{P}_\vartheta\}_{\vartheta\in\theta}$ auf Ω, $\tilde{P}_\vartheta : \mathcal{A}_\vartheta \to [0,1]$
   4. einer Zufallsvariablen $X : \Omega \to M$ mit Werten in einem messbaren Raum (M, M)
   5. einer echten nichtleeren Teilmenge $\emptyset \neq H \subsetneq \theta$ (genannt Hypothese oder Nullhypothese)

   Im Folgenden sei $(\Omega, \{\tilde{P}_\vartheta\}_{\vartheta\in\theta}, X, H)$ ein Testproblem.
2. Für jedes $\vartheta \in \theta$ betrachten wir die Wahrscheinlichkeitsverteilung $P_\vartheta : \mathrm{M} \to [0,1]$ $N \mapsto \tilde{P}_\vartheta(X^{-1}(N))$.
   Die Menge $K = \theta \setminus H$ heißt die ***Alternative***.
3. Jede Abbildung $T : X(\Omega) \to \{H, K\}$ heißt ein ***Test*** (oder ein ***Testverfahren***). Gilt $T(X(\omega)) = H$ für ein $\omega \in \Omega$, so nennen wir dies die ***Annahme der Hypothese***.
   Gilt $T(X(\omega)) = K$ für ein $\omega \in \Omega$, so nennen wir dies das ***Verwerfen der Hypothese***.
4. Ist $T : X(\Omega) \to \{H, K\}$ ein Test, so heißt $R = T^{-1}(K)$ der ***Verwerfungsbereich*** (oder der ***kritische Bereich***).
   Die Angabe von R ist äquivalent zur Angabe von T.
5. Ist $X(\Omega) \subset \mathbb{R}$ und gibt es eine Funktion $\tilde{T} : \mathbb{R} \to \mathbb{R}$ mit $R = \{x \in \mathbb{R} \mid \tilde{T}(x) \geq t\}$ für ein $t \in \mathbb{R}$, so heißt $\tilde{T}$ eine ***Teststatistik für T*** und t ein ***kritischer Wert***.
6. Die Abbildung $\beta : \theta \to [0,1]$ $\vartheta \mapsto P_\vartheta(X \in R)$ heißt der ***Gütefaktor von T***.

Beispiel 15.3 In der Situation von 15.1 gilt:
$\Omega = \{0,1\}^n$, $\theta = \left[\frac{1}{2}, 1\right]$, $\tilde{P}_\vartheta = n$-te Bernoulli-Verteilung zum Parameter $\vartheta \in \theta$.
$X : \Omega \to \mathbb{R}$ $(\omega_1, \ldots, \omega_n) \mapsto \omega_1 + \ldots + \omega_n$, $X(\Omega) = \{0, \ldots, n\}$,

$P_\vartheta = B_{n,\vartheta}$ =Binomialverteilung
$H = \{\frac{1}{2}\}$, $K =]\frac{1}{2}, 1]$, $T : X(\Omega) \to \{H, K\}$ sei gegeben durch $R = T^{-1}(K) = \{X \geq t\}$ für ein $t \in \mathbb{R}$ (mit Teststatistik $\tilde{T} : \mathbb{R} \to \mathbb{R}\ x \mapsto x$), d.h. die Behauptung des Wünschelrutengängers wird nur dann akzeptiert, wenn er mindestens t korrekte Ortungen vornimmt.
Die Gütefunktion $\beta : [\frac{1}{2}, 1] \to [0, 1]$ erfüllt $\beta(\vartheta) = P_\vartheta(X \geq t) = b_{n,\vartheta}(t) + b_{n,\vartheta}(t + 1) + \ldots + b_{n,\vartheta}(n)$.
Wünscht man im Fall $n = 5$ etwa $\beta(\vartheta) \leq 0,05$ für alle $\vartheta \in \theta$, so muss man $t = 5$ wählen, denn für $t = 4$ ist $P_\vartheta(X \geq 4) \approx 0,187 > \alpha$.

Definition 15.4 Sei $(\Omega, \{\tilde{P}_\vartheta\}_{\vartheta\in\theta}, X, H)$ ein Testproblem, d.h. θ Menge, $(\Omega, \mathcal{A}_\vartheta, \tilde{P}_\vartheta)_{\vartheta\in\theta}$ Familie von Wahrscheinlichkeitsräumen, $X : \Omega \to \Omega'$ Zufallsvariable, $P_\vartheta = \tilde{P}_\vartheta \circ X^{-1}$ für $\vartheta \in \theta$, $H \subset \theta$ Hypothese, $K = \theta \setminus H$ Alternative.

1. Ein ***randomisierter Test*** T ist eine messbare Abbildung $T : X(\Omega) \to [0, 1]$. Wird dabei $x \in X(\Omega)$ beobachtet, so sei $T(x)$ die Wahrscheinlichkeit, mit der H verworfen wird.
2. Ist $\vartheta \in H$ und wird H verworfen, so spricht man von einem ***Fehler 1. Art***.
3. Ist $\vartheta \in K$ und wird H akzeptiert, so spricht man von einem ***Fehler 2. Art***.
4. Ist $\alpha \in \mathbb{R}_+$ mit $\beta(\vartheta) \leq \alpha$ für alle $\vartheta \in H$, so sagt man, der Test habe das ***Signifikanzniveau*** α.

   Die Wahrscheinlichkeit eines Fehlers 1. Art ist dann maximal α.
5. Ist $\alpha \in \mathbb{R}_+$, so sagt man, der Test habe die Macht $1 - \alpha$, wenn für alle $\vartheta \in K$ gilt $\beta(\vartheta) \geq 1 - \alpha$.

Bemerkung 15.5 *Im Folgenden sei $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, X, H)$ ein Testproblem.*

1. *Ist $T : X(\Omega) \to \{H, K\}$ ein Test im Sinne von Def. 15.2, so ist $T : X(\Omega) \to \{H, K\} \to [0, 1]\ H \mapsto 0,\ K \mapsto 1$.*

   In diesem Sinne werden wir ab jetzt nur noch randomisierte Tests betrachten.
2. *Ist $T : X(\Omega) \to [0, 1]$ ein randomisierter Test, so ist die Gütefunktion $\beta : \theta \to [0, 1]$ gegeben durch $\vartheta \mapsto E_\vartheta(T) = \sum\limits_{x\in X(\Omega)} T(x)P_\vartheta(x)$, denn mit Wahrscheinlichkeit $P_\vartheta(x)$ wird x beobachtet und dann wird H mit Wahrscheinlichkeit $T(x)$ verworfen.*
3. *Hat ein (gewöhnlicher) Test $T : X(\Omega) \to [0, 1]$ das Signifikanzniveau $\alpha \in \mathbb{R}_+$,*

so ist die Wahrscheinlichkeit eines Fehlers 1. Art kleiner oder gleich α, denn Für $\vartheta \in H$ gilt $P_\vartheta(\{x \in X(\Omega) \mid T(x) = 1\}) = \beta(\vartheta) \leq \alpha$.

4. *Hat ein (gewöhnlicher) Test $T : X(\Omega) \to [0,1]$ die Macht $1-\alpha$ für ein $\alpha \in \mathbb{R}_+$, so ist die Wahrscheinlichkeit eines Fehlers 2. Art kleiner oder gleich α, denn für $\vartheta \in K$ gilt $P_\vartheta(\{x \in X(\Omega) \mid T(x) = 0\}) = 1 - P_\vartheta(\{x \in X(\Omega) \mid T(x) = 1\}) = 1 - \beta(\vartheta) \leq \alpha$*

Beispiel 15.6 (Fortsetzung von 15.1 „Wünschelrutengänger“)
Sei $\Omega = \{0,1\}^n$, $\theta = [\frac{1}{2}, 1]$, $\tilde{P}_\vartheta = n$-te Bernoulli-Verteilung zum Parameter ϑ, $X : \Omega \to \mathbb{R}$ $(\omega_1, \ldots, \omega_n) \mapsto \omega_1 + \ldots + \omega_n$ mit $X(\Omega) = \{0, \ldots, n\}$ und $P_\vartheta = b_{n,\vartheta}$, $H = \{\frac{1}{2}\}$, $K =]\frac{1}{2}, 1]$.
Sei $T : X(\Omega) \to [0,1]$ mit $\operatorname{im} T \subset \{0,1\}$ und $R = T^{-1}(1) = \{X \geq t\}$ für ein $t \in \mathbb{R}$, def. durch $x \mapsto \begin{cases} 1 & \text{falls } x \geq t \\ 0 & \text{sonst} \end{cases}$.
Im Folgenden sei $n = 5$ und $t = 5$ (notwendig für Niveau $\alpha = 0,05$)
Die Gütefunktion ist $\beta : \theta \to [0,1]$ $\vartheta \mapsto P_\vartheta(X \geq t)$.

1. $\beta(0,6) = 0,08$ d.h. bei Trefferquote 0,6 wird die Hypothese nur mit Wahrscheinlichkeit 0,08 verworfen.
2. $\beta(0,9) = 0,59$ d.h. …

Man kann so herausfinden, wie groß n sein muss, damit eine Abweichung von der Nullhypothese mit großer Wahrscheinlichkeit festgestellt wird.
Beim Fehler 2. Art kann man in diesem Beispiel nicht für alle $\vartheta \in K$ erreichen, dass $1 - \beta(\vartheta)$ klein ist, denn $\lim\limits_{\vartheta \overset{>}{\to} \frac{1}{2}} (1 - \beta(\vartheta)) = 1 - \beta(\frac{1}{2}) \geq 1 - \alpha$ für kleine α unmöglich.
Man kannn aber z.B. erreichen, dass für alle $\vartheta \geq 0,6$ ein Fehler 2. Art mit Wahrscheinlichkeit$\leq 0,05$ eintritt.

Definition 15.7

1. Die Hypothese $H \subset \theta$ heißt ***einfach***, wenn es ein $\vartheta_H \in \theta$ gibt mit $H = \{\vartheta_H\}$.
   In diesem Fall schreiben wir $P_H = P_{\vartheta_H}$.
   Andernfalls heißt H ***zusammengesetzt***.
   Analog heißt die Alternative $K = \theta \setminus H$ ***einfach***, wenn es ein $\vartheta_K \in \theta$ gibt mit $K = \{\vartheta_K\}$.
   Wieder schreiben wir $P_K = P_{\vartheta_K}$.

2. Sind $H, K \subset \theta$ einfach, so heißt ein Test $T : X(\Omega) \to [0,1]$ ein ***Neyman-Pearson-Test***, wenn es ein $c \in [0,\infty]$ gibt mit $T : X(\Omega) \to [0,1]$ $x \mapsto \begin{cases} 1 & \text{falls} \quad P_K(x) > cP_H(x) \\ 0 & \text{falls} \quad P_K(x) < cP_H(x) \end{cases}$
(Für $x \in X(\Omega)$ mit $P_K(x) = cP_H(x)$ kann man $T(x)$ beliebig fortsetzen).

Bemerkung 15.8 *Sind $H, K \subset \theta$ einfach und ist $T : X(\Omega) \to [0,1]$ ein Test, so ist $E_H(T) = \sum\limits_{x \in X(\Omega)} T(x)P_H(x)$ das minimale Niveau von T und $E_K(T) = \sum\limits_{x \in X(\Omega)} T(x)P_K(x)$ die minimale Macht von T.*
*Dabei heißt T **schärfer** als ein Test $T' : X(\Omega) \to [0,1]$, wenn $E_K(T) > E_K(T')$*

Satz 15.9 *(Der Satz von Neyman-Pearson)*
Sei $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, X, H)$ ein Testproblem mit $H, K \subset \theta$ einfach (und mit Ω abzählbar).

1. *Ist $T : X(\Omega) \to [0,1]$ ein Neyman-Pearson-Test, so ist T mindestens so scharf wie alle anderen Tests $T' : X(\Omega) \to [0,1]$ mit $E_H(T') = E_H(T)$.*
2. *Zu jedem $\alpha \in [0,1]$ gibt es einen Neyman-Pearson-Test $T : X(\Omega) \to [0,1]$ mit $E_H(T) = \alpha$.*

Beweis:

1. Sei $A = \{x \in X(\Omega) \mid T(x) > T'(x)\}$.
   Für $x \in A$ gilt $T(x) > 0$, also $P_K(x) \geq cP_H(x)$.
   Sei $B = \{x \in X(\Omega) \mid T(x) < T'(x)\}$.
   Für $x \in B$ gilt $T(x) < 1$, also $P_K(x) \leq cP_H(x)$.
   $\Rightarrow E_K(T) - E_K(T') = \sum\limits_{x\in X(\Omega)} (T(x) - T'(x))P_K(x) \geq$
   $\sum\limits_{x\in A} (T(x) - T'(x))cP_H(x) + \sum\limits_{x\in B} (T(x) - T'(x))cP_H(x) =$
   $c \sum\limits_{x\in X(\Omega)} (T(x) - T'(x))P_H(x) = c(E_H(T) - E_H(T')) = 0$
   $\Rightarrow E_K(T) \geq E_K(T')$
2. 1. Fall: $\alpha = 0 \Rightarrow c = \infty$ liefert $E_H(T) = 0$
   2. Fall: $\alpha > 0$:
   Betrachte $f : [0,\infty[\to \mathbb{R}$ $c \mapsto P_H(\{x \in X(\Omega) \mid P_K(x) > cP_H(x)\})$ mit $q(x) = \frac{P_K(x)}{P_H(x)}$ und $g : [0,\infty[\to \mathbb{R}$ $c \mapsto P_H(\{x \in X(\Omega) \mid q(x) \geq c\})$
   Es gilt:
   1. $g(0) = 1$
   2. f monoton fallend, $\lim\limits_{c\to\infty} f(c) = 0$

3. f rechtsstetig
4. $\lim\limits_{\tilde{c} \overset{<}{\to} c} f(\tilde{c}) = g(c)$

Dann sei $c^* = \inf\{c \in \mathbb{R} \mid f(c) \leq \alpha\}$
$\Rightarrow g(c^*) \geq \alpha \geq f(c^*)$
Im Fall $g(c^*) = f(c^*)$ sei $\gamma := 0$.
Im Fall $g(c^*) - f(c^*)$ sei $\gamma := \frac{\alpha - f(c^*)}{g(c^*) - f(c^*)}$.
Damit definieren wir $T : X(\Omega) \to \mathbb{R}$ durch

$$x \mapsto \begin{cases} 1 & \text{falls} \quad P_K(x) > c^* P_H(x) \\ 0 & \text{falls} \quad P_K(x) < c^* P_H(x) \\ \gamma & \text{sonst} \end{cases}$$

Es folgt $E_H(T) = \sum\limits_{x \in X(\Omega)} T(x) P_H(x) =$
$P_H(\{x \in X(\Omega) \mid q(x) > c^*\}) + \gamma P_H(\{x \in X(\Omega) \mid q(x) = c^*\}) =$
$f(c^*) + \gamma(g(c^*) - f(c^*)) = \alpha$

□

Beispiel 15.10 Sei wieder $\Omega = \{0,1\}^n$, $X : \Omega \to \mathbb{R}$ $(\omega_1, \ldots, \omega_n) \mapsto \omega_1 + \ldots + \omega_n$ mit $X(\Omega) = \{0, \ldots, n\}$, $P_\vartheta = b_{n,\vartheta}$ $\vartheta \in \theta = \{\frac{1}{2}, 1\}$.
Zuerst behandeln wir den Fall $H = \{\frac{1}{2}\} = \{p_H\}$, $K = \{p_k\}$ mit $\frac{1}{2} < p_K \leq 1$ $\theta' = \{H, K\}$.
Wie im Beweis von 15.9 2. betrachte $q(x) = \frac{P_K(x)}{P_H(x)} = \frac{\binom{n}{x} p_K^x (1-p_K)^{n-x}}{\binom{n}{x} p_H^x (1-p_H)^{n-x}} = (2p_K)^x (2 - 2p_K)^{n-x} = (2p_K)^x (2 - 2p_k)^{n-x}$
Die Funktion $q(x)$ ist wegen $p_K > \frac{1}{2}$ streng monoton wachsend für $x \in X(\Omega) = \{0, \ldots, n\}$
Also gilt $\{x \in X(\Omega) \mid q(x) \geq c\} = \{t, t+1, \ldots, n\}$ für ein $t \in \{0, \ldots, n+1\}$.
Der optimale Test erfüllt somit $T : X(\Omega) \to [0,1]$ $x \mapsto \begin{cases} 1 & \text{falls} \quad x \geq t \\ 0 & \text{falls} \quad x < t-1 \\ \gamma & \text{falls} \quad x = t-1 \end{cases}$

mit $\gamma = \begin{cases} 0 & \text{falls} \quad q(t-1) < c \\ \in [0,1] \text{ beliebig} & \text{falls} \quad q(t-1) = c \end{cases}$

Sein Niveau ist $E_H(T) = P_H(X \geq t) + \gamma P_H(X = t-1)$.
Hiermit kann man t und γ aus $E_H(t) = \alpha$ berechnen.
In der Gleichung $E_H(t) = \alpha$ kommt aber p_K gar nicht vor.
$\Rightarrow$ der Test T ist schärfster Test für die Hypothese $H = \{\frac{1}{2}\}$ gegen $K =]\frac{1}{2}, 1]$ zum Niveau α.

Bemerkung 15.11 *(Allgemeine Vorgehensweise zur Bearbeitung eines Testpro-*

blems)

1. *Schritt: Vorüberlegungen: Was soll getestet werden ?*
   Welches $X : \Omega \to \Omega'$ entspricht der realen Hypothese ?
   Welche Verteilungen kommen für X in Frage ?
   Welches Signifikanzniveau ist gefragt ?
2. *Schritt: Erhebung der Daten*
3. *Schritt: Auswertung der Daten und Interpretation*

Definition 15.12 Sei $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, X, H)$ ein Testproblem, für das $X : \Omega \to \mathbb{R}$ eine Dichte besitzt.
Für jedes $x \in X(\Omega)$ sei $L_x : \Omega \to [0, \infty[\ \vartheta \mapsto P_\vartheta(x)$ die Likelihood-Funktion an der Stelle x.

1. Die Abbildung $X(\Omega) \to \mathbb{R}\ x \mapsto \frac{\sup\{L_x(\vartheta)|\vartheta\in K\}}{\sup\{L_x(\vartheta)|\vartheta\in H\}}$ heißt der ***Likelihood-Quotient***.
2. Ein Test $T : X(\Omega) \to [0, 1]$ heißt ***Likelihood-Quotienten-Test***, wenn es ein $c \in \mathbb{R}$ gibt mit $T(x) = \begin{cases} 1 & \text{falls} \quad q(x) > c \quad \text{(Nullmenge wird verworfen)} \\ 0 & \text{falls} \quad q(x) < c \end{cases}$.

Korollar 15.13 *Ist $(\Omega, \{P_\vartheta\}_{\vartheta\in\theta}, X, H)$ ein Testproblem, für das $X : \Omega \to \mathbb{R}$ eine Dichte besitzt und sind $H, K \subset \theta$ einfach, so existiert zu jedem $\alpha \in [0, 1]$ ein Likelihood-Quotienten-Test zum Niveau α.*

Beweis:
Der Beweis von Satz 15.9 überträgt sich auf die vorliegende Situation. □
Zahlreiche wichtige Tests lassen sich als Likelihood-Quotiententests darstellen und so als „optimal“ nachweisen.

Definition 15.14

1. Für $n \in \mathbb{N}_+$ heißt die durch die Dichte
   $f : \mathbb{R} \to [0, \infty[\ x \mapsto \frac{1}{\sqrt{n\pi}}\Gamma\left(\frac{n}{2}\right)^{-1}\Gamma\left(\frac{n+1}{2}\right)\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$ die ***t_n-Verteilung***.
2. Für $n \in \mathbb{N}_+$ heißt die durch die Dichte
   $f : \mathbb{R} \to [0, \infty[\ x \mapsto \begin{cases} \Gamma\left(\frac{n}{2}\right)^{-1} 2^{-\frac{n-2}{2}} x^{-\frac{x}{2}} & \text{für} \quad x \geq 0 \\ 0 & \text{für} \quad x < 0 \end{cases}$ gegebene Verteilung auf $\mathbb{R}$ die ***χ^2_n-Verteilung***.

Satz 15.15 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.*

1. *Sind $X_1, \dots, X_n : \Omega \to \mathbb{R}$ unabhängige, $N(0,1)$-verteilte Zufallsvariable, so ist $X_1^2 + \dots + X_n^2 : \Omega \to \mathbb{R}$ eine χ_n^2-verteilte Zufallsvariable.*
2. *Ist $X : \Omega \to \mathbb{R}$ eine $N(0,1)$-verteilte Zufallsvariable und $Y : \Omega \to \mathbb{R}$ eine von X unabhängige χ_n^2-verteilte Zufallsgröße, so ist $Z : \Omega \to \mathbb{R}$*
   $$\omega \mapsto \begin{cases} \frac{X(\omega)}{\sqrt{Y(\omega)/n}} & \textit{falls} \quad Y(\omega) > 0 \\ 0 & \textit{sonst} \end{cases}$$
   eine t_n-verteilte Zufallsvariable.
3. *Sind $X_1, \dots, X_n : \Omega \to \mathbb{R}$ $N(\mu, \sigma^2)$-verteilte Zufallsvariable und ist $Y : \Omega \to \mathbb{R}$*
   $\omega \mapsto \frac{\sqrt{n}}{s(\omega)} \left[\frac{1}{n}(X_1(\omega) + \dots + X_n(\omega)) - \mu\right]$ *mit $s : \Omega \to \mathbb{R}$*
   $\omega \mapsto \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i(\omega) - \frac{1}{n}(X_1(\omega) + \dots + X_n(\omega)))}$, *so ist Y eine t_{n-1}-verteilte Zufallsvariable.*

Beispiel 15.16 (Der t-Test)

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und seien $X_1, \dots, X_n : \Omega \to \mathbb{R}$ unabhängige, $N(\mu, \sigma^2)$-verteilte Zufallsvariable mit unbekannten Parametern $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+$.

Für ein geeignetes $\mu_0 \in \mathbb{R}$ ist zu testen, ob $\mu = \mu_0$ gilt.

$\Rightarrow \theta = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\}$,

$H = \{(\mu, \sigma^2) \in \theta \mid \mu = \mu_0\}$,

$K = \{(\mu, \sigma^2) \in \theta \mid \mu \neq \mu_0\}$

Sei $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$.

Dann hat X die Dichte $f : \mathbb{R}^n \to [0, \infty[$ $(x_1, \dots, x_n) \mapsto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \mathrm{e}^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2}$

$\Rightarrow$ man kann den Likelihood-Quotiententest ausrechnen und erhält $T : X(\Omega) \to [0,1]$ $x \mapsto \begin{cases} 1 & \text{falls} \quad |Y(\omega)| > t \\ 0 & \text{falls} \quad |Y(\omega)| < t \end{cases}$ für ein gewisses $t \in \mathbb{R}$, wobei Y die Funktion aus 15.15 3. bezeichnet.

(vgl. Krengel S. 171)

Der Test T heißt der t-Test.

Bestimmung von t:

Für $\beta \in \mathbb{R}$ heißt die Zahl $t_{n-1,\beta}$ mit $P(Y \leq t_{n-1,\beta}) = \beta$ das ***β-Quantil*** von t_{n-1}.

Um einen Test zum Niveau $\alpha \in [0,1]$ zu erhalten, muss man $t = t_{n-1,1-\frac{\alpha}{2}}$ wählen.

(vgl. Krengel S.174)

Beispiel 15.17 (Der χ^2-Test)

Sei $\tilde{\Omega} = \{1, \dots, r\}$ und $\tilde{P} : \tilde{\Omega} \to [0,1]$ mit $\tilde{P}(i) = p_i$ und $p_1 + \ldots + p_r = 1$.
Für $i = 1, \dots, n$ seien $X_i : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen mit $X(\Omega) = \tilde{\Omega}$ und Verteilung $P_{X_i} = \tilde{P}$.
Die Wahrscheinlichkeiten $p_1, \dots, p_r$ seien daher unbekannt, d.h.
$\theta = \{(p_1, \dots, p_r) \in [0,1]^n \mid p_1 + \dots, p_r = 1\}$
Für einen gegebenen Vektor $(\pi_1, \dots, \pi_r) \in [0,1]^r$ mit $\pi_1 + \ldots + \pi_r = 1$ ist zu testen, ob $(p_1, \dots, p_r) = (\pi_1, \dots, \pi_r)$ gilt.
Dazu wird $n \gg 0$ gewählt und eine Approximation durchgeführt.
Für $i = 1, \dots, r$ sei $Y_i : \Omega \to \mathbb{R}$ $\omega \mapsto \#\{j \mid X_j(\omega) = i\}$ die Häufigkeit des i-ten Versuchsergebnisses.
Weiter sei $X = (Y_1, \dots, Y_r) \to \mathbb{R}^r$.
Dann gilt $\tilde{\Omega} = X(\Omega) = \{(x_1, \dots, x_r) \in \mathbb{N}^r \mid x_1 + \ldots + x_r = n\}$.
Die Verteilung P_ϑ für $\vartheta = (p_1, \dots, p_r) \in \theta$ ist die Multinomialverteilung mit Parametern $p_1, \dots, p_r$.
Für $n \gg 0$ gilt $P_\vartheta((x_1, \dots, x_r)) \approx \left(\frac{1}{\sqrt{2\pi n}}\right)^{r-1} \frac{1}{\sqrt{p_1 \cdot \ldots \cdot p_r}} \mathrm{e}^{-\frac{1}{2}\sum\limits_{i=1}^{r} \frac{1}{np_i}(X_i - np_i)^2}$.
(vgl. Zitate im Krengel)
Wieder kann man für $n \gg 0$ den Likelihood-Quotienten-Test ausrechnen und erhält $T : X(\Omega) \to [0,1]$ $(x_1, \dots, x_r) \mapsto \begin{cases} 1 & \text{falls} \quad V^r(x_1, \dots, x_r) > c_\alpha \\ 0 & \text{sonst} \end{cases}$ mit einer Funktion $V^r : \mathbb{R}^r \to \mathbb{R}$ $(x_1, \dots, x_r) \mapsto \sum\limits_{i=1}^{r} \frac{(x_i - n\pi_i)^2}{n\pi_i}$
(vgl. Krengel S. 179)
Der Test T heißt der χ^2-Test.
Die Zahl c_α hängt vom gewählten Signifikanzniveau α ab.
Ist $F_{r-1} : \mathbb{R} \to [0,1]$ die Verteilungsfunktion der χ^2_{r-1}-Verteilung, so gilt $P_\vartheta(V^r \leq n) \to F_{r-1}(n)$ für $n \to \infty$.
Zur Bestimmung von c_α kann man das $(1-\alpha)$-Quantil von χ^2_{r-1} wählen, d.h. den Wert $F_{r-1}(c_\alpha) = 1 - \alpha$.
Die Approximation ist hinreichend gut, wenn $n \cdot \max\{\pi_1, \dots, \pi_r\} \geq 3$ gilt.
Für $r \geq 10$ genügt auch schon $n \cdot \max\{\pi_1, \dots, \pi_r\} \geq 1$.

Fakultät für Mathematik — 6. Mai 1998
Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 1: Kombinatorik

Aufgabe 1: (Einfache Urnenmodelle)
In einer Urne befinden sich $m \geq 1$ paarweise verschiedene Kugeln, aus denen Stichproben bestehend aus je $n \in \{1, \ldots, m\}$ Kugeln gezogen werden. Geben Sie in den folgenden Fällen eine Menge an, die die möglichen Versuchsergebnisse beschreibt, und bestimmen Sie die Anzahl der möglichen Versuchsergebnisse.
a) Geordnete Stichproben mit Rücklegen.
b) Geordnete Stichproben ohne Rücklegen.
c) Ungeordnete Stichproben mit Rücklegen.
d) Ungeordnete Stichproben ohne Rücklegen.

Aufgabe 2: (Eigenschaften von Binomialkoeffizienten)
Seien $m, n \in \mathbb{N}$. Beweisen Sie die folgenden Formeln.
a) $\binom{m}{n} = \binom{m}{m-n}$
b) $\binom{m}{n} = \binom{m-1}{n} + \binom{m-1}{n-1}$
c) $\sum\limits_{n=0}^{m} \binom{m}{n} = 2^m$
d) $\sum\limits_{n=0}^{m} (-1)^n \binom{m}{n} = 0$ für $m \geq 1$
e) $\sum\limits_{n=0}^{m} n \binom{m}{n} = m\, 2^{m-1}$

Aufgabe 3: Seien $k, \ell, m, n \in \mathbb{N}$ mit $k \leq \ell + m$. Beweisen Sie die Formel

$$\sum_{n=0}^{m} \binom{m}{n} \binom{\ell}{k-n} = \binom{\ell+m}{k}$$

Aufgabe 4: (Siebformel) Beweisen Sie Satz 1.10 der Vorlesung.
Sei M eine Menge, seien $r, n \in \mathbb{N}_+$, und seien $N_1, \ldots, N_r \subseteq M$ Teilmengen. Ferner sei eine Funktion $f : \mathfrak{P}(M) \longrightarrow \mathbb{R}$ gegeben mit $f(A \cup B) = f(A) + f(B)$ für disjunkte Teilmengen $A, B \in \mathfrak{P}(M)$. Schließlich sei

$$S_n = \{x \in M \mid x \in N_i \text{ für genau } n \text{ Indizes } 1 \leq i \leq r\}$$

Dann gilt

$$f(S_n) = \sum_{i=n}^{r} (-1)^{i-n} \binom{i}{n} \sum_{\substack{\{j_1,\ldots,j_i\}\subseteq\{1,\ldots,r\} \\ \text{i-elementige Teilmenge}}} f(N_{j_1} \cap \cdots \cap N_{j_i})$$

Aufgabe 5: (Das 8-Damenproblem)

a) Wie viele Möglichkeiten gibt es, acht Türme so auf ein Schachbrett zu stellen, daß sich keine zwei gegenseitig bedrohen?

b) Schreiben Sie ein Maple-Programm `Rooks`(), das die Möglichkeiten in a) alle berechnet und ihre Summe ausgibt. Repräsentieren Sie dazu jedes Feld als Paar (i, j) mit $1 \leq i, j \leq 8$ und jede Lösung von a) als `array`(1..8, 1..2). Die Summe ist komponentenweise zu bilden, und die Felder (i, j) jeder Lösung sind nach aufsteigendem i anzuordnen.

c) Schreiben Sie ein Maple-Programm `Queens`(), das alle Möglichkeiten berechnet und ausgibt, acht Damen so auf ein Schachbrett zu stellen, daß sich keine zwei gegenseitig bedrohen.

Hinweise: Verwenden Sie a). Schreiben Sie erst ein Programm `DiagCheck`($Feld1$, $Feld2$), das prüft, ob sich zwei Damen diagonal bedrohen. Letzteres ist genau dann der Fall, wenn $Feld1 - Feld2$ ein Vielfaches von $(1, 1)$ oder $(1, -1)$ ist.

Fakultät für Mathematik 15. Mai 1998
Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 2: Bedingte Wahrscheinlichkeiten

Aufgabe 6: a) Ein Prüfer hat 50 Standardfragen, von denen er in jeder Prüfung sechs zufällig auswählt. Ein Student kennt die Antwort auf 30 Fragen. Wie groß ist die Wahrscheinlichkeit, daß er die Prüfung besteht, wenn er dazu mindestens drei Fragen richtig beantworten muß?

b) In einer Überraschungseiersaison gibt es 120 verschiedene Spielzeuginhalte. Wie groß ist die Wahrscheinlichkeit, daß man von jedem Inhalt mindestens ein Exemplar hat, wenn man 500 Überraschungseier kauft?

Aufgabe 7: a) Von drei Münzen hat eine zwei silberne Seiten, eine zwei goldene, und eine eine silberne und eine goldene. Die Münzen werden zufällig auf drei Fächer verteilt. Man öffnet ein Fach und erblickt eine goldene Seite. Mit welcher Wahrscheinlichkeit ist die andere Seite dieser Münze ebenfalls golden?

b) Bei einem Fernsehquiz befindet sich hinter genau einer von drei Türen ein Preis, und der Showmaster weiß hinter welcher. Der Kandidat wählt eine Tür. Dann öffnet der Showmaster eine der beiden anderen Türen, hinter der sich kein Preis verbirgt. Der Kandidat kann nun die gewählte Tür öffnen oder die andere noch verschlossene Tür. Lohnt es sich für ihn zu wechseln?

Aufgabe 8: Karl und Zacharias werfen wiederholt eine Münze. Wenn Kopf kommt, erhält Karl einen Punkt, wenn Zahl kommt Zacharias. Wer zuerst N Punkte hat gewinnt. Wie groß ist Karls Gewinnwahrscheinlichkeit, wenn er gerade n Punkte hat und Zacharias m?

Aufgabe 9: Gewisse Würfel haben jeweils gleichwahrscheinliche Seiten mit Augenzahlen aus $\{1, 2, \dots, 6\}$, die sich zu 21 addieren. Wir sagen, ein Würfel besiege den anderen, wenn er bei genügend vielen Würfen öfter eine höhere Augenzahl zeigt als der andere. Entscheiden Sie (evtl. mit Hilfe von Maple), ob es drei solcher Würfel gibt, die sich reihum besiegen.

Aufgabe 10: (Der große Aktienzock)

Ein Sparer besitze $x \in \{0, 1, 2, \ldots, 1000000\}$ Euro. Um sein Sparziel "1 Million Euro" zu erreichen, habe er zwei Strategien zur Auswahl:

(I) *Der große Aktienzock:* Ist $x \leq 500000$ Euro, so kauft der Sparer für x Euro eine riskante Aktie. Nach einem Jahr steigt die Aktie mit Wahrscheinlichkeit p auf das Doppelte, und mit Wahrscheinlichkeit $1-p$ verliert der Sparer sein Geld vollständig. Ist $500000 < x < 1000000$, so riskiert der Sparer $1000000 - x$ Euro. Nach einem Jahr wiederholt er das Verfahren.

(II) *Die Sparbuchtaktik:* Der Sparer legt alles aufs Sparbuch. Mit Wahrscheinlichkeit p erhält er $1,5$ Prozent Zins im Jahr, und mit Wahrscheinlichkeit $1-p$ verliert er $1,5$ Prozent des Angesparten, weil mittlerweile die Inflation stark angezogen hat.

Sei $f_p : [0, 1000000] \longrightarrow [0, 1]$ die Funktion, für die $f_p(x)$ angibt, mit welcher Wahrscheinlichkeit der Sparer im Fall (I) irgendwann sein Ziel erreicht.

a) Beschreiben Sie $f_p(x)$ rekursiv. Berechnen Sie $f_p(400000)$ explizit.

b) In Abhängigkeit von p, bestimmen Sie die maximale Anzahl von Rekursionen, die notwendig ist, um $f_p(x)$ auf zehn Nachkommastellen genau zu berechnen.

c) Schreiben Sie eine Maple-Prozedur `Zock`(p, x), die $f_p(x)$ auf zehn Nachkommastellen genau berechnet. (Können Sie auch eine Prozedur schreiben, die $f_p(x)$ für ganze Zahlen x exakt berechnet?)

d) Beweisen Sie, daß die Funktion $f_p(x)$ stetig und monoton steigend ist.

e)* Zeigen Sie, daß die Funktion $f_p(x)$ fast überall differenzierbar ist und $f_p'(x) = 0$ erfüllt.

f)* In Abhängigkeit von p, bestimmen sie die für den Sparer günstigere Strategie. Überlegen Sie sich zuerst, was günstiger hier überhaupt heißen könnte. Implementieren Sie auch die Gewinnwahrscheinlichkeit bei (II) und vergleichen Sie sie für verschiedene p und x mit $f_p(x)$.

Fakultät für Mathematik
Universität Regensburg

20. Mai 1998

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 3: Klassische Wahrscheinlichkeitsverteilungen

Aufgabe 11: Beim Spiel "Mensch ärgere Dich nicht" darf die Figur eines Spielers erst dann auf das Feld, wenn dieser eine Sechs würfelt.
a) Wie groß ist die Wahrscheinlichkeit, daß ein Spieler nach zehn Runden immer noch nicht draußen ist?
b) Wieviele Runden müssen vier Spieler mindestens spielen, bis mit 90 Prozent Wahrscheinlichkeit alle vier draußen sind?

Aufgabe 12: Für $n \geq 2$ sei $\Omega = \{0,1\}^n$ der n-te Bernoulli-Raum mit Bernoulli-Verteilung P zum Parameter $p \in [0,1]$. Wir betrachten die Ereignisse $A_i = \{X_i = 1\}$ für $i = 1, \ldots, n$ und $B = \{(\omega_1, \ldots, \omega_n) \in \Omega \mid \omega_1 + \cdots + \omega_n$ ungerade $\}$.
a) Berechnen Sie $P(B)$. Stellen Sie $P(B)$ mit Hilfe der Werte der Binomialverteilung dar und folgern Sie eine Formel für Binomialkoeffizienten.
b) Welche der Familien $F_1 = \{A_1, \ldots, A_n\}$, $F_2 = F_1 \cup \{B\}$ und $F_3 = \{A_2, \ldots, A_n, B\}$ sind unabhängig?

Aufgabe 13: Der Korrektor einer Staatsexamensklausur stellte in 10 Prozent der Fälle einen Fehler in der Lösung einer Aufgabe fest. Ein Nachkorrektor fand heraus, daß dabei eine fehlerhafte Lösung mit 94 Prozent Wahrscheinlichkeit als solche erkannt wurde, und eine korrekte Lösung wurde mit 4,2 Prozent Wahrscheinlichkeit als falsch bezeichnet.
a) Wie groß ist die Wahrscheinlichkeit dafür, daß eine als falsch bezeichnete Lösung tatsächlich falsch ist?
b) Mit welcher Wahrscheinlichkeit kommt eine fehlerhafte Lösung unerkannt durch?
c) Arbeitet das Korrekturverfahren zufriedenstellend?
d) Wie ändern sich die Ergebnisse von a), b) und c), wenn es einen unabhängigen Zweitkorrektor gibt, der mit derselben (Un-) Zuverlässigkeit arbeitet, und wenn

nur Lösungen verworfen werden, die von beiden Korrektoren bemängelt wurden?

Aufgabe 14: Sei (Ω, P) der Stufenraum, der das Urnenmodell von Polya beschreibt, sei $\omega = (\omega_1, \dots, \omega_n) \in \Omega$, und sei $k = \#\{i \mid \omega_i = w\}$.
a) Beweisen Sie die Formel

$$P(\omega) = \frac{w(w+c)\cdots(w+(k-1)c)\cdot s(s+c)\cdots(s+(n-k-1)c)}{(w+s)(w+s+c)\cdots(w+s+(n-1)c}$$

b) Folgern Sie, daß für jede Permutation σ von $\{1, \dots, n\}$ die Gleichung $P(\omega) = P((\omega_{\sigma(1)}, \dots, \omega_{\sigma(n)}))$ gilt.

Aufgabe 15: (Das Josephus-Problem)
Während des jüdischen Aufstands gegen Rom (70 n. Chr.) wurden n Juden in einer Höhle eingeschlossen. Um der Sklaverei zu entgehen, vereinbarten sie sich gegenseitig wie folgt zu vernichten. Sie stellten sich entlang eines Kreises auf und numerierten sich mit 1 bis n durch. Beginnend mit Nummer k, wurde jede k-te Person niedergemacht, wobei sich der Kreis sofort wieder schloß. Der spätere Geschichtsschreiber Flavius Josephus stellte sich so auf, daß er als letzter übrig blieb, schreckte dann aber vor dem vereinbarten Selbstmord zurück.
a) Zeigen Sie, daß man die Nummer der s-ten eliminierten Person mit folgendem Algorithmus finden kann:

1) Setze $x_0 := 1 + k(n-s)$ und $q := \frac{k}{k-1}$.

2) Für $i = 1, 2, \dots$ berechne $x_i := \lceil q x_{i-1} \rceil$, wobei $\lceil y \rceil$ für $y \in \mathbb{R}$ die kleinste ganze Zahl $\geq y$ bezeichne.

3) Ist $a = \max\{x_i \mid x_i \leq kn\}$, so gib $kn + 1 - a$ als Ergebnis zurück.

b) Schreiben Sie eine Maple-Prozedur $\texttt{Josephus}(n, k, s)$, die die Nummer der s-ten umgebrachten Person ermittelt und ausgibt. Berechnen Sie

$$\texttt{Josephus}(40, 7, 40),$$

die Nummer, die angeblich dem historischen Fall entspricht.
c) Beweisen Sie, daß im Fall $k = 2$ der letzte Überlebende die Nummer $1 + 2n - 2^m$ hat, wobei $m = \lfloor \log_2(n) \rfloor + 1$ ist. (Für $y \in \mathbb{R}$ sei dabei $\lfloor y \rfloor$ die größte ganze Zahl $\leq y$.)
d)* Die Nummer einer bestimmten Person bei der ersten Zählung sei y_1, bei der zweiten Zählung y_2, etc. Beweisen Sie $y_{i+1} = y_i + n - \lfloor \frac{y_i}{k} \rfloor$, falls y_{i+1} definiert ist.

Zeigen Sie hiermit $y_i = \lfloor \frac{(y_{i+1}-n)k-1}{k-1} \rfloor$ und folgern Sie einen schnellen Algorithmus zur Berechnung von `Josephus`(n, k, s). Implementieren Sie diesen.

Fakultät für Mathematik — 28. Mai 1998

Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 4: Diskrete Wahrscheinlichkeitsräume

Aufgabe 16: Beweisen Sie, daß die Pascal-Verteilung normiert ist.

Aufgabe 17: Ein Buch mit 400 Seiten enthalte 400 Druckfehler, die zufällig verteilt sind.
a) Wie groß ist (approximativ) die Wahrscheinlichkeit dafür, daß auf den beiden ersten Seiten je genau ein Druckfehler vorkommt.
b) Geben Sie eine Abschätzung für die Wahrscheinlichkeit dafür an, daß auf den letzten 100 Seiten mehr als 120 oder weniger als 80 Druckfehler sind.

Aufgabe 18: (Verdoppelungsstrategie beim Roulette)
Beim Roulette ("Monte-Carlo-Roulette") wird eine der Zahlen in $\{0, 1, \ldots, 36\}$ ausgespielt. Setzt man auf "einfache Chance" (z.B. auf die ungeraden Zahlen "Impair"), so erhält man das Doppelte des Einsatzes im Erfolgsfall und die Hälfte des Einsatzes, wenn null ("Zero") kommt. Andernfalls ist der Einsatz weg.
Ein Spieler geht folgendermaßen vor: er setzt stets auf "Impair" und beginnt mit einer Einheit ("Jeton"). Verliert er seinen Einsatz ganz oder teilweise, so setzt er in der nächsten Runde das Doppelte. Gewinnt er, so beginnt er wieder mit einer Einheit.
Die Zufallsvariable X beschreibe den Reingewinn einer Serie von Spielen, die bis zum ersten Erfolg oder bis zum Erreichen des Höchsteinsatzes des Casinos dauert.
Berechnen Sie die Verteilung und den Erwartungswert von X, wenn
a) es keinen Höchsteinsatz und unbeschränktes Kapital gibt,
b) der Höchsteinsatz 1024 Jetons beträgt.

Aufgabe 19: (Die Ü-Eier-Sammlung)

a) In einer bestimmten Ü-Eier-Saison gebe es n verschiedene Spielzeuginhalte. Ein Sammler besitze bereits m verschiedene Inhalte. Wieviele Ü-Eier muß er im Durchschnitt kaufen, bis er ein neues Spielzeug findet?

b) Zeigen Sie, daß der Sammler im Durchschnitt $n(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n})$ Eier kaufen muß, bis seine Sammlung vollständig ist, wenn er bei null anfängt. (Was ergibt sich im realistischen Fall $n = 120$?)

Aufgabe 20: (Die Vietnam-Lotterie 1970)

Im Jahre 1970 wurde in den USA die Einberufung von Rekruten zum Einsatz im Vietnamkrieg folgendermaßen durchgeführt. Den Tagen des Jahres wurden "zufällig" die Ränge $1, 2, \ldots, 366$ zugeordnet. Die beigefügte Maple-Datei *Vietnam.mws* enthält diesen "Zufallskalender". Nacheinander wurden nun die Männer mit den Rängen $1, 2, \ldots$ einberufen, solange bis die Armee genügend Rekruten hatte. Männer mit Geburtstagen niedriger Ränge wurden fast sicher einberufen, solchen mit hohen Rängen praktisch nie.

a) Schreiben Sie ein Maple-Programm `Mittelwerte`(K), das die Liste der zwölf Mittelwerte der Ränge der Tage jedes Monats berechnet. Diese Mittelwerte sollten "zufällig" um 183 schwanken. Formulieren Sie einige Beobachtungen.

b) Nun sei σ die Permutation von $\{1, 2, \ldots, 12\}$, für die $\sigma(i)$ die Anzahl der Monate angibt, deren Mittelwert kleiner oder gleich dem Mittelwert des i-ten Monats ist. Berechnen Sie σ sowie die Anzahl C der Paare $(i, j) \in \{1, 2, \ldots, 12\}^2$ mit $i < j$ und $\sigma(i) < \sigma(j)$.

c) Beweisen Sie, daß die durchschnittliche Anzahl solcher Paare bei einer zufälligen Permutation von $\{1, 2, \ldots, 12\}$ gleich 33 ist.

d) Mit welcher Wahrscheinlichkeit hat eine zufällig gewählte Permutation von $\{1, 2, \ldots, 12\}$ höchstens C solcher Paare?

e) Schreiben Sie ein Maple-Programm `SimulPairs`$()$, die möglichst viele Zufallskalender erzeugt, die Monatsmittelwerte berechnet und die Anzahl der steigenden Paare bestimmt. Finden Sie so einen Näherungswert für die Wahrscheinlichkeit dafür, daß sich höchstens C steigende Paare ergeben. Welche Schlußfolgerungen ziehen Sie?

Fakultät für Mathematik
Universität Regensburg

4. Juni 1998

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 5: Erwartungswerte und Standardabweichungen

Aufgabe 21: Sei $r \in \mathbb{R}$ vorgegeben. Gibt es einen diskreten Wahrscheinlichkeitsraum (Ω, P) und eine Folge von Zufallsvariablen $X_n : \Omega \longrightarrow \mathbb{R}$ für $n \in \mathbb{N}$, so daß für alle $n \in \mathbb{N}$ der Erwartungswert $E(X_n)$ existiert und gleich r ist, und so daß $\lim\limits_{n\to\infty} E(\exp \circ X_n) = \infty$ gilt? Hierbei bezeichne $\exp : \mathbb{R} \longrightarrow \mathbb{R}$ die Exponentialfunktion.

Aufgabe 22: Eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$ sei hypergeometrisch verteilt mit Parametern n, w, s, d.h. für $i = 0, \ldots, n$ gelte $P_X(i) = \binom{w}{i}\binom{s}{n-i}/\binom{w+s}{n}$. Beweisen Sie:

a) $E(X) = \frac{nw}{w-s}$

b) $\mathrm{Var}(X) = \frac{nws(w+s-n)}{(w+s)^2(w+s-1)}$

Aufgabe 23: Zu der Weihnachtsfeier eines Vereins bringt jedes der n erschienenen Mitglieder ein Geschenk mit. Diese Geschenke werden dann unter den Anwesenden verlost, so daß jeder ein Geschenk erhält.

a) Mit welcher Wahrscheinlichkeit erhalten genau m Mitglieder ihr eigenes Geschenk zurück?

b) Wie kann man diese Wahrscheinlichkeit im Fall $n >> 0$ (z.B. 1. FC Bayern München) approximieren?

c) Wieviele Mitglieder erhalten durchschnittlich ihr Geschenk zurück?

Aufgabe 24: In der Losschachtel des Standes "Glückshafen" beim Volksfest befinden sich n Lose, unter denen sich m Gewinne (und $n-m$ Nieten!) verbergen. Ein Volksfestbesucher kauft solange ein Los und öffnet es, bis er (endlich!) einen Gewinn findet. Wieviele Lose muß er durchschnittlich kaufen?

Aufgabe 25: (Die moderne Portfoliotheorie (MPT) nach Markowitz)
Eine Summe von 1 Million Euro (EU) soll auf n verschiedene Anlageformen (z.B. Aktien, festverzinsliche Wertpapiere, Festgelder, Immobilien) verteilt werden. Der Ertrag einer Anlage in einem gewissen Zeitraum sei definiert als ihr Wert am Ende des Zeitraums geteilt durch den Anfangswert. Für eine Reihe von Zeiträumen $\{t_1, \ldots, t_r\}$ seien die Erträge der n Anlageformen bekannt und werden durch Zufallsvariable $X_i : \{t_1, \ldots, t_r\} \longrightarrow \mathbb{R}$ beschrieben ($i = 1, \ldots, n$). Das Ziel der MPT ist es, die Summe so auf die Anlageformen zu verteilen ("ein Portfolio zusammenzustellen"), daß der erwartete Gesamtertrag maximiert und die Standardabweichung ("das Risiko des Portfolios") minimiert wird. Man geht dabei von der naiven Annahme aus, daß sich von den vergangenen Erträgen und Korrelationen auf die zukünftigen schließen läßt.
a) Berechnen Sie den erwarteten Gesamtertrag und die Varianz des Portfolios aus den einzelnen Anlagesummen. Stellen Sie ein Gleichungssystem auf, das die "optimalen" Portfolios charakterisiert.
b) Zeigen Sie, daß es i.a. genau ein minimales Risiko σ_0 für das Portfolio gibt, und daß es für jedes $\sigma \geq \sigma_0$ genau ein Portfolio mit Risiko σ gibt, für das der erwartete Ertrag maximal ist.
c) Berechnen Sie mit Hilfe des Maple-Pakets `stats` die erwarteten Erträge, Standardabweichungen und Korrelationskoeffizienten der folgenden Anlageformen:

1) US-Aktien auf DM-Basis (S&P 500 Index),
2) deutsche Aktien (CDax Index bzw. Vorläufer davon),
3) deutsche Anleihen (5%ige Pfandbriefe),
4) das gute alte Sparbuch,
5) Gold !!!

Die Erträge dieser Anlageformen in den Jahren 1950 bis 1997 finden Sie in der beigefügten Datei "Markowitz.mws".
d) Berechnen Sie mit Maple im Fall c) das risikoärmste Portfolio sowie die Kurve der Erträge der optimalen Portfolios $E(\sigma)$ (die sogenannte "efficient frontier"). Verwenden Sie das Paket `plot`, um letztere graphisch darzustellen. Bestimmen Sie für einige Werte $\sigma \geq \sigma_0$ die zugehörigen optimalen Portfolios, z.B. für $\sigma = 0,1$ und $\sigma = 0,15$.

Fakultät für Mathematik 10. Juni 1998
Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 6: Erzeugende Funktionen

Aufgabe 26: Eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$, die auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) definiert ist, sei Pascal-verteilt zu den Parametern $p \in]0,1]$ und $r \in \mathbb{N}_+$, d.h. es gelte $X(\Omega) \subseteq \mathbb{N}$ und $P_X(k) = \binom{k+r-1}{k} p^r (1-p)^k$ für jedes $k \in \mathbb{N}$.

a) Bestimmen Sie eine geschlossene Formel für die erzeugende Funktion g_X von X.

b) Verwenden Sie g_X, um den Erwartungswert $E(X)$ und die Varianz $\mathrm{Var}(X)$ zu berechnen.

Aufgabe 27: Eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$, die auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) definiert ist, sei hypergeometrisch verteilt mit Parametern n, w, s.

a) Bestimmen Sie eine Formel für die erzeugende Funktion g_X von X.

b) Verwenden Sie g_X, um den Erwartungswert $E(X)$ und die Varianz $\mathrm{Var}(X)$ zu berechnen.

c) Zu einer vorgegebenen Zahl $p \in]0,1]$ sei $s = \frac{1-p}{p} w$. Beweisen Sie, daß die Formel $\lim\limits_{w \to \infty} P_X(k) = b_{n,p}(k)$ für $k = 0, \ldots, n$ gilt, d.h. daß die Binomialverteilung ein Grenzwert der hypergeometrischen Verteilung ist.

d) Interpretieren Sie das Ergebnis von c) anschaulich mit Hilfe eines geeigneten Urnenmodells.

Aufgabe 28: Die Lösungszeit, die ein Student durchschnittlich für eine Übungsaufgabe benötigt, sei zwei Stunden. Die Standardabweichung liege dabei aber bei sechs Stunden. Wieviele Übungsaufgaben kann der Professor seinen Studenten jede Woche aufbrummen, wenn er wünscht, daß durchschnittlich mindestens 95 % der Studenten alle Aufgaben lösen können?

Hinweis: Der Professor nimmt an, daß seine Studenten Tag und Nacht arbeiten, abzüglich täglich acht Stunden Schlaf, wöchentlich sechs Stunden Vorlesung bzw. Übungen, sowie zwei Stunden Biergarten nach der Freitagsvorlesung.

Aufgabe 29: Sei (Ω, P) ein diskreter Wahrscheinlichkeitsraum, sei $X : \Omega \longrightarrow \mathbb{R}$ eine reellwertige Zufallsvariable, sei $\epsilon > 0$, und sei $f : [0,\infty[\longrightarrow [0,\infty[$ eine monoton wachsende Funktion mit $f(\epsilon) > 0$. Zeigen Sie

$$P(\mid X - E(X) \mid \geq \epsilon) \leq \frac{1}{f(\epsilon)} E(f(\mid X - E(X) \mid)).$$

Aufgabe 30: (Simulation einer Kernreaktion)

Trifft ein Neutron auf den Kern eines Uranatoms, so kann es diesen spalten und dabei neue Neutronen erzeugen. Für jedes $k \in \mathbb{N}$ sei p_k die Wahrscheinlichkeit, daß genau k Neutronen entstehen. Diese treffen nach τ Sekunden auf weitere Uranatome. Für $n \in \mathbb{N}$ beschreibe die Zufallsvariable $X_n : \Omega \longrightarrow \mathbb{R}$ die Anzahl der Neutronen, die bei den Kernspaltungen nach $n\tau$ Sekunden entstehen, wobei wir von $X_0 = 1$ ausgehen.

a) Bestimmen Sie $g_{X_1}(t)$ und beweisen Sie, daß die erzeugende Funktion von X_n gegeben ist durch $g_{X_n}(t) = \underbrace{g_{X_1} \circ \cdots \circ g_{X_1}}_{n \text{ mal}}(t)$

b) Berechnen Sie $E(X_n)$ aus $E(X_1)$ sowie $\mathrm{Var}(X_n)$ aus $\mathrm{Var}(X_1)$, falls die letztere Varianz existiert.

c) Zeigen Sie, daß die Kettenreaktion mit der Wahrscheinlichkeit $q = \inf\{t \geq 0 \mid g_{X_1}(t) - t = 0\}$ zum Stillstand kommt. (Letzteres bedeutet $X_n(\Omega) = \{0\}$ für $n >> 0$.) Bestimmen Sie diese Wahrscheinlichkeit für den Fall, daß X_1 geometrisch verteilt ist.

d) Zeigen Sie, daß die Bedingung $q < 1$ damit äquivalent ist, daß $E(X_1) > 1$ oder $p_1 = 1$ gilt.

e) Schreiben Sie eine Maple-Prozedur `Kettenreaktion`(L), die die Werte der Zufallsvariablen $X_1, X_2, \ldots, X_{100}$ bei einer vorgegebenen endlichen Liste von Wahrscheinlichkeiten $L = [p_0, p_1, \ldots, p_n]$ simuliert. Verifizieren Sie durch oftmalige Anwendung dieser Prozedur das Ergebnis von c).

$f)^*$ Im folgenden sei $E(X_1) > 1$ vorausgesetzt. Betrachten Sie für $i \geq 1$ die Zufallsvariable $Y_i : \Omega \longrightarrow \mathbb{R}$ mit $Y_i(\omega) = X_i(\omega)/E(X_1)^i$ für $\omega \in \Omega$. Überzeugen Sie sich durch Simulation mittels `Kettenreaktion`(L), daß die Folge $Y_1(\omega), Y_2(\omega), \ldots$

für eine Menge Ω' von Elementen $\omega \in \Omega$ mit $P(\Omega') = 1$ gegen einen festen Wert $Y(\omega)$ konvergiert. Können Sie Formeln für den Erwartungswert und die Varianz der Zufallsvariablen $Y : \Omega' \longrightarrow \mathbb{R}$ erraten?
$g)^*$ Beweisen Sie die in f) empirisch gefundenen Formeln. Welche Folgerungen ergeben sich für das Wachstumsverhalten der Folge $X_0, X_1, X_2, \ldots$?

Fakultät für Mathematik 17. Juni 1998
Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 7: Stetige Wahrscheinlichkeitsverteilungen

Aufgabe 31: Sei Ω eine Menge und $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$. Zeigen Sie, daß $\mathcal{A}$ genau dann eine σ-Algebra ist, wenn die folgenden vier Eigenschaften erfüllt sind.
1) $\Omega \in \mathcal{A}$
2) Ist $A \in \mathcal{A}$, so gilt auch $\Omega \setminus A \in \mathcal{A}$.
3) Sind $A, B \in \mathcal{A}$, so gilt auch $A \cap B \in \mathcal{A}$.
4) Sind für $n \geq 1$ paarweise disjunkte Mengen $A_n \in \mathcal{A}$ gegeben, so gilt $\biguplus_{n\geq 1} A_n \in \mathcal{A}$.

Aufgabe 32: a) Zeigen Sie, daß alle offenen und alle abgeschlossenen Teilmengen von $\mathbb{R}$ Borel-meßbar sind.
b) Beweisen Sie, daß die von allen offenen Intervallen und die von allen offenen Teilmengen von $\mathbb{R}$ erzeugten σ-Algebren in $\mathbb{R}$ mit $\mathcal{B}(\mathbb{R})$ übereinstimmen.

Aufgabe 33: Zeigen Sie, daß durch die folgenden Funktionen Dichten gegeben sind, die Wahrscheinlichkeitsverteilungen auf $\mathbb{R}$ definieren.
a) Für $\alpha, \beta > 0$ ist die *Weibull-Verteilung* mit Parametern α, β gegeben durch die Dichte

$$\begin{aligned} f : \mathbb{R} &\longrightarrow [0, \infty[\\ x &\longmapsto \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} & \text{für } x \geq 0, \\ 0 & \text{für } x < 0. \end{cases} \end{aligned}$$

b) Für $\kappa > 0$ und $\varphi \in\,]0, 2\pi]$ ist die *von-Mises-Verteilung* mit Parametern κ, φ gegeben durch die Dichte

$$\begin{aligned} f : \mathbb{R} &\longrightarrow [0, \infty[\\ x &\longmapsto \begin{cases} \frac{1}{2\pi I_0(\kappa)} \cdot e^{\kappa \cos(x-\varphi)} & \text{für } x \in [-\pi, \pi[, \\ 0 & \text{sonst}, \end{cases} \end{aligned}$$

wobei $I_0(\kappa) = \sum_{i=0}^{\infty} \frac{1}{(i!)^2} (\frac{\kappa}{2})^{2i}$ gelte.

Aufgabe 34: a) Die Abbildung $\Gamma : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ mit $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ heißt die *Gammafunktion*. Zeigen Sie $\Gamma(n) = (n-1)!$ für $n \geq 1$ sowie $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, und finden Sie eine Formel für $\Gamma(\frac{n}{2})$ bei ungeradem $n \in \mathbb{N}_+$.

Beweisen Sie, daß durch die folgenden Funktionen Dichten gegeben sind, die Wahrscheinlichkeitsverteilungen auf $\mathbb{R}$ definieren.

b) Für $\alpha, \nu \in \mathbb{R}_+$ ist die *Gamma-Verteilung* mit Parametern α, ν gegeben durch die Dichte

$$\begin{aligned} f : \mathbb{R} &\longrightarrow [0, \infty[\\ x &\longmapsto \begin{cases} \Gamma(\nu)^{-1}\alpha^\nu x^{\nu-1}e^{-\alpha x} & \text{für } x \geq 0, \\ 0 & \text{für } x < 0. \end{cases} \end{aligned}$$

c) Für $n \in \mathbb{N}_+$ ist die χ_n^2*-Verteilung* gegeben durch die Dichte

$$\begin{aligned} f : \mathbb{R} &\longrightarrow [0, \infty[\\ x &\longmapsto \begin{cases} \Gamma(\frac{n}{2})^{-1}2^{-n/2}x^{(n-2)/2}e^{-x/2} & \text{für } x \geq 0, \\ 0 & \text{für } x < 0. \end{cases} \end{aligned}$$

d) Für $n \in \mathbb{N}_+$ ist die *t-Verteilung* (oder *Student-Verteilung*) mit n Freiheitsgraden gegeben durch die Dichte

$$\begin{aligned} f : \mathbb{R} &\longrightarrow [0, \infty[\\ x &\longmapsto \frac{1}{\sqrt{n\pi}}\Gamma(\tfrac{n}{2})^{-1}\Gamma(\tfrac{n+1}{2})(1 + \tfrac{x^2}{n})^{-(n+1)/2} \end{aligned}$$

Aufgabe 35: (Simulation geometrischer Wahrscheinlichkeiten)

Im folgenden sind die passenden Zufallsexperimente als Maple-Prozeduren zu implementieren und jeweils mindestens 10000mal zu wiederholen, um so gute Schätzungen für die gesuchten Erwartungswerte zu erhalten.

a) Wählt man zwei Punkte zufällig im Einheitskreis, so beträgt ihr erwarteter Abstand etwa $d_1 \approx 0,90541$.

b)* Beweisen Sie die exakte Formel $d_1 = \frac{128}{45\pi}$.

c) Wählt man zwei Punkte zufällig im Einheitsquadrat, so beträgt ihr erwarteter Abstand etwa $d_2 \approx 0,52141$.

d)* Beweisen Sie die exakte Formel $d_2 = \frac{1}{15}(2 + \sqrt{2} + 5\ln(1 + \sqrt{2}))$.

e) Wählt man zwei Punkte zufällig in einem gleichseitigen Dreieck mit Seitenlänge eins, so beträgt ihr erwarteter Abstand etwa $d_3 \approx 0,36479$.

f)* Beweisen Sie die exakte Formel $d_3 = \frac{1}{5} + \frac{3}{20}\ln 3$.

g) Wählt man zwei Punkte zufällig im Einheitswürfel, so beträgt ihr erwarteter Abstand etwa $d_4 \approx 0,5844$.
h)* Beweisen Sie die exakte Formel $d_4 = \frac{1}{105}(4 + 17\sqrt{2} - 6\sqrt{3} + 21\ln(1+\sqrt{2}) + 42\ln(2+\sqrt{3}) - 7\pi)$.

Fakultät für Mathematik
Universität Regensburg

25. Juni 1998

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 8: Stetige Zufallsvariable

Aufgabe 36: (Woher der Wind weht)
Die an einer Wetterstation gemessenen Windrichtungen genügen der von-Mises-Verteilung mit Parametern $\kappa = \frac{1}{2}$ und $\varphi = 2\pi$, wobei die vorherrschende Windrichtung Westen sei. Mit welcher Wahrscheinlichkeit stimmt ein zufällig erzeugter Wetterbericht, der "Wind aus östlichen Richtungen" (also zwischen SO und NO) vorhersagt?

Aufgabe 37: (Der Super-GAU bei der Fußball-WM)
Für ein Fernsehgerät, das bereits t Jahre alt ist, sei die Wahrscheinlichkeit, innerhalb der nächsten kurzen Zeit Δt auszufallen, annähernd gleich $r(t)\Delta t$. Die Funktion $r : \mathbb{R}_+ \longrightarrow [0, \infty[$ heißt dabei die *Ausfallrate.* Zeigen Sie die folgenden Behauptungen.
a) Besitzt die Verteilung der Lebensdauer des Fernsehers eine stetige Dichte f, so gilt $r(t) = f(t)/P(]t, \infty[)$ für alle $t \in \mathbb{R}_+$ mit $P(]t, \infty[) \neq 0$.
b) Ist $r(t)$ proportional zu t^ϱ für ein $\varrho > -1$, so ist die Lebensdauer des Fernsehers Weibull-verteilt.
c) Die Lebensdauer eines t Minuten alten Fernsehgeräts sei Weibull-verteilt mit Parametern $\alpha = 1$ und $\beta = 2$. Mit welcher Wahrscheinlichkeit fällt das Gerät gerade während der 90 Minuten des Endspiels der Fußball-WM aus?

Aufgabe 38: (Die Dauer des Elfmeterschießens)
Bei einem Elfmeterschießen werden n Elfmeter ausgeführt. Bestimmen Sie die Dichte, den Erwartungswert und die Varianz der Zufallsvariablen, die die Dauer des Elfmeterschießens beschreibt, wenn die Dauer eines einzelnen Elfmeters exponentialverteilt ist zum Parameter $\lambda \in \mathbb{R}_+$.

Aufgabe 39: Beweisen Sie, daß jede Verteilungsfunktion auf $\mathbb{R}$ höchstens an abzählbar vielen Stellen unstetig ist.

Aufgabe 40: (Die Irrfahrt des Raumschiffs Voyager)
Der Weltraum. Unendliche Weiten. Wir schreiben das 24. Jahrhundert. Dies sind die Abenteuer des Sternenschiffes U.S.S. Voyager, welches sich nach dem Zusammentreffen mit einer riesigen Strahlungswelle in einer fernen Galaxis wiederfindet, mehr als 70000 Lichtjahre von zu Hause entfernt.

a) Die Wissenschaftler an Bord der U.S.S. Voyager haben einen neuen Hyper-Warp-Antrieb entdeckt, der das gesamte Schiff innerhalb einer Folge der Sendung an einen 10000 Lj. entfernten Ort beamt. Leider gibt es noch keinerlei Möglichkeit, die Richtung der Fortbewegung zu kontrollieren. Schreiben Sie eine Maple-Prozedur `HyperWarp`(P), die zufällig einen Ort auf der Sphäre mit Radius 10000 Lj. um den Punkt $P = [x, y, z]$ auswählt und ausgibt. (Verwenden Sie die Gleichverteilung auf der Sphäre.)
b) Finden Sie durch Simulation mittels einer Maple-Prozedur `EThome`(N) die Wahrscheinlichkeit dafür, daß die Serie nach höchstens $N = 50$ Folgen abgebrochen werden muß, weil das Raumschiff zufällig in die Nähe der Heimat gebeamt wird. (Am Anfang der Serie sei die U.S.S. Voyager am Nullpunkt. Die Umgebung des Sonnensystems bilde eine Kugel vom Radius 10000 Lj. um den Punkt $[70000, 0, 0]$.)
c) Sei D_n die Distanz der U.S.S. Voyager von ihrem Startpunkt $[0, 0, 0]$ nach n Folgen. Bestimmen Sie durch Simulation Näherungswerte für den Erwartungswert $E(D_n)$ und die Standardabweichung $\sigma(D_n)$ für einige n . Können Sie von n abhängige Formeln erraten?

Fakultät für Mathematik — 3. Juli 1998

Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 9: Erwartungswerte stetiger Zufallsvariablen

Aufgabe 41: (Warten auf das "Golden Goal")
Bei einem Spiel der Fußball-WM kommt es zur Verlängerung. Die Zeit, die es dauert, bis Spieler $i \in \{1, \ldots, 22\}$ ein Tor erzielt, sei exponentiell verteilt zum Parameter $\lambda_i > 0$. Bestimmen Sie die durchschnittliche Dauer der Verlängerung sowie die Wahrscheinlichkeit dafür, daß es zu einem Elfmeterschießen kommt.

Aufgabe 42: Eine reellwertige Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ besitze eine der folgenden Verteilungen. Entscheiden Sie jeweils, ob ihr Erwartungswert und ihre Varianz existieren, und berechnen Sie diese gegebenenfalls.
a) X ist Gamma-verteilt mit Parametern $\alpha, \nu \in \mathbb{R}_+$.
b) X ist χ_n^2-verteilt mit $n \in \mathbb{N}_+$.
c) X ist t-verteilt mit $n \in \mathbb{N}_+$ Freiheitsgraden.

Aufgabe 43: Sei $n \in \mathbb{N}_+$, und seien $X_1, \ldots, X_n : \Omega \longrightarrow \mathbb{R}$ unabhängige, standardnormalverteilte Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Berechnen Sie die Dichte der Zufallsvariablen $X = X_1^2 + \cdots + X_n^2 : \Omega \longrightarrow \mathbb{R}$. Um welche Verteilung handelt es sich bei P_X?

Aufgabe 44: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, und seien $X, Y : \Omega \longrightarrow \mathbb{R}$ zwei unabhängige Zufallsvariable, deren Erwartungswerte existieren. Beweisen Sie, daß dann auch $E(XY)$ existiert, und es gilt $E(XY) = E(X)E(Y)$.
(Hinweis: Betrachten Sie zunächst den Fall, daß es ein $n \in \mathbb{N}_+$ gibt mit $X(\Omega) \cup Y(\Omega) \subseteq [1, n]$.)

Aufgabe 45: (Das Blitzschachmatch)

Beim Blitzschach erhält jeder Spieler eine Gesamtbedenkzeit von fünf Minuten pro Partie, so daß an einem Abend eine Reihe von Partien ausgetragen werden kann. Im direkten Vergleich zweier Schachspieler besitze Spieler A die Gewinnwahrscheinlichkeit $p \in]0,1[$ und Spieler B die Gewinnwahrscheinlichkeit $1-p$, wobei Remisen (d.h. unentschiedene Partien) nicht mitgerechnet werden. Bestimmen Sie in den folgenden Fällen jeweils durch Simulation mit Hilfe geeigneter Maple-Prozeduren für möglichst viele Fälle p, m, n die gesuchten Wahrscheinlichkeiten.

a) **(Das Standardmatch)** Der Verlierer zahlt dem Gewinner einer Partie eine DM. Mit welcher Wahrscheinlichkeit P_0 verliert Spieler A irgendwann sein ganzes Geld, wenn er m DM dabei hat und Spieler B hat n DM.

b) Beweisen Sie $P_0 = \frac{n}{m+n}$ im Fall $p = \frac{1}{2}$ sowie $P_0 = \frac{(1-p)^{m+n}-(1-p)^m p^n}{(1-p)^{m+n}-p^{m+n}}$ im Fall $p \neq \frac{1}{2}$.

c) **(Das Zockermatch)** Um den Umsatz und den Nervenkitzel zu erhöhen, beschließen die Spieler nach einer Weile, die Gesamtbedenkzeit pro Partie und Spieler auf zwei Minuten zu verkürzen und den Einsatz auf zwei DM zu erhöhen. Später spielen Sie sogar mit einer Minute Bedenkzeit um fünf DM Einsatz. Berechnen Sie wieder die Wahrscheinlichkeit dafür, daß sich Spieler A ruiniert. Bestimmen Sie auch das Grenzwertverhalten dieser Wahrscheinlichkeit bei weiteren Erhöhungen des Einsatzes.

d) **(Raketenschach)** Die beiden Spieler starten mit je fünf Minuten Bedenkzeit. Der Sieger erhält bei der nächsten Partie jeweils eine Minute Bedenkzeit weniger. Besitzt Spieler A irgendwann i Minuten Bedenkzeit und Spieler B hat j Minuten, so gewinnt Spieler A mit Wahrscheinlichkeit $\frac{pi}{pi+(1-p)j}$. Das Match ist beendet, sobald einer der Spieler auf fünf Siege kommt. Mit welcher Wahrscheinlichkeit gewinnt Spieler A das Match?

Fakultät für Mathematik 9. Juli 1998
Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 10: Grenzwertsätze

Aufgabe 46: Eine reellwertige Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ sei Weibull-verteilt mit Parametern $\alpha, \beta \in \mathbb{R}_+$. Zeigen Sie, daß die Zufallsvariable $X^\beta : \Omega \longrightarrow \mathbb{R}$ exponentialverteilt ist.

Aufgabe 47: Der Einstellungstest einer Firma besteht aus 99 Fragen, die nach aufsteigendem Schwierigkeitsgrad angeordnet sind. Die Wahrscheinlichkeit dafür, daß ein gewisser Kandidat die i-te Frage richtig beantwortet, sei $1 - \frac{i}{100}$. Mit welcher Wahrscheinlichkeit wird er eingestellt, wenn dazu die korrekte Antwort auf mindestens 60 Fragen verlangt wird?

Aufgabe 48: (Bier ist Bier und Schnaps ist Schnaps)
Ein Bartender schenkt Schnaps aus einer großen 3-Liter-Flasche aus. Ein Drink ("Stamperl") enthält durchschnittlich 5 cl Schnaps mit einer Standardabweichung von 0,5 cl. Mit welcher Wahrscheinlichkeit reicht die Flasche für 65 Drinks?

Aufgabe 49: Ein Bernoulli-Versuch mit Erfolgswahrscheinlichkeit $p \in]0, 1[$ werde unendlich oft wiederholt. Für $n \in \mathbb{N}$ sei A_n das Ereignis, daß zwischen dem 2^n-ten Versuch (einschließlich) und dem 2^{n+1}-ten Versuch (ausschließlich) ein Run der Länge n auftritt. Mit welcher Wahrscheinlichkeit treten unendlich viele der Ereignisse A_n ein?

Aufgabe 50: (Die MEL-Regel)
In der Situation der modernen Portfoliotheorie (vgl. Übungsaufgabe 25) sei $X : \Omega \longrightarrow \mathbb{R}$ die Zufallsvariable, die für $\Omega = \mathbb{N}$ den Gesamtertrag des Portfolios im n-ten Anlagezeitraum beschreibt, und $W : \Omega \longrightarrow \mathbb{R}$ gebe den Wert des Portfolios nach n Anlagezeiträumen an.

a) Zeigen Sie, daß fast sicher

$$\lim_{n \to \infty} \frac{1}{n} \ln(\frac{W_n}{W_0}) = E(\ln(X))$$

gilt. Folgern Sie, daß ein Portfolio mit Ertrag $X : \Omega \longrightarrow \mathbb{R}$ ein anderes mit Ertrag $Y : \Omega \longrightarrow \mathbb{R}$ auf lange Sicht fast sicher übertrifft (d.h. sein Gesamtwert nach $n >> 0$ Anlageperioden ist höher), wenn $E(\ln(X)) > E(\ln(Y))$ gilt.
Die *MEL-Regel* ("Maximum Expected Log") besagt, daß unter den optimalen Portfolios aus Aufgabe 25 dasjenige zu wählen ist, für das $E(\ln(X))$ maximal ist.
b) Beweisen Sie die Näherungsformel

$$E(\ln(X)) \approx \ln E(X) - \frac{1}{2} \frac{\mathrm{Var}(X)}{E(X)^2}$$

(Hinweis: Schreiben Sie $X = 1 + Y$ und berechnen Sie die Taylor-Entwicklung an der Stelle $E(X) = 1 + E(Y)$.)
c) Stellen Sie sich vor, Sie seien verantwortlich für die Verwaltung eines Aktienfonds, der nur in Indexzertifikate der folgenden Länder investieren darf:
1) Deutschland 2) Frankreich 3) Großbritannien 4) Japan 5) USA
Die historischen Jahreserträge (von 1950 bis 1997, in DM) dieser Wertpapiere liegen in der Datei "laender.mws" vor. Schreiben Sie eine Maple-Prozedur `MELrule`(L), die das optimale Portfolio berechnet, das die MEL-Regel erfüllt. Wie ist die Anlagesumme prozentual auf die einzelnen Indexzertifikate zu verteilen?

Fakultät für Mathematik
Universität Regensburg

17. Juli 1998

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 11: Schätztheorie

Aufgabe 51: Bei einer Qualitätsprüfung wurden 100 Computerchips geprüft, von denen sich vier als defekt erwiesen. Bestimmen Sie eine Maximum-Likelihood-Schätzung für die Anzahl der defekten Chips in einer Lieferung von 10000 Stück.

Aufgabe 52: (Schätzung eines Fischbestandes)
Ein Teich enthält eine unbekannte Zahl N von Fischen, die geschätzt werden soll. Dazu werden n Fische gefangen, markiert und wieder ausgesetzt. Nach einer Weile werden wieder n Fische gefangen, unter denen m markierte Fische sind.
a) Bestimmen Sie den Maximum-Likelihood-Schätzer T für N.
b) Ist T erwartungstreu? Berechnen Sie den Bias b_T.
c) Berechnen Sie die Varianz von T sowie den mittleren quadratischen Fehler R_T.
d) Geben Sie für $n \leq 5$ Konfidenzintervalle zum Niveau 0,95 an.

Aufgabe 53: Ein Gerät enthält die störanfälligen Teile A_1, A_2 und B. Es funktioniert, wenn B und mindestens eines der Teile A_1, A_2 funktioniert. Aus Untersuchungen der einzelnen Bauteile sind Konfidenzintervalle für die Wahrscheinlichkeiten p_A und p_B bekannt, daß ein Teil vom Typ A bzw. B nach einem Jahr noch funktioniert.
Leiten Sie unter der Annahme, daß Störungen in den Bauteilen unabhängig voneinander auftreten, zu einem eventuell kleineren Niveau ein möglichst kleines Konfidenzintervall für die Wahrscheinlichkeit her, daß das Gerät nach einem Jahr noch funktioniert.

Aufgabe 54: Unter der Annahme, daß Knaben- und Mädchengeburten gleich wahrscheinlich sind, sei $1-\vartheta$ die Wahrscheinlichkeit, daß ein Knabe rot-grün-blind ist. Nach einem genetischen Modell ist dann $(1-\vartheta)^2$ die Wahrscheinlichkeit dafür, daß ein Mädchen rot-grün-blind ist. Es sind 8870 Männer und 9140 Frauen untersucht worden, und davon waren 670 Männer und 40 Frauen rot-grün-blind. Bestimmen Sie eine Maximum-Likelihood-Schätzung für ϑ.

Aufgabe 55: Je zehn Messungen der Winkel α, β, γ eines Dreiecks ergeben folgende Mittelwerte: $\overline{\alpha} = 61,3°$, $\overline{\beta} = 44,6°$, $\overline{\gamma} = 73,4°$. Die Winkelsumme ist auf Grund der Meßfehler nicht 180°. Berechnen Sie mit Hilfe der Methode der kleinsten Quadrate Schätzungen für α, β, γ mit Winkelsumme 180°.

Fakultät für Mathematik 24. Juli 1998
Universität Regensburg

Übungen zur Vorlesung
Einführung in die Wahrscheinlichkeitstheorie

Blatt 12: Testtheorie

Aufgabe 56:* Bestimmen Sie unter Verwendung der Normalapproximation den Verwerfungsbereich eines Tests zum Niveau 0,05, der bei 1000 Würfen einer Münze entscheiden soll, ob die Münze fair ist.

Aufgabe 57:* Es wird vermutet, daß bei gewissen Pferderennen auf einer kreisförmigen Rennbahn die Startposition einen Einfluß auf die Gewinnchancen hat. In 144 Rennen hatten die Sieger die Startpositionen $1, 2, \ldots, 8$ mit den folgenden Häufigkeiten: 29,19,18,25,17,10,15,11. Testen Sie die Hypothese, daß alle Startpositionen die gleiche Siegwahrscheinlichkeit besitzen, zum Signifikanzniveau 0,05.

Aufgabe 58:* In einer fernen Galaxie bricht zu einem unbekannten Zeitpunkt ϑ eine Supernova aus und sendet einen Lichtblitz der Dauer t. Es sei bekannt, daß die Dauer solcher Lichtblitze exponentialverteilt ist zu einem bereits gemessenen Parameter λ. Nur das Ende des Lichtblitzes konnte zeitlich bestimmt werden. Beschreiben Sie einen schärfsten Test zum Niveau $\alpha = 0,05$ für die Hypothese $\vartheta \leq 3$ gegen die Alternative $\vartheta > 3$.

Aufgabe 59:* Bei Rouletterädern tritt manchmal ein gewisser Defekt auf, so daß eine bestimmte Zahlenkombination nicht mit Wahrscheinlichkeit p, sondern mit Wahrscheinlichkeit $q \neq p$ ausgespielt wird. Zeigen Sie, daß es für $n >> 0$ stets einen Test gibt, der aus n unabhängigen Wiederholungen des Spiels besteht, und der mit einem Niveau $\leq 0,05$ sowie einer Macht $\geq 0,95$ prüft, ob ein gegebenes Rouletterad in Ordnung ist.

Aufgabe 60:* **(Darwins Versuch)**
Charles Darwin hatte 15 Paare von Maissamen in 15 Töpfe gepflanzt. Der eine Samen war jeweils durch Kreuzung, der andere durch Selbstbefruchtung entstanden. Für $i = 1, \ldots, 15$ verglich er nach einiger Zeit die Höhe x_i der gekreuzten

Pflanze mit der Höhe y_i der selbstbefruchteten Pflanze und erhielt die folgenden Unterschiede $z_i = x_i - y_i$ (in $\frac{1}{8}$-inch):

$$6, 8, 14, 16, 23, 24, 28, 29, 41, -48, 49, 56, 60, -67, 75$$

Es soll die Hypothese "Es gibt keinen Unterschied zwischen den beiden Samenarten" gegen die Alternative "Gekreuzte Samen entwickeln stärkere Pflanzen" getestet werden.

a) Ist $n \in \mathbb{N}_+$ und $D = \{d_1, \ldots, d_n\} \in \mathbb{Z}$, so sei $q(t, D)$ für $t \in \mathbb{Z}$ die Anzahl aller Teilmengen von D, deren Summe höchstens t ist. Finden Sie eine rekursive Beschreibung der Zahlen $q(t, D)$.

b) Schreiben Sie eine Maple-Prozedur, die die Funktion $q(t, D)$ berechnet.

c) Berechnen Sie die Wahrscheinlichkeit dafür, daß unter der Hypothese die Summe der negativen Höhendiffenrenzen höchstens 115 ist.

d) Formulieren Sie ganz allgemein eine Testmethode, die die obige Vorgehensweise mathematisch beschreibt.