Wahrscheinlichkeitstheorie und Statistik A

Vorlesung von Prof. Dr. N. Bäuerle
Mitschrift: Moritz Diehle[1]
Bilder und Layout: Daniel Zaum[2]
Version 0.9.1
Vielen Dank an alle, die Fehler gemeldet haben.
Alle gefundenen Fehler bitte melden.
Alle Angaben ohne Gewähr.

7. Februar 2003

[1]Moritz.Diehle@stud.uni-hannover.de
[2]cowhen@cowhen.com

Inhaltsverzeichnis

1 Wahrscheinlichkeitstheorie **2**

Wahrscheinlichkeitstheorie **2**
§1 Wahrscheinlichkeitsräume und Kombinatorik 2
Satz 1.2: Siebformel 4
§2 Unabhängigkeit von Ereignissen und bedingte Wahrscheinlichkeit 8
Lemma 2.1: Multiplikationssatz 9
Satz 2.2: Satz von der totalen Wahrscheinlichkeit 9
Satz 2.3: Bayes´sche Formel 10
§3 Zufallsvariablen und deren Verteilung 11
Satz 3.5: Dichtetransformation 22
§4 Charakteristiken von Zufallsvariablen und Zufallsvektoren 24
4.1 Erwartungswert und Momente von Zufallsvariablen 24
4.2 Charakteristiken mehrdimensionaler Verteilungen 27
§5 Abschätzung für Erwartungswerte und Wahrscheinlichkeiten 31
Satz 5.1: Ungleichung von Cauchy-Schwarz 31
Satz 5.2: Ungleichung von Chebyshev 31
§6 Erzeugende Funktionen 33
Satz 6.2: Eindeutigkeitssatz für erzeugende Funktionen 33
Satz 6.4: Stetigkeitssatz für erzeugende Funktionen 34
Satz 6.7: Eindeutigkeitssatz 36
Satz 6.8: Stetigkeitssatz 36
§7 Grenzwertsätze 37
7.1 Der zentrale Grenzwertsatz 37
Satz 7.1: Zentraler Grenzwertsatz 37
Satz 7.3: Berry Esséen 39
7.2 Gesetze der großen Zahlen 39
Satz 7.4: Schwaches Gesetz der großen Zahlen 39
Satz 7.5: Starkes Gesetz der großen Zahlen 40
§8 Erzeugung von Zufallszahlen 42
§9 Randomisierte Algorithmen 44
Index 45

Kapitel 1

Wahrscheinlichkeitstheorie

[1] = mathematische Analyse zufälliger Erscheinungen, die

1. Keine deterministische Regularität besitzen
2. eine statistische Regularität besitzen

Beispiel: Wahrscheinlichkeit, daß ich im Lotto 6 Richtige habe ist sehr klein.
Wahrscheinlichkeit, daß irgendjemand 6 Richtige hat ist groß.
Beispiel: Wie viele Leute müssen in einer Gruppe sein, damit die Wahrscheinlichkeit $\geq \frac{1}{2}$ ist, daß 2 am gleichen Tag Geburtstag haben? $n = 23$.

§1 Wahrscheinlichkeitsräume und Kombinatorik

Gegeben sei ein Experiment (z.B. Würfeln):
Ω = Raum der Versuchsausgänge (z.B. $\Omega = \{1, 2, 3, 4, 5, 6\}$)
$\omega \in \Omega$ Elementarereignis
$A \subset \Omega =$ Ereignis (z.B. $A = \{2, 4, 6\}$ „gerade Zahl fällt“)

Beispiel: Zuerst wird eine Münze geworfen. Fällt Kopf wird mit einem Würfel geworfen, fällt Zahl, so wird nochmal mit der Münze geworfen.

$$\Omega = \{K1, K2, K3, K4, K5, K6, ZK, ZZ\}$$

Ereignissen $A \subset \Omega$ werden jetzt Wahrscheinlichkeiten zugeordnet. Ist Ω nicht abzählbar, z.B. $\Omega = [0, 1]$, so können nicht alle Mengen $A \subset [0, 1]$ Wahrscheinlichkeiten zugeordnet werden, um ein sinnvolles mathematisches Modell zu bekommen.

Definition:

a) Eine Familie $\mathcal{A}$ von Teilmengen von Ω heißt *Algebra*, falls gilt:

(i) $\Omega \in \mathcal{A}$
(ii) $A \in \mathcal{A} \Longrightarrow A^c = \Omega - A \in \mathcal{A}$
(iii) $A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A}$

[1] 17.10.2002

b) Eine Algebra heißt σ-*Algebra*, falls zusätzlich gilt:

$$A_1, A_2, \ldots \in \mathcal{A} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

Beispiele: „Würfeln“
$\Omega = \{1,2,3,4,5,6\}$
$\mathcal{A}_1 = \{\emptyset, \Omega\}$
$\mathcal{A}_2 = \{\emptyset, \{2,4,6\}, \{1,3,5\}, \Omega\}$
$\mathcal{A}_3 = \mathcal{P}(\Omega) =$ Menge aller möglichen Teilmengen.
Sind σ-Algebren.

Bemerkung:

(i) Das Paar $(\Omega, \mathcal{A})$ mit $\mathcal{A}$ σ-Algebra heißt *Messraum*.

(ii) Jede σ-Algebra ist abgeschlossen gegen abzählbare Durchschnitte.

(iii) Sind $A, B \subset \Omega$ und $A \cap B = AB = \emptyset$ (A und B sind disjunkt), so schreiben wir $A + B$ statt $A \cup B$.

(iv) $\mathcal{A} = \mathcal{P}(\Omega)$ ist stets eine σ-Algebra.

Definition:
Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine Funktion $P : \mathcal{A} \to [0,1]$ heißt Wahrscheinlichkeitsmaß, falls

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

(ii) $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, falls $A_i \in \mathcal{A}$, $i = 1,2,3,\ldots$
Das Tripel $(\Omega, \mathcal{A}, P)$ heißt *Wahrscheinlichkeitsraum*.

Lemma 1.1:
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $A, A_1, A_2, \ldots \in \mathcal{A}$.
Dann gilt:

a) $P(\emptyset) = 0$

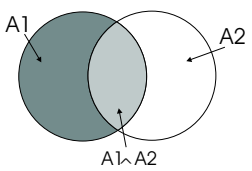
b) $P(A^c) = 1 - P(A)$

c) $A_1 \subset A_2 \Rightarrow P(A_1) \leq P(A_2)$

d) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

e) $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$

Beweis:
a) Setze $A_1 = A_2 = \ldots = \emptyset$ in (ii):
$P(\emptyset) = \sum_{i=1}^{\infty} P(\emptyset) = \lim_{n\to\infty} nP(\emptyset) \Rightarrow P(\emptyset) = 0.$
d)$A_1 \cup A_2 = [A_1 - (A_1 \cap A_2)] + A_2 + \emptyset + \ldots$



$$\left.\begin{array}{rcl} \Rightarrow P(A_1 \cup A_2) & = & P(A_1 - (A_1 \cap A_2)) + P(A_2) + 0 + \ldots \\ P(A_1) & = & P(A_1 - (A_1 \cap A_2)) + P(A_1 \cap A_2) \end{array}\right\} \Rightarrow \text{Beh.}$$

Der klassische Wahrscheinlichkeitsbegriff bezieht sich auf sogenannte Laplacesche Wahrscheinlichkeiträume.

Definition:
Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ heißt *Laplacescher Wahrscheinlichkeitraum*, falls gilt:

(i) Ω ist endlich, d.h. $\Omega = \{\omega_1, \ldots, \omega_n\}$

(ii) $\mathcal{A} = \mathcal{P}(\Omega)$

(iii) $P(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{n}$ für $\omega \in \Omega$, d.h. alle Versuchsausgänge sind gleich wahrscheinlich

Bemerkung:

(i) Für $A \subset \Omega$ gilt $P(A) = P\left(\sum_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}) = \frac{|A|}{|\Omega|}$

(ii) P ist ein Wahrscheinlichkeitsmaß, denn

$$P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$$

$$P\left(\sum_{i=1}^{k} A_i\right) = \frac{|A_1 + \ldots + A_k|}{|\Omega|} = \frac{|A_1|}{|\Omega|} + \ldots + \frac{|A_k|}{|\Omega|} = P(A_1) + \ldots + P(A_k)$$

Hauptproblem hier: Bestimmung der Elementzahl von Teilmengen.

Eine nützliche Verallgemeinerung von Lemma 1.1 d) ist die sogenannte Siebformel von Sylvester - Poincaré:

Satz 1.2: (Siebformel)
Ist $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, so gilt für beliebige $A_1, A_2, \ldots, A_n \in \mathcal{A}$:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik})$$

Für $n = 2$ ergibt sich Lemma 1.1 d).

[2] Permutationen

I. Gegeben: n verschiedene Dinge = Elemente
Eine *Permutation* der Elemente ist eine beliebige Aufreihung der Elemente.

Beispiel: Gegeben: a,b,c
Mögliche Permutationen: abc, acb, bac, bca, cab, cba

Lemma 1.3:
Die Anzahl der Permutationen von n verschiedenen Elementen ist $n! = n(n-1)\ldots 1$.

Beweis: Für den ersten Platz hat man n Möglichkeiten.
Für den zweiten Platz hat man $(n-1)$ Möglichkeiten. etc. □

Beispiel: 10 Personen sitzen an einem langen Tisch.
Wie viele verschiedene Sitzordnungen gibt es? Antwort: $10!$

[2] 24.10.2002

II. Gegeben: n Elemente, die nicht alle verschieden sind.
Es seien n_1 Elemente vom Typ 1, n_2 vom Typ 2, ..., n_k vom Typ k.

$$n_1 + n_2 + \ldots + n_k = n$$

Lemma 1.4: Die Anzahl der Permutationen von n Elementen mit jeweils $n_1, n_2, \ldots, n_k$ gleichen Elementen ist

$$\frac{n!}{n_1! n_2! \ldots n_k!}$$

Beweis: Zunächst nummerieren wir die gleichen Elemente durch, um sie zu unterscheiden. Dann gibt es nach Lemma 1.3 $n!$ mögliche Permutationen. Zu einer Klasse K_i fassen wir die Permutationen zusammen bei denen die Elemente vom Typ i die Plätze tauschen. $\Rightarrow |K_i| = n_i!$
$\Rightarrow$ Behauptung

Beispiele: Eine Urne enthält 10 Kugeln. 5 rote Kugeln, 2 schwarze und drei Blaue. Die Kugeln werden nacheinander gezogen. Wie groß ist die Wahrscheinlichkeit, das zuerst die roten, dann die schwarzen und zuletzt die Blauen Kugeln gezogen werden?
Antwort: $\left(\frac{10!}{5!2!3!}\right)^{-1} = 0,0004$

Kombination = Einfache Urnenmodelle
Kombinationen: Urne mit n Elementen, numeriert $1, 2, \ldots, n$.
$k \leq n$ Elemente werden „zufällig“gezogen.

I **Ziehen mit Zurücklegen, mit Berücksichtigung der Reihenfolge**
Lemma 1.5: Die Anzahl der Kombinationen k-ter Ordnung von n verschiedenen Elementen, die beliebig oft wiederholt werden können mit Berücksichtigung der Reihenfolge ist:

$$n^k$$

Beispiel: Mit den Ziffern $1, 2, \ldots, 9$ kann man $9^3 = 729$ verschiedene 3-stellige Zahlen bilden.$(n = 9, k = 3)$

II **Ziehen ohne Zurücklegen, mit Berücksichtigung der Reihenfolge**
Lemma 1.6: Die Anzahl der Kombinationen k-ter Ordnung von n verschiedenen Elementen mit Berücksichtigung der Reihenfolge ist:

$$\frac{n!}{(n-k)!}$$

Beweis: Beim ersten Zug gibt es n Möglichkeiten.
Beim Zweiten Zug $(n-1)$ Möglichkeiten.
$\vdots$
Beim k-ten Zug gibt es $(n-l+1)$ Möglichkeiten.
Beispiel: Anzahl der 3-stelligen Zahlen mit verschiedenen Ziffern $1, \ldots, 9$ ist $9 \cdot 8 \cdot 7 = 504$

III **Ziehen ohne Zurücklegen, ohne Berücksichtigung der Reihenfolge**
Lemma 1.7: Die Anzahl der Kombinationen k-ter Ordnung von n verschiedenen Elementen ohne Berücksichtigung der Reihenfolge ist:

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

Beweis: Bei Berücksichtigung der Reihenfolge ergeben sich $\frac{n!}{(n-k)!}$ Möglichkeiten. Die k Elemente kann man auf k! verschiedene Weisen anordnen.
$\Rightarrow$ ohne Berücksichtigung der Reihenfolge $\frac{n!}{k!(n-k)!}$ $\square$
Beispiel: Lotto: 6 aus 49. Es gibt 13.983.816 verschiedene Möglichkeiten für die 6 Zahlen.

IV **Ziehen mit Zurücklegen, ohne Berücksichtigung der Reihenfolge**
Lemma 1.8: Die Anzahl der Kombination k-ter Ordnung von n verschiedenen Elementen, die beliebig oft wiederholt werden können, ohne Berücksichtigung der Reihenfolge ist:

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$$

Beweis: Bei III bestimmen wir eigentlich die Anzahl aller verschiedenen k-Tupel natürlicher Zahlen $(n_1, n_2, \ldots, n_k)$ mit

$$1 \leq n_1 < n_2 < \ldots < n_k \leq n \quad (III)$$

Bei IV suchen wir die Anzahl aller verschiedener k-Tupel natürlicher Zahlen $(n_1, n_2, \ldots, n_k)$ mit

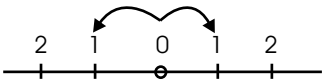
$$1 \leq n_1 \leq n_2 \leq \ldots \leq n_k \leq n \quad (IV)$$

Durch die Abbildung $m_j = n_j + j - 1, \quad j = 1, \ldots, k$ bilden wir (IV) auf (III) ab mit $\hat{n} = n + k - 1$:
Es gilt $1 \leq m_1 < m_2 < \ldots < m_k \leq (n + k - 1)$

n1	n2	n3	n4
1	1	3	7
↓	↓	↓	↓
1	2	5	10
m1	m2	m3	m4

Diese Abbildung ist bijektiv, da durch $n_j = m_j - j + 1$ die Umkehrabbildung gegeben ist.
Die Anzahl der Elemente in den Mengen ist also gleich. $\overset{L.1.7}{\Rightarrow}$ Anzahl $= \binom{n+k-1}{k}$ $\square$

Beispiel:
Von 0 aus kann man auf dem Zahlenstrahl entweder eine Position nach rechts oder eine nach links gehen. Wieviele Zahlen sind nach k Zügen erreichbar?



Antwort: $n = 2, k$
$\binom{k+1}{k}$

Anzahl der Stichproben Umfang k aus $[1, \ldots, n]$	mit Zurücklegen	ohne Zurücklegen
mit Reihenfolge	n^k	$\frac{n!}{(n-k)!}$
ohne Reihenfolge	$\binom{n+k-1}{k}$	$\binom{n}{k}$

[3] Weitere Beispiele:

1. Auf wieviele Arten lassen sich k nicht unterscheidbare Marken auf n Schubladen verteilen?
   Nummeriere die Schubladen $1, \ldots, n$ und reihe die Marken auf.
   Ordne jeder Marke eine Schubladennummer zu.
   Also: Ziehen mit Zurücklegen ohne Reihenfolge. $\binom{n+k+1}{k}$

2. Das Schaufenster eines Elektrogeschäfts soll mit Glühbirnen, die sich nur in ihrer Farbe unterscheiden, dekoriert werden. Es stehen 4 rote, 5 grüne, 3 blaue und 6 gelbe Glühbirnen zur Verfügung.
   Wie viele Anordnungsmöglichkeiten bestehen wenn die Glühbirnen

[3] 31.10.2002

a) in Reihe angeordnet werden und Gleichfarbige beisammen stehen sollen.

b) in beliebiger Reihenfolge in einer Reihe angeordnet werden?

Antwort: a) 4! b) $\frac{18!}{4!\cdot 5!\cdot 3!\cdot 6!} \approx 5,1459 \cdot 10^8$

3. Es sei $S_n =$ Menge aller Permutationen der Zahlen $\{1, 2, \ldots, n\}$
Eine Permutation heisst absolut, falls sie keine einzige Zahl fest läßt.
(z.B. 2413 absolute Permutation, 21$\underline{34}$ *Keine* absolute Permutation)
Abs = Menge aller absoluter Permutationen.
Wie groß ist die Wahrscheinlichkeit, daß eine absolute Permutation auftritt?
$(Abs)^c = \bigcup_{i=1}^{n} A_k$ $A_k =$Menge aller Permutationen,die die Zahl k festlassen.

$$\begin{aligned}
\Rightarrow P(Abs) &= 1 - P((Abs)^c) = 1 - P\left(\bigcup_{i=1}^{n} A_k\right) \\
&= 1 - \sum_{k=1}^{n} (-1)^{k-1} \underbrace{\sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} P(A_{i1} \cap \ldots \cap A_{ik})}_{=:S_k} \\
&= 1 + \sum_{k=1}^{n} (-1)^k \cdot S_k
\end{aligned}$$

$A_{i1} \cdot \ldots \cdot A_{ik} =$ Menge der Permutationen, die die Zahlen $i_1, \ldots, i_k$ festlassen.

$$|A_{i1} \cdot \ldots \cdot A_{ik}| = (n-k)!$$
$$P(A_{i1} \cdot \ldots \cdot A_{ik}) = \frac{(n-1)!}{n!}$$

Es gibt $\binom{n}{k}$ Möglichkeiten $1 \le i_1 < i_2 < \ldots < i_k \le n$ zu erfüllen. (Typ III)

$$\Rightarrow S_k = \binom{n}{k} \cdot \frac{(n-k)!}{n!} = \frac{1}{n!}$$

$$P(Abs) = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

§2 Unabhängigkeit von Ereignissen und bedingte Wahrscheinlichkeit

Im folgenden sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ gegeben. Alle auftretenden Ereignisse seien $\in \mathcal{A}$.

Definition:

a) Zwei Ereignisse A und B heißen *unabhängige Ereignisse* falls gilt:

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

b) n Ereignisse $A_1, \ldots, A_n$ heißen unabhängig, falls für alle $k = 1, \ldots, n$ und für alle k-Tupel $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ stets gilt:

$$P\left(A_{i1} \cdot \ldots \cdot A_{ik}\right) = P\left(A_{i1}\right) \cdot \ldots \cdot P\left(A_{ik}\right)$$

c) unendlich viele Ereignisse heißen unabhängig, falls je endlich viele von ihnen unabhängig sind.

Bemerkung:

(i) In der Regel stimmt der intuitive Begriff der Unabhängigkeit mit dem mathematischen Begriff überein. So sind z.B. beim Ziehen *mit* Zurücklegen die Ereignisse $A_i = \{$Zahl beim i-ten Zug$\}$ $i = 1, \ldots, n$ im Laplace-Experiment unabhängig. Ohne zurücklegen liefert abhängige Ereignisse. Nicht so intuitiv ist das nächste Beispiel.

(ii) Teilsysteme von unabhängigen Ereignissen sind unabhängig.
Beispiel: Werfen eines „fairen“Tetraeders. Eine Seite ist rot, eine grün, eine Blau und eine Seite besitzt alle 3 Farben. Es seien folgende Ereignisse definiert: (1 Wurf)

$$\begin{aligned} A &= \text{„Rot kommt“} \\ B &= \text{„Grün kommt“} \\ C &= \text{„Blau kommt“} \end{aligned}$$

Dann gilt offensichtlich:

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

$$P(AB) = P(AC) = P(BC) = \frac{1}{4}$$

$$P(ABC) = \frac{1}{4}$$

$\Rightarrow$ Die Ereignisse A und B sind unabhängig.
Die Ereignisse A und C sind unabhängig.
Die Ereignisse B und C sind unabhängig.

Jedoch: die Ereignisse A, B und C sind *nicht* unabhängig.

Frage: wie ändert sich die Wahrscheinlichkeit für das Eintreten eines Ereignisses A, wenn zusätzlich bekannt wird, daß das Ereignis B eingetreten ist?
Beispiel: 1 mal Würfeln
A = Es wird eine 1 gewürfelt, B = es wird eine ungerade Zahl gewürfelt.
Definition: Es sei B ein Ereignis mit $P(B) > 0$. Für jedes Ereignis A heißt die Größe $P^*(A) = P(A|B) := \frac{P(AB)}{P(B)}$ die *bedingte Wahrscheinlichkeit von A* (unter der Bedingung B).
Bemerkung: $(B, \mathcal{A}^*, P^*)$, wobei $\mathcal{A}^* = \{A \in \mathcal{A} | A \subseteq B\}$ ist wieder ein Wahrscheinlichkeitsraum.
Beispiel: 1 mal Würfeln (siehe vorher)

$$P(B) = \frac{1}{2}, P(AB) = P(\text{Es wird eine 1 gewürfelt}) = \frac{1}{6}$$

$$\Rightarrow P(A|B) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Oft ist die bedingte Wahrscheinlichkeit $P(A|B)$ gegeben und man muss $P(AB)$ bestimmen. Also $P(AB) = P(A|B) \cdot P(B)$.
Eine Verallgemeinerung hiervon ist
Lemma 2.1: Multiplikationssatz
Es seien $A_1, A_2, \ldots, A_n \in \mathcal{A}$ Ereignisse mit $P(A_1 \cdot \ldots \cdot A_n) > 0$.
Dann gilt:

$$\begin{aligned} P(A_1 \cdot \ldots \cdot A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdot \ldots \cdot P(A_n|A_1 \ldots A_{n-1}) \\ &= \prod_{k=1}^{n} P(A_k|A_1 \ldots A_{k-1}) \end{aligned}$$

Beweis erfolgt durch vollständige Induktion nach n.
[4] **Beispiel:** Von einem Kartenspiel mit 32 Karten bekommt jeder von 3 Spielern 10 Karten. Wie groß ist die Wahrscheinlichkeit, daß jeder der 3 Spieler genau ein As erhält?
A_i = Spieler i erhält genau ein As, $i = 1, 2, 3$.
Gesucht: $P(A_1A_2A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1A_2)$

$$P(A_1) = \frac{\binom{4}{1}\binom{28}{9}}{\binom{32}{10}}, P(A_2|A_1) = \frac{\binom{3}{1}\binom{19}{9}}{\binom{22}{10}}, P(A_3|A_1A_2) = \frac{\binom{2}{1}\binom{10}{9}}{\binom{12}{10}}$$

Satz 2.2: (Satz von der totalen Wahrscheinlichkeit)
Es seien $B_1, B_2, \ldots$ endlich viele oder abzählbar unendlich viele paarweise disjunkte Ereignisse mit

(i) $P(B_i) > 0 \quad \forall\, i = 1, 2, \ldots$

(ii) $\bigcup_{i=1}^{\infty} B_i = \Omega$

d.h. diese (B_i) bilden eine Zerlegung von Ω.
Dann gilt für jedes $A \in \mathcal{A}$:

$$P(A) = \sum_i P(B_i) \cdot P(A|B_i)$$

[4] 07.11.2002

Beweis: $P(A) = P(A \cap \Omega) = P\left(A \cap \left(\bigcup_i B_i\right)\right) = P\left(\bigcup_i (A \cap B_i)\right) = \sum_i P(A \cap B_i) = \sum_i P(B_i)\frac{P(AB_i)}{P(B_i)} = \sum_i P(B_i)P(A|B_i)$ $\square$

Beispiel: Hannover 96 vs. Kaiserslautern
Falls Fredi Bobic eine Trefferquote $\geq 30\%$, dann hat Hannover eine Siegchance von 75%. Ist die Trefferquote $< 30\%$, so hat Hannover eine Siegchance von 40% . Bei Spielen gegen KL ist die Wahrscheinlichkeit, daß F.B. eine Trefferquote von $\geq 30\%$ hat, 70%. Wie groß ist jetzt die Wahrscheinlichkeit, daß H96 gegen KL gewinnt?
Es sei $T =$ F.B. hat Trefferquote von $\geq 30\%$
$T^c =$ F.B. hat Trefferquote von $< 30\%$
$S =$ Sieg für H96

$$\begin{aligned} P(S) &= P(T)P(S|T) + P(T^c)P(S|T^c) \\ &= 0,7 \cdot 0,75 + 0,3 \cdot 0,4 \\ &= 0,645 \end{aligned}$$

Satz 2.3: (Bayes´sche Formel)
Es seien $B_1, B_2, \ldots$ eine endliche oder abzählbar unendliche Zerlegung von Ω wie in Satz 2.2 [Seite 9]. Dann gilt für jedes $A \in \mathcal{A}$ mit $P(A) > 0$:

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_j P(B_j) \cdot P(A|B_j)}$$

Beweis: $P(B_i|A) = \frac{P(B_i A)}{P(A)} = \frac{P(B_i) \cdot P(A|B_i)}{\sum_j P(B_j) \cdot P(A|B_j)}$ $\square$

Beispiel: Bei der binären Übertragung von Nachrichten werden durch Störungen 5% der gesendeten Nullen zu Einsen und umgekehrt 3% der gesendeten Einsen zu Nullen verfälscht. Das Verhältnis der gesendeten Nullen zu den gesendeten Einsen betrage 3:5.
Wie groß ist die Wahrscheinlichkeit, daß eine empfangene Null richtig ist?
Mathematisches Modell: $\Omega = \{(i,j) \mid i,j \in \{0,1\}\}$

1.Komponente: gesendetes Signal

2.Komponente: empfangenes Signal

Ereignisse: S_0: Eine Null wird gesendet $= \{(0,1),(0,0)\}$
S_1: Eine Eins wird gesendet $= \{(1,0),(1,1)\}$
E_0: Eine Null wird empfangen $= \{(0,0),(1,0)\}$
Bekannt sind: $P(E_0|S_0) = 0,95 \quad P(E_0|S_1) = 0,03 \quad \frac{P(S_0)}{P(S_1)} = 0,6$
$\Omega = S_0 + S_1$

$$\begin{aligned} \Rightarrow P(S_0|E_0) &= \frac{P(S_0)P(E_0|S_0)}{P(S_0)P(E_0|S_0) + P(S_1)P(E_0|S_1)} = \frac{1}{1 + \frac{P(S_1)P(E_0|S_1)}{P(S_0)P(E_0|S_0)}} \\ &= \frac{1}{1 + \frac{10}{6} \cdot \frac{0.03}{0.95}} = 0,95 \end{aligned}$$

§3 Zufallsvariablen und deren Verteilung

Bisher haben wir uns für Teilmengen $A \in \Omega$ interessiert bzw. spezielle $\omega \in \Omega$. Oft interessiert aber nicht Ω selbst sondern gewisse Kennzahlen von ω.

Beispiel:

1. $\Omega =$ Menge aller Eintragungen in einem Telefonbuch
   $\omega =$ Familienname, $X(\omega) =$ Anzahl der Buchstaben von ω.
   $\omega =$ Telefonnummer, $X(\omega) =$ Anzahl der Ziffer „1“ in ω.

2. 2 mal Würfeln
   $\Omega = \{\omega = (w_1, w_2) \mid w_i \in \{1, \ldots, 6\}\ i = 1, 2\}$
   Augensumme $\omega \mapsto X(\omega) = w_1 + w_2$
   $\{\omega | X(\omega) = 10\} = \{(6, 4), (5, 5), (4, 6)\}$

Definition: Sei $(\Omega, \mathcal{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum. Die Abbildung $X : \Omega \to \mathbb{R}$ heißt *Zufallsvariable*, falls für jedes $x \in \mathbb{R}$:

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{A}$$

Wir unterscheiden:

1. diskrete Zufallsvariable: X nimmt hier nur abzählbar viele verschiedene Werte an
Beispiel: Augensumme bei 2 mal Würfeln

2. stetige Zufallsvariable: X hat einen überabzählbaren Wertebereich z.B. $[a, b], \mathbb{R}, \mathbb{R}_+$ etc.
Beispiel: Lebensdauer eines Radios.

Ist $\Omega = \mathbb{R}$, so können wir *nicht* $\mathcal{A} = \mathcal{P}(\Omega)$ wählen. Hier wählt man die Borelsche σ-Algebra.
$\mathfrak{B}(\mathbb{R}) = \sigma(\{(a_i, b_i], -\infty \leq a_i < b_i \leq \infty\})$

Verteilungsfunktion und Dichte
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.
Definition: Die Funktion $F_X : \mathbb{R} \to [0, 1]$ sei $F_X(x) := P(X \leq x) = P(\{\omega \mid X(\omega) \leq x\})$ heißt *Verteilungsfunktion*.
Eigenschaften von Verteilungsfunktionen F_X:

1. $F_X(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = 0.$

2. $F_X(\infty) = \lim\limits_{x \to \infty} F_X(x) = 1.$

3. $F_X(x) \leq F_X(x + h) \quad \forall\, x \in \mathbb{R}$ und $h \geq 0$, d.h. F_X ist immer nicht fallend.

4. $F_X(x)$ ist rechtsseitig stetig, d.h. für jede Folge $\{h_n\}$ mit $h_n \geq 0$ und $\lim\limits_{n \to \infty} h_n = 0$ gilt: $\lim\limits_{n \to \infty} F_X(x + h_n) = F_X(x)\ \forall\, x \in \mathbb{R}$

[5] **Bemerkungen:**

(i) Mit Hilfe der Verteilungsfunktion F_X lassen sich folgende Wahrscheinlichkeiten ausdrücken:

$$\begin{aligned} P(a \leq X \leq b) &= P(\{X \leq b\} - \{X < a\}) = P(\{X \leq b\} - \{X \leq a\} + \{X = a\}) \\ &= P(X \leq b) - P(X \leq a) + P(X = a) \\ &= F_X(b) - F_X(a) + P(X = a) \end{aligned}$$

(ii) Sei jetzt X diskret und nehme die Werte $x_1, x_2, \ldots$ an.
Weiter sei $p(x_i) := P(X = x_i)$. Dann gilt:

$$F_X(x) = P(X \leq x) = P\left(\sum_{x_i \leq x} \{X = x_i\}\right) = \sum_{x_i \leq x} \underbrace{P(X = x_i)}_{=p(x_i)}$$

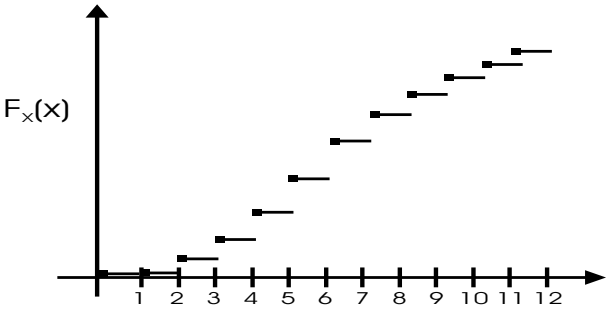
Die Abbildung $x_i \mapsto p(x_i)$ heißt *Zähldichte*.

Beispiel: X = Augensumme bei 2 mal Würfeln.
X ist diskret mit Werten $2, 3, \ldots, 12$.

Zähldichte

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

Verteilungsfunktion



Wichtige diskrete Verteilungsfunktionen / Zähldichten sind:

1. **Binomialverteilung** $B_{n,p}$
Beispiel: Münze wird n-mal geworfen.

$$\Omega = \{(w_1, \ldots, w_n) | w_i \in \{K, Z\}\ i = 1, \ldots, n\}$$

$$X : \Omega \to \{0, 1, \ldots, n\} \text{ mit } X(\Omega) = \text{Anzahl der Kopf-Würfe } = \sum_{i=1}^{n} 1_{w_i = K}$$

Weiter seien die Ereignisse $A_i = \{w \mid w_i = K\},\ i = 1, \ldots, n$ unabhängig und $P(A_i) = p,\ i = 1, \ldots, n$.

$P(X = k) = ?$

$$P(X = k) = P\left(\sum_{1 \leq i_1 < \ldots < i_k \leq n} A_{i_1} \ldots A_{i_k} \cdot \prod_{j \neq i_k} A_j^c\right)$$

[5] 14.11.2002

$$P\left(A_{i_1} \dots A_{i_k} \cdot \prod_{j \neq i_k} A_j^c\right) = p^k \cdot (1-p)^{n-k}$$

$$\Rightarrow P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

X nimmt hier die Werte $0, 1, \dots, n$ an.
Zähldichte: $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ Parameter $n \in \mathbb{N}, p \in (0,1)$

2. **Hypergeometrische Verteilung** $H_{n,m,r}$
Beispiel: Urne mir r roten und s schwarzen Kugeln. $r + s = n$ ($\Rightarrow s = n - r$)
Es werden m Kugeln (ohne Zurücklegen) gezogen.
$\omega =$ Ergebnis der Ziehung
$X(\omega) =$ # der gezogenen roten Kugeln.

$$P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

X nimmt die Werte $k = \max\{0, m - \underbrace{(n-r)}_{=s}\}, \dots, \min\{r, m\}$ an.

Zähldichte: $p(k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$ $r, n, m \in \mathbb{N}$ Parameter

3. **Poisson-Verteilung** Π_λ
X nimmt die Werte $0, 1, 2, \dots$ an.

Zähldichte: $p(k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!},\ k = 0, 1, \dots$

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

Bemerkung: Falls n groß, ist die Berechnung der Verteilungsfunktion der Binomialverteilung schwierig, da viele Summanden aufzusummieren sind.
Dann: Approximation durch Poisson-Verteilung.
Es gilt:

$$\begin{aligned} \text{Sei} \quad F_\lambda(x) &= \text{Verteilungsfunktion von } \Pi_\lambda \\ F_{n,\frac{\lambda}{n}}(x) &= \text{Verteilungsfunktion von } B_{n,\frac{\lambda}{n}} \end{aligned}$$

Dann gilt: $F_{n,\frac{\lambda}{n}}(x) \to F_\lambda(x)$ für $n \to \infty, \forall\, x \in \mathbb{N}_0$

Lemma 3.1: Sei $F_\lambda(x)$ die Verteilungsfunktion der Poisson-Verteilung mit Parameter λ und $F_{n,p}(x)$ sei die Verteilungsfunktion der Binomialverteilung mit Parametern n und p, dann gilt:

$$F_{n,\frac{\lambda}{n}}(x) \to F_\lambda(x) \text{ für } n \to \infty$$

Beweis: Da $F_\lambda(x)$ und $F_{n,p}(x)$ nur bei $x \in \mathbb{N}_0$ springen, genügt es, die Aussage für die Zähldichten zu zeigen, d.h. zu zeigen ist

$$\begin{aligned} p_k &= \binom{n}{k} p^k (1-p)^{n-k} \to \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} \\ p_k &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!}\frac{n!}{(n-k)!n^k}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \\ &= \frac{\lambda^k}{k!}\frac{\overbrace{n(n-1)\cdots(n-k+1)}^{\text{k-Faktoren}\to 1}}{n\cdot n\cdots n}\frac{\overbrace{\left(1-\frac{\lambda}{n}\right)^n}^{\to \mathrm{e}^{-\lambda}}}{\underbrace{\left(1-\frac{\lambda}{n}\right)^k}_{\to 1}} \\ &\to \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \text{ für } n \to \infty \end{aligned}$$

4. **Geometrische Verteilung**
   Beispiel: Wir würfeln bis erstmals eine 6 fällt:
   $X(\omega)$ =Anzahl der Würfe
   $P(X=k) = P$(Wurf 1 bis (k-1) keine 6, dann 6) $= \left(\frac{5}{6}^{k-1}\right)\cdot\frac{1}{6} = p^{k-1}(1-p)$
   mit $p = \frac{5}{6}$
   X nimmt wieder Werte $0, 1, 2, \ldots$ an.
   Zähldichte: $p(k) = (1-p)p^k \quad , k = 0, 1, \ldots$ mit $0 < p < 1$

5. **Negative Binomialverteilung**
   Beispiel: Münzwurf $(k+m)$-mal. Wahrscheinlichkeit für „Zahl“= p. Wie groß ist die Wahrscheinlichkeit im $(k+m)$-ten Wurf zum m-ten mal „Zahl“zu werfen $= p(k)$.

$$p(k) = \binom{k+m-1}{m-1}(1-p)^k p^m = \binom{k+m-1}{k}(1-p)^k p^m$$

   X nimmt die Werte $0, 1, 2, \ldots$ an.

   Zähldichte: $p(k) = \binom{k+m-1}{m-1}p^m(1-p)^k \quad , k = 0, 1, \ldots$ mit $0 < p < 1, m \in \mathbb{N}$ Parameter

[6] Im folgenden betrachten wir stetige Zufallsvariablen:

Definition: Eine Zufallsvariable X heißt *absolut stetige Zufallsvariable,* falls es eine integrierbare Funktion $f : \mathbb{R} \to [0, \infty)$ gibt, so daß die Verteilungsfunktion F_X von X die folgende Darstellung hat:

$$F_X(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y$$

f nennt man *Dichte* von X.

Bemerkung:

[6] 21.11.2002

(i) Es muss gelten: $\int_{-\infty}^{\infty} f(y)\mathrm{d}y = 1,\ f(y) \geq 0,\ \forall\, y \in \mathbb{R}$.

(ii) X absolut stetig $\Rightarrow F_X(x)$ stetig.

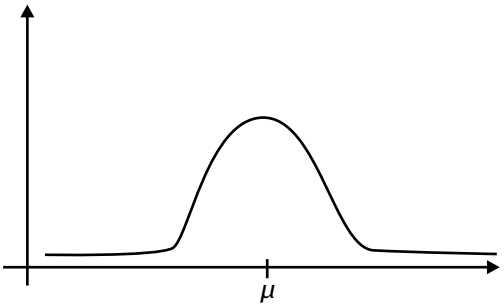
(iii)

$$\begin{aligned} \text{X absolut stetig } \Rightarrow P(X = x) &= \lim_{h \downarrow 0} P(x - h < X \leq x + h) = \\ &= \lim_{h \downarrow 0} (F_X(x + h) - F_X(x - h)) = 0 \end{aligned}$$

Wichtige (absolut) stetige Verteilungsfunktionen / Dichten sind:

1. **Normalverteilung** $\mathcal{N}(\mu, \sigma^2)$

   Parameter $\mu \in \mathbb{R},\ \sigma > 0$.
   Dichte: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$



   Verteilungsfunktion:

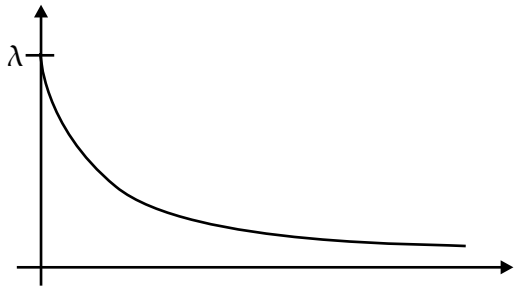
$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \mathrm{d}y$$

   Standardnormalverteilung: $\mu = 0,\ \sigma = 1,$ d.h. $\mathcal{N}(0, 1)$
   Test:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \mathrm{d}y &\overset{\frac{y-\mu}{\sigma} =: u,\ \frac{\mathrm{d}u}{\mathrm{d}y} = \frac{1}{\sigma}}{=} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u = \\ &= \frac{1}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u}_{=\sqrt{2\pi}} = 1 \end{aligned}$$

2. **Exponentialverteilung** Exp(λ)

   Parameter $\lambda > 0$.
   Dichte: $f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & \text{, für } x \geq 0 \\ 0 & \text{, für } x < 0 \end{cases}$



Verteilungsfunktion:

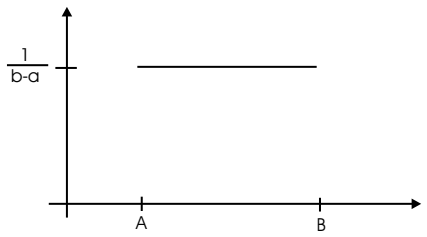
$$\begin{aligned} F_X(x) &= \int_0^x \lambda \mathrm{e}^{-\lambda y} \mathrm{d}y = 1 - \mathrm{e}^{-\lambda x} \quad \text{für } x \geq 0 \\ F_X(x) &= 0, \text{ für } x < 0 \end{aligned}$$

Die Exp(λ)-Verteilung besitz die Eigenschaft der sogenannten „Gedächtnislosigkeit“.
D.h. sei $s < t$

$$\begin{aligned} P(X \geq t \mid X \geq s) &= \frac{P(X \geq t, X \geq s)}{P(X \geq s)} = \frac{P(X \geq t)}{P(X \geq s)} = \frac{1 - F_X(t)}{1 - F_X(s)} \\ &= \frac{\mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda s}} = \mathrm{e}^{-\lambda(t-s)} = P(X \geq t - s) \end{aligned}$$

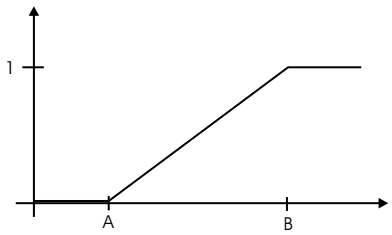
3. **Gleichverteilung** $\mathcal{U}([a,b]) = \mathrm{unif}([a,b])$

Dichte: $f(x) = \begin{cases} \frac{1}{b-a} & \text{, für } a \leq x \leq b \\ 0 & \text{, sonst , } a < b \end{cases}$



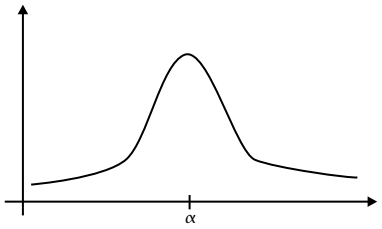
Verteilungsfunktion:

$F_X(x) = \int_a^x \frac{1}{b-a} \mathrm{d}y = \frac{x-a}{b-a}$ für $a \leq x \leq b$



4. **Cauchyverteilung**

Parameter $\alpha \in \mathbb{R},\ \beta > 0$
Dichte: $f(x) = \frac{1}{\pi\beta} \cdot \frac{1}{1+\left(\frac{x-\alpha}{\beta^2}\right)}$



Zufallsvektoren
Mehrere (zufällige) Merkmale sollen jetzt gleichzeitig untersucht werden.
Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

Definition: Es seien $X_1, \ldots, X_n$ Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$.

a) Die Abbildung $X : \Omega \to \mathbb{R}^n,\ X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ heißt (n-dimensionaler) *Zufallsvektor*.

b) Die Abbildung $F_X = F_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$ mit $F_X(\vec{x}) = F_X(x_1,\ldots,x_n) = P(X_1 \leq x_1,\ldots,X_n \leq x_n)$ heißt *Verteilungsfunktion* des Zufallsvektors X oder auch *gemeinsame Verteilung* von $X_1,\ldots,X_n$.

Eigenschaften von (n-dimensionalen) Verteilungsfunktionen F_X.

1. $F_X(x) \to 0$, falls eine Komponente von $\vec{x}$ gegen $-\infty$ geht, während die übrigen Komponenten festgehalten werden.

2. $F_X(x) \to 1$, falls alle Komponenten von $\vec{x}$ gegen ∞ gehen.

3. $F_X(x)$ ist Δ-monoton in $\vec{x}$, d.h.

$$\vec{x}^{(1)} \leq \vec{x}^{(2)} \text{ (Komponentenweise)}$$

$$\Rightarrow \sum_{\{(\varepsilon_1+\ldots+\varepsilon_n) \mid \varepsilon_i \in \{1,2\}\ \forall\, i\}} (-1)^{(\varepsilon_1+\ldots+\varepsilon_n)} F_X(x_1^{(\varepsilon_1)},\ldots,x_n^{(\varepsilon_n)}) \geq 0$$

   z.B. n=2

$$F_X(x_1^{(2)},x_2^{(2)}) - F_X(x_1^{(1)},x_2^{(2)}) - F_X(x_1^{(2)},x_2^{(1)}) + F_X(x_1^{(1)},x_2^{(1)}) \geq 0$$

4. $F_X(\vec{x})$ ist rechtseitig stetig, d.h. $\forall\, \varepsilon >=, \exists\, \delta(\varepsilon,\vec{x}) > 0$ mit $\vec{x} \leq \vec{y} \leq \vec{x} + \delta \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \Rightarrow 0 \leq F_X(\vec{y}) - F_X(\vec{x}) \leq \varepsilon$

Beispiel: Ein Experiment (z.B. Würfeln) hat r mögliche Ausgänge $E_1,\ldots,E_r$ mit jeweiligen Wahrscheinlichkeiten $p_1,\ldots,p_r$, wobei $p_1+\ldots+p_r = 1$.
Dieser Versuch wird n-mal (unabhängig) wiederholt.
Mathematisches Modell:
$\Omega = \{\omega = (\omega_1,\ldots,\omega_n) \mid \omega_i \in \{E_1,\ldots,E_r\}\ \forall\, i\}$
$\mathcal{A} = \mathcal{P}(\Omega)$
Für $i = 1,\ldots,r$ sei $X_i(\omega) = X_i((\omega_1,\ldots,\omega_n)) =$ Anzahl der E_i-Ausgänge $= \sum_{k=1}^{n} 1_{\{\omega_k = E_i\}}$
$P(\{\omega\}) = P(\{(\omega_1,\ldots,\omega_n)\}) = p_1^{X_1(\omega)} \cdot \ldots \cdot p_r^{X_r(\omega)}$
Sei nun $k_1,\ldots,k_r \in \mathbb{N}_0,\ k_1+\ldots+k_r = n$
Gesucht ist $P(X_1 = k_1,\ldots,X_r = k_r) = p_1^{k_1} \cdot \ldots \cdot p_r^{k_r} \cdot$ Anzahl der $(\omega_1,\ldots,\omega_n)$ bei denen k_i Komponenten den Wert E_i haben, $i = 1,\ldots,r$.

$$= p_1^{k_1} \cdot \ldots \cdot p_r^{k_r} \cdot \frac{n!}{k_1! \cdot \ldots \cdot k_r!}$$

$$p_X(k_1,\ldots,k_r) = \begin{cases} \frac{n!}{k_1! \cdot \ldots \cdot k_r!} p_1^{k_1} \cdot \ldots \cdot p_r^{k_r} & \text{, falls } k_j \in \mathbb{N}_0,\ k_1+\ldots+k_r = n \\ 0 & \text{, sonst} \end{cases}$$

ist die Zähldichte der sogenannten *Multinomialverteilung* $\mathcal{M}_{n,r,p_1,\ldots,p_r}$
Bemerkung: Für $n = 2$ erhalten wir die Binomialverteilung.

[7] $X = (X_1,\ldots,X_n)$

$$\begin{aligned} F_X(\vec{x}) &= F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 \leq x_1,\ldots,X_n \leq x_n) = \\ &= P(\{\omega \mid X_1(\omega) \leq x_1\} \cap \{\omega \mid X_2(\omega) \leq x_2\} \cap \ldots \cap \{\omega \mid X_n(\omega) \leq x_n\}) \end{aligned}$$

[7] 28.11.2002

Definition: Wenn man in einer gemeinsamen Verteilungsfunktion $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ bei $(n-1)$ Variablen x_j den Grenzübergang $x_j \to \infty$ ausführt, so heißt die entstehende (1-dimensionale) Verteilungsfunktion eine *Randverteilung*.

Lemma 3.2: Die 1-dimensionalen Randverteilungen einer Multinomialverteilung sind Binomialverteilungen.

Definition:

a) Der Zufallsvektor $X = (X_1,\ldots,X_n)$ heißt *diskreter Zufallsvektor*, falls X nur abzählbar viele verschiedene Werte annimmt. $p_X(x_1,\ldots,x_n) = P(X_1 = x_1,\ldots,X_n = x_n)$ heißt gemeinsame Zähldichte von X.

b) Der Zufallsvektor $X = (X_1,\ldots,X_n)$ heißt *absolut stetiger Zufallsvektor*, falls es eine integrierbare Funktion $f_x : \mathbb{R}^n \to [0,\infty)$ gibt, so daß

$$F_X(x_1,\ldots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_x(y_1,\ldots,y_n)\mathrm{d}y_n,\ldots,\mathrm{d}y_1$$

Bemerkung:

(i) Hier gilt wieder: $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_x(y_1,\ldots,y_n)\mathrm{d}y_n,\ldots,\mathrm{d}y_1 = 1$

(ii) Die Dichte von X_i (sogenannte Randdichte) ist dann

$$f_{X_i}(y_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_x(y_1,\ldots,y_n)\mathrm{d}y_n \ldots \mathrm{d}y_{i+1}\mathrm{d}y_{i-1}\ldots\mathrm{d}y_1$$

Beispiel: $f_{X_1,X_2}(x_1,x_2) = \begin{cases} 4x_1x_2 & \text{, falls } 0 \le x_1,x_2 \le 1 \\ 0 & \text{, sonst} \end{cases}$

Test: $\int_0^1\int_0^1 4x_1x_2\,\mathrm{d}x_2\mathrm{d}x_1 = \int_0^1 2x_1\,[x_2^2]_0^1\,\mathrm{d}x_1 = [x_1^2]_0^1 = 1$

Randdichte $f_{X_1}(x_1) = \int_0^1 4x_1x_2\,\mathrm{d}x_2 = 2x_1$

Beispiel: Der Zufallsvektor (X_1,X_2) heißt standardnormalverteilt $\left(\mathcal{N}(\vec{0},E), \vec{0} = (0,0), E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$, falls (X_1,X_2) absolut stetig ist mit Dichte

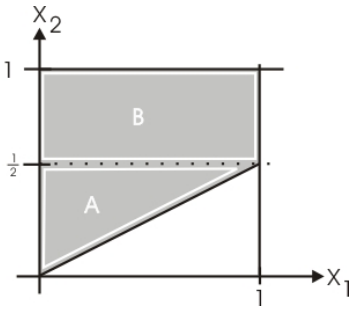
$$f_X(x_1,x_2) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(x_1^2+x_2^2\right)\right)$$

Lemma 3.3:
Sei $X = (X_1,\ldots,X_n)$ ein absolut stetiger Zufallsvektor mit Dichte f_x und $G \in \mathbb{R}^n$ ein Gebiet. Dann gilt:

$$P(X \in G) = \int_G \cdots \int f_X(\vec{x})\mathrm{d}x_n,\ldots,\mathrm{d}x_1$$

Beispiel: siehe Beispiel oben

$$\begin{aligned} P(X_1 \leq 2X_2) &= \int_0^{\frac{1}{2}} \int_0^{2t_2} 4t_1t_2 \, \mathrm{d}t_1\mathrm{d}t_2 + \int_{\frac{1}{2}}^{1} \int_0^{1} 4t_1t_2 \, \mathrm{d}t_1\mathrm{d}t_2 \\ &= \int_0^{\frac{1}{2}} 2t_2 \left[t_1^2\right]_0^{2t_2} \, \mathrm{d}t_2 + \int_{\frac{1}{2}}^{1} 2t_2 \left[t_1^2\right]_0^{1} \, \mathrm{d}t_2 \\ &= 2\left[t_2^4\right]_0^{\frac{1}{2}} + \left[t_2^2\right]_{\frac{1}{2}}^{1} = \frac{7}{8} \end{aligned}$$



Unabhängigkeit von Zufallsvariablen

Definition: Es seien $X_1, \ldots, X_n$ Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. $X_1, \ldots, X_n$ heißen unabhängige Zufallsvariablen, falls

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

für alle $x_1, \ldots, x_n \in \mathbb{R}$.

Bemerkung:

(i) Anders ausgedrückt die Zufallsvariablen $X_1, \ldots, X_n$ sind unabhängig falls

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \ldots \cdot P(X_n \leq x_n).$$

(ii) Ist $(X_1, \ldots, X_n)$ ein diskreter Zufallsvektor mit Werten in $M \subset \mathbb{R}^n$, dann sind $X_1, \ldots, X_n$ genau dann unabhängig falls

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) \; \forall \; (x_1, \ldots, x_n) \in M.$$

(iii) Ist $(X_1, \ldots, X_n)$ ein absolut stetiger Zufallsvektor, dann sind $X_1, \ldots, X_n$ genau dann unabhängig, falls

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \cdots \cdot f_{X_n}(x_n) \; \forall \; (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

(iv) Sind $X_1, \ldots, X_n$ unabhängig, so auch $g_1(X_1), \ldots, g_n(X_n)$ oder auch $h(X_1, X_2), X_3, \ldots, X_n$.

Beispiel: 1 Münzwurf.
$\Omega = \{K, Z\}$
$P(\{K\}) = P(\{Z\}) = \frac{1}{2}$
Definiere 2 Zufallsvariablen X_1 und X_2 wie folgt:
$X_1(K) = 1 \quad , X_1(Z) = -1$
$X_2(K) = -1 \quad , X_2(Z) = 1$
$X_1, X_2 : \Omega \to \{-1, 1\}$
Sind X_1 und X_2 unabhängig?
Nein, denn $P(X_1 = 1, X_2 = 1) = 0 \neq \frac{1}{4} = P(X_1 = 1)P(X_2 = 1)$

Beispiel: Sei (X_1, X_2) ein absolut stetiger Zufallsvektor mit Dichte

$$f_{(X_1,X_2)}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right), 0 \leq \rho < 1$$

Die Dichte von X_1 ist dann gegeben durch (Randdichte):

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x_1,x_2)\mathrm{d}x_2 \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\frac{x_1^2(1-\rho^2)}{1-\rho^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(x_2-\rho x_1)^2}{1-\rho^2}\right) \mathrm{d}x_2 \end{aligned}$$

Setze $u = \frac{x_2-\rho x_1}{\sqrt{1-\rho^2}} \Rightarrow \frac{\mathrm{d}u}{\mathrm{d}x_2} = \frac{1}{\sqrt{1-\rho^2}}$

$$\Rightarrow f_{X_1}(x_1) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}x_1^2\right) \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}u^2\right)\mathrm{d}u}_{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_1^2\right)$$

Wegen Symetrie gilt:
$f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_2^2\right)$
Sind die Zufallsvariablen X_1 und X_2 unabhängig?
Es muss gelten $f_{(X_1,X_2)}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$

$$f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}(x_1^2+x_2^2)\right)$$

Unabhängigkeit liegt also genau dann vor, wenn $\rho = 0$.

[8] **Funktionen von Zufallsvariablen und Zufallsvektoren**

Lemma 3.4: Sei X eine Zufallsvariable und $a, b \in \mathbb{R}$ beliebig, $a \neq 0$.
Dann ist $aX + b$ eine Zufallsvariable und es gilt:

a) $F_{aX+b} = \begin{cases} F_X\left(\frac{x-b}{a}\right) & \text{, falls } a > 0 \\ 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right) & \text{, falls } a < 0 \end{cases}$

b) Ist X absolut stetig mit Dichte f_X, so ist auch $aX + b$ absolut stetig mit Dichte $f_{aX+b}(x) = \frac{1}{|a|}f_x\left(\frac{x-b}{a}\right)$.

Beweis:

a) Sei $a > 0$:

$$F_{aX+b}(x) = P\left(aX + b \leq x\right) = P\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right).$$

Sei $a < 0$:

$$F_{aX+b}(x) = P\left(aX + b \leq x\right) = P\left(X \geq \frac{x-b}{a}\right) = 1 - \underbrace{P\left(X \leq \frac{x-b}{a}\right)}_{F_X\left(\frac{x-b}{a}\right)} + P\left(X = \frac{x-b}{a}\right).$$

[8] 05.12.2002

b) Sei $x \in \mathbb{R}$. Z.z.: $\int_{-\infty}^{x} \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right) \mathrm{d}t \stackrel{!}{=} F_{aX+b}\left(\frac{x-b}{a}\right)$

Sei $a > 0$:

$$\frac{1}{a} \int_{-\infty}^{x} f_X\left(\frac{t-b}{a}\right) \mathrm{d}t = \int_{-\infty}^{\frac{x-b}{a}} f_X(u) \mathrm{d}u = F_X\left(\frac{x-b}{a}\right)$$

mit $\frac{t-b}{a} =: u \quad \frac{\mathrm{d}u}{\mathrm{d}t} = \frac{1}{a}$

Analog $a < 0$. □

Beispiel: Sei $X \sim \mathcal{N}(0,1)$, d.h. X hat die Dichte $f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$.
Welche Dichte hat $\sigma X + \mu$ mit $\sigma > 0$, $\mu \in \mathbb{R}$?
Lemma 3.4 $\Rightarrow f_{\sigma X+\mu}(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
d.h. $\sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$

Lemma 3.5:

a) Sei $X = (X_1, X_2)$ ein absolut stetiger Zufallsvektor mit gemeinsamer Dichte $f_X(x_1, x_2)$. Dann ist $Z = X_1 + X_2$ eine Zufallsvariable mit Dichte

$$f_Z(x) = \int_{-\infty}^{\infty} f_X(t, x-t) \mathrm{d}t$$

Falls X_1 und X_2 unabhängig sind, gilt:

$$f_Z(x) = \int_{-\infty}^{\infty} f_{X_1}(t) f_{X_2}(x-t) \mathrm{d}t \quad \text{Faltungsformel}$$

b) Sei $X = (X_1, X_2)$ ein diskreter Zufallsvektor mit Werten in $\mathbb{N}_0^2$.
Dann ist $Z := X_1 + X_2$ eine diskrete Zufallsvariable mit

$$P(Z = k) = \sum_{i=0}^{k} P(X_1 = i, X_2 = k - i)$$

Falls X_1 und X_2 unabhängig sind, gilt:

$$P(Z = k) = \sum_{i=0}^{k} P(X_1 = i) P(X_2 = k - i)$$

Beispiel: Es sei $X_1 \sim \Pi(\lambda_1)$, $X_2 \sim \Pi(\lambda_2)$ Poisson verteilt, X_1 und X_2 unabhängig.
Welche Zähldichte hat $X_1 + X_2$?

$$\begin{aligned} P(X_1 + X_2 = k) &= \sum_{i=0}^{k} P(X_1 = i) P(X_2 = k - i) = \sum_{i=0}^{k} \mathrm{e}^{-\lambda_1} \frac{\lambda_1^i}{i!} \cdot \mathrm{e}^{-\lambda_2} \frac{\lambda_2^{k-i}}{(k-i)!} = \\ &= \mathrm{e}^{-(\lambda_1+\lambda_2)} \frac{1}{k!} \underbrace{\sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i}}_{(\lambda_1+\lambda_2)^k} = \mathrm{e}^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^k}{k!} \end{aligned}$$

Also $X_1 + X_2 \sim \Pi(\lambda_1 + \lambda_2)$

Satz 3.5: (Dichtetransformation)
Es sei also $X = (X_1, \ldots, X_n)$ ein absolut stetiger Zufallsvektor mit Dichte $f_X(x_1, \ldots, x_n)$ und Werten in $S \subset \mathbb{R}^n$. $g = (g_1, \ldots, g_n)$ sei eine bijektive Abbildung von S nach $g(S) \subset \mathbb{R}^n$ mit stetigen partiellen Ableitungen und es gelte:

$$J_g(\vec{t}) := \det \begin{pmatrix} \frac{\partial}{\partial t_1} g_1(\vec{t}) & \ldots & \frac{\partial}{\partial t_n} g_1(\vec{t}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial t_1} g_n(\vec{t}) & \ldots & \frac{\partial}{\partial t_n} g_n(\vec{t}) \end{pmatrix} \neq 0 \quad \forall\, \vec{t} \in S$$

dann lautet die Dichte von $Y = g(x)$:

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot |J_{g^{-1}}(y)| \quad , y \in g(S)$$

Bemerkung: Es gilt $J_{g^{-1}}(y) = \frac{1}{J_g(g^{-1}(y))}$.

Beispiel: Erzeugung von $\mathcal{N}(0,1)$-verteilten Zufallsvariablen mit der Box-Muller-Methode.
Seien $U_1, U_2 \sim \mathcal{U}(0,1)$ und unabhängig. Dann sind

$$\begin{aligned} X_1 &:= \sqrt{-2\log(U_1)} \cdot \cos(2\pi U_2) \\ X_1 &:= \sqrt{-2\log(U_1)} \cdot \sin(2\pi U_2) \end{aligned}$$

unabhängig und $X_1, X_2 \sim \mathcal{N}(0,1)$.
Beweis: Es sei $\Theta := 2\pi U_2 \Rightarrow \Theta \sim \mathcal{U}(0, 2\pi)$

$$\frac{1}{2}V^2 := -\log(U_1) \Rightarrow \frac{1}{2}V^2 \sim \exp(1) \quad \text{vgl. Übungsaufgabe 41}$$

$$\begin{aligned} \text{Also:} \quad X_1 &= V \cdot \cos\Theta = g_1(V, \Theta) \\ X_2 &= V \cdot \sin\Theta = g_2(V, \Theta) \end{aligned}$$

Welche Dichte hat $f_{V,\Theta}(v, \vartheta)$?

$$\begin{aligned} P(\frac{1}{2}V^2 \leq x) = 1 - \mathrm{e}^{-x} \quad &\Leftrightarrow \quad P\left(V \leq \sqrt{2x}\right) = 1 - \mathrm{e}^{-x} \\ &\Rightarrow \quad P\left(V \leq y\right) = 1 - \mathrm{e}^{-\frac{1}{2}y^2} \end{aligned}$$

$\Rightarrow$ Dichte von V : $f_V(v) = v\mathrm{e}^{-\frac{v^2}{2}}$
Wegen der Unabhängigkeit von Θ und V ist die gemeinsame Dichte

$$f_{V,\Theta}(v, \vartheta) = \begin{cases} \frac{1}{2\pi} \cdot \mathrm{e}^{-\frac{v^2}{2}} \cdot v & \text{, für } v \geq 0,\ 0 \leq \theta \leq 2\pi \\ 0 & \text{, sonst} \end{cases}$$

$$\frac{\partial g_1}{\partial v} = \cos(\vartheta), \frac{\partial g_1}{\partial \vartheta} = v\sin(\vartheta), \frac{\partial g_2}{\partial v} = \sin(\vartheta), \frac{\partial g_2}{\partial \vartheta} = v\cos(\vartheta)$$

$$\Rightarrow J_g(t) = \det \begin{pmatrix} \cos(\vartheta) & -v\sin(\vartheta) \\ \sin(\vartheta) & v\cos(\vartheta) \end{pmatrix} = v\cos^2(\vartheta) + v\sin^2(\vartheta) = v$$

und $v = g^{-1}(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$
Nach Satz 3.5 gilt also:

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= f_{V,\Theta}(\sqrt{x_1^2 + x_2^2}) \cdot \frac{1}{\sqrt{x_1^2 + x_2^2}} = \\ &= \frac{1}{2\pi} \cdot \mathrm{e}^{-\frac{1}{2}(x_1^2 + x_2^2)} \cdot \frac{\sqrt{x_1^2 + x_2^2}}{\sqrt{x_1^2 + x_2^2}} = \\ &= \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{1}{2}x_1^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{1}{2}x_2^2} \end{aligned}$$

§4 Charakteristiken von Zufallsvariablen und Zufallsvektoren

4.1 Erwartungswert und Momente von Zufallsvariablen

Definition:

a) Sei X eine diskrete Zufallsvariable mit Werten $x_1, x_2, \ldots$ und einer Zähldichte $p(x_1), p(x_2), \ldots$.Dann heißt

$$E(X) := \sum_{i=1}^{\infty} x_i \cdot p(x_i)$$

Erwartungswert von X, falls $\sum_{i=1}^{\infty} |x_i| \cdot p(x_i) < \infty$

b) Sei X eine absolut stetige Zufallsvariable mit Dichte $f_X(x)$.Dann heißt

$$E(X) := \int_{-\infty}^{\infty} x \cdot f_X(x)\mathrm{d}x$$

Erwartungswert von X, falls $\int_{-\infty}^{\infty} |x| \cdot f_X(x)\mathrm{d}x < \infty$

[9] **Bemerkungen:**

(i) Manchmal sagt man statt Erwartungswert auch Mittelwert der Verteilung.

(ii) Später werden wir sehen, daß $E(X)$ den langfristigen, durchschnittlichen Ausgang des Zufallsexperiments X beschreibt. D.h. $\frac{1}{n}(X_1, \ldots, X_n) \approx E(X)$ für n groß, falls $X_1, \ldots, X_n$ unabhängige Wiederholungen von X sind.

Beispiel: 1 mal Würfeln
$\Omega = \{1, \ldots, 6\},\ X(\omega) = \omega$
Zähldichte: $p(1) = \ldots = p(6) = \frac{1}{6}$
Erwartungswert: $E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = 3,5$
Beispiel: $X \sim \mathcal{N}(\mu, \sigma^2)$
Dichte: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$
$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \mathrm{d}x$
Substitution: $\frac{x-\mu}{\sigma} =: u \quad \Rightarrow \frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{\sigma}$

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma u + \mu) \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u \\ &= \frac{1}{\sqrt{2\pi}} \underbrace{\left(\int_{-\infty}^{\infty} \sigma u \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u\right)}_{=0\text{(Integrand punktsymetrisch)}} + \mu \underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u}_{=\sqrt{2\pi}} = \mu \end{aligned}$$

[9] 12.12.2002

Beispiel: Sei $X \sim B_{n,p}$ binomialverteilt
Zähldichte $p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = np \cdot \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = \\ &= np \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}}_{=1} = np \end{aligned}$$

Lemma 4.1:

a) Ist X eine diskrete Zufallsvariable mit Werten $x_1, x_2, \ldots$ und Zähldichte $p(x_1), p(x_2), \ldots$ und $\varphi : \{x_1, x_2, \ldots\} \to \mathbb{R}$. Dann ist $E(\varphi(x)) = \sum_{i=1}^{\infty} \varphi(x_i) p(x_i)$, falls $E(.)$ existiert.

b) Ist X eine absolut stetige Zufallsvariable mit Dichte $f_X(x)$ und $\varphi : \mathbb{R} \to \mathbb{R}$. Dann ist $E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f_X(x) \mathrm{d}x$, falls $E(.)$ existiert.

Definition: Sei X eine Zufallsvariable.

a) Ist $k \in \mathbb{N}$ und existiert $E(|X|^k)$, dann heißt

$$\mu_k := E(X^k)$$

k-tes Moment von X

b) Ist $k \in \mathbb{N}$ und existiert $E(|X|^k)$, dann heißt

$$\bar{\mu}_k := E((X - \mu_1)^k)$$

k-tes zentriertes Moment von X.

c) $\bar{\mu}_2 = E((X - \mu_1)^2)$ heißt auch *Varianz* von X.
Schreibweise $\mathrm{Var}(x) = \bar{\mu}_2$
$\sigma_X := \sqrt{\mathrm{Var}(x)}$ heißt *Standardabweichung* von X.

Beispiel: 1 mal Würfeln, X = Augenzahl

$$\mathrm{Var}(X) = E((X - 3,5)^2) = \sum_{i=1}^{6} (i - 3,5)^2 \cdot \frac{1}{6} = \frac{35}{12}$$

Beispiel: $X \sim \mathcal{N}(\mu, \sigma^2)$
$E(X) = \mu$ und
$\mathrm{Var}(X) = E\left((X-\mu)^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \mathrm{d}x$
Mit Substitution: $\frac{x-\mu}{\sigma} =: u$
$\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 u^2 \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u = \frac{1}{\sqrt{2\pi}} \cdot 2 \cdot \int_{0}^{\infty} \sigma^2 u^2 \exp\left(-\frac{1}{2}u^2\right) \mathrm{d}u$

Mit Substitution: $\frac{1}{2}u^2 = t \quad \frac{\mathrm{d}t}{\mathrm{d}u} = u \quad u = \sqrt{2t}$

$$\text{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}} \cdot 2 \cdot \int_0^\infty 2t \exp(-t) \frac{1}{\sqrt{2t}} \mathrm{d}t = \frac{\sigma^2}{\sqrt{2\pi}} \cdot 2 \cdot \sqrt{2} \underbrace{\int_0^\infty \sqrt{t} \exp(-t) \mathrm{d}t}_{\Gamma\left(\frac{3}{2}\right) = \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}} = \sigma^2$$

Bemerkung: Die Varianz bzw. die Standardabweichung ist ein Maß dafür, wie weit die Zufallsvariable um den Erwartungswert herum streut.

Lemma 4.2: Sei X eine Zufallsvariable mit $E(|X|^2) < \infty$

a) Dann existiert $\text{Var}(X)$ und es gilt:

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

b) Für $a, b \in \mathbb{R}$ gilt:

$$E(aX + b) = aE(X) + b$$
$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Beweis:

b) (i) Sei X diskret.

$$\begin{aligned} E(aX+b) &= \sum_{i=1}^\infty (ax_i + b) p(x_i) = a \cdot \underbrace{\sum_{i=1}^\infty x_i \cdot p_i(x_i)}_{=E(X)} + b \cdot \underbrace{\sum_{i=1}^\infty p(x_i)}_{=1} \\ &= aE(X) + b \end{aligned}$$

(ii) Sei X stetig.

$$\begin{aligned} E(aX+b) &= \int_{-\infty}^\infty x f_{aX+b}(x) \mathrm{d}x = \int_{-\infty}^\infty x \cdot \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right) \mathrm{d}x \\ &\quad \text{mit Substitution:} \quad \frac{x-b}{a} =: u \quad , \frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{a} \\ &= \int_{-\infty}^\infty (au + b) f_X(u) \mathrm{d}u \\ &= a \underbrace{\int_{-\infty}^\infty u f_X(u) \mathrm{d}u}_{=E(X)} + b \underbrace{\int_{-\infty}^\infty f_X(u) \mathrm{d}u}_{=1} = aE(X) + b \end{aligned}$$

a)

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

b)

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

□

4.2 Charakteristiken mehrdimensionaler Verteilungen

Definition:

a) Sei $X = (X_1, \ldots, X_n)$ ein diskreter Zufallsvektor mit Werten $x_1, x_2, \ldots \in \mathbb{R}^n$ und Zähldichte $p_1(x)_1, p_2(x_2), \ldots$.Weiter sei $\varphi : \mathbb{R}^n \to \mathbb{R}$. Dann heißt

$$E(\varphi(\vec{x})) := \sum_{i=1}^{\infty} \varphi(x_i)p(x_i)$$

Erwartungswert von $\varphi(\vec{x})$, falls $\sum\limits_{i=1}^{\infty} |\varphi(x_i)|p(x_i) < \infty$

b) Sei $X = (X_1, \ldots, X_n)$ ein absolut stetiger Zufallsvektor mit Dichte $f_X(x_1, \ldots, x_n)$. Weiter sei $\varphi : \mathbb{R}^n \to \mathbb{R}$. Dann heißt

$$E(\varphi(\vec{x})) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \varphi(x_1, \ldots, x_n) f_X(x_1, \ldots, x_n) \mathrm{d}x_1, \ldots, \mathrm{d}x_n$$

Erwartungswert von $\varphi(\vec{x})$, falls $E(\varphi(\vec{x})) = \int \ldots \int |\varphi(\vec{x})| f_X(\vec{x}) \mathrm{d}x_1, \ldots, \mathrm{d}x_n < \infty$

[10] **Lemma 4.3:** Seien $X_1, \ldots, X_n$ Zufallsvariablen, deren Erwartungswerte existieren.

a) Dann existiert auch der Erwartungswert von $X_1 + \ldots + X_n$ und es gilt:

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$$

b) Sind $X_1, \ldots, X_n$ *unabhängig*, dann existiert auch der Erwartungswert von $X_1 \cdot \ldots \cdot X_n$ und es gilt:

$$E(X_1 \cdot \ldots \cdot X_n) = E(X_1) \cdot \ldots \cdot E(X_n)$$

Beweis: o.B.d.A. n=2.
Wir betrachten nur den Fall $X = (X_1, X_2)$ ist diskret. X nehme die Werte $\left(x_1^{(i)}, x_2^{(j)}\right)$ $i, j \in \mathbb{N}$ mit Wahrscheinlichkeit P_{ij} an.

a) $$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \underbrace{\left|x_1^{(i)} + x_2^{(j)}\right|}_{\leq\left(\left|x_1^{(i)}\right|+\left|x_2^{(j)}\right|\right)} P_{ij} \leq \sum_{i=1}^{\infty} \left|x_1^{(i)}\right| \underbrace{\sum_{j=1}^{\infty} P_{ij}}_{P\left(X_1=x_1^{(i)}\right)} + \sum_{j=1}^{\infty} \left|x_2^{(j)}\right| \underbrace{\sum_{i=1}^{\infty} P_{ij}}_{P\left(X_2=x_2^{(j)}\right)} \overset{\text{n. Vor.}}{<} \infty$$

$\Rightarrow$ Erwartungswert existiert.

[10] 19.12.2002

$$\begin{aligned} E(X_1+X_2) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\left(x_1^{(i)}+x_2^{(j)}\right)P_{ij} \\ &\overset{\text{s.o.}}{=} \sum_{i=1}^{\infty}\left(x_1^{(i)}\right)P\left(X_1=x_1^{(i)}\right)+\sum_{j=1}^{\infty}\left(x_2^{(j)}\right)P\left(X_2=x_2^{(j)}\right) \\ &= E(X_1)+E(X_2) \end{aligned}$$

b) Erwartungswert existiert: Nachweis wie in a).

$$\begin{aligned} E(X_1\cdot X_2) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}x_1^{(i)}x_2^{(j)}P_{ij}=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}x_1^{(i)}x_2^{(j)}\cdot P\left(X_1=x_1^{(i)},X_2=x_2^{(j)}\right) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}x_1^{(i)}x_2^{(j)}\cdot P\left(X_1=x_1^{(i)}\right)\cdot P\left(X_2=x_2^{(j)}\right) \\ &= \sum_{i=1}^{\infty}x_1^{(i)}P\left(X_1=x_1^{(i)}\right)\cdot\sum_{j=1}^{\infty}x_2^{(j)}P\left(X_2=x_2^{(j)}\right)=E(X_1)E(X_2) \end{aligned}$$

Beispiel: Sei $X\sim B_{n,p}$ binomialverteilt.
$\Rightarrow X=X_1+\ldots+X_n$ wobei $X_i\sim B(1,p)\quad i=1,\ldots,n$ und $X_1,\ldots,X_n$ unabhängig.
$\Rightarrow E(X)=\sum\limits_{i=1}^{n}E(X_i)=n\cdot E(X_1)=n\cdot p$

Definition: Seien X_1,X_2 Zufallsvariablen mit $\mu_1=E(X_1),\mu_2=E(X_2)$.

a) Die *Kovarianz* von X_1 und X_2 ist definiert durch

$$\text{Cov}(X_1,X_2)=E((X_1-\mu_1)(X_2-\mu_2))=E(X_1X_2)-\mu_1\mu_2$$

falls der Erwartungswert existiert.

b) Der *Korrelationskoeffizient* von X_1 und X_2 ist definiert durch

$$\rho_{X_1,X_2}=\frac{\text{Cov}(X_1,X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}}$$

falls alle auftretenden Erwartungswerte existieren.

c) X_1 und X_2 heißen *unkorreliert*, falls $\rho_{X_1,X_2}=0$.

Bemerkungen:

(i) Es gilt stets $-1\le\rho_{X_1,X_2}\le 1$.

(ii) Sind X_1,X_2 unabhängig, so sind sie auch unkorreliert, da

$$\text{Cov}(X_1,X_2)=E\left((X_1-\mu_1)(X_2-\mu_2)\right)=\underbrace{E(X_1-\mu_1)}_{=0}\cdot\underbrace{E(X_2-\mu_2)}_{=0}=0$$

Die Umkehrung gilt i.a. nicht (siehe Beispiel unten)

(iii) Der Korrelationskoeffizient ist ein Maß für die lineare Abhängigkeit der zwei Zufallsvariablen.
Extremfälle: $X_2 = aX_1 + b \quad , \ a \neq 0 \ , \ b \in \mathbb{R}$.

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E((X_1 - \mu_1)(aX_1 + b - a\mu_1 - b)) = aE((X_1 - \mu_1)^2) \\ &= a\text{Var}(X_1) \end{aligned}$$

$$\begin{aligned} \rho_{X_1,X_2} = \frac{a\text{Var}(X_1)}{\sqrt{\text{Var}(X_1) \cdot a^2 \text{Var}(X_1)}} &= 1 \quad \text{, falls } a > 0 \text{ positiv korreliert} \\ &= -1 \quad \text{, falls } a < 0 \text{ negativ korreliert} \end{aligned}$$

Beispiel: X, Y unkorreliert, aber stochastisch abhängig.
$\Omega = \{\omega_1, \omega_2, \omega_3\}$

ω	$P(\omega)$	$X(\omega)$	$Y(\omega)$	$X(\omega) \cdot Y(\omega)$
ω_1	$\frac{1}{3}$	0	-1	0
ω_2	$\frac{1}{3}$	1	0	0
ω_3	$\frac{1}{3}$	0	1	0

$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0 - \frac{1}{3} \cdot \left(-\frac{1}{3} + \frac{1}{3}\right) = 0$
Also X,Y unkorreliert.

Aber: $\left.\begin{array}{rcl} P(X \leq 0) &=& P(\{\omega_1, \omega_3\}) = \frac{2}{3} \\ P(Y \leq 0) &=& P(\{\omega_1, \omega_2\}) = \frac{2}{3} \end{array}\right\} \Rightarrow P(X \leq 0, Y \leq 0) \neq P(X \leq 0)P(Y \leq 0)$

d.h. X, Y sind stochastisch abhängig.

Lemma 4.4: $X_1, \ldots, X_n$ Zufallsvariablen mit $E(X_i^2) < \infty \quad , \ i = 1, \ldots, n$

a) $\text{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k)$

b) Falls $X_1, \ldots, X_n$ paarweise unkorreliert (oder unabhängig), so gilt:

$$\text{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i)$$

Beweis:

a)

$$\begin{aligned} \text{Var}(X_1 + \ldots + X_n) &= E\left(\left(\sum_{i=1}^{n} X_i - E(X_i)\right)^2\right) \\ &= E\left(\left(\sum_{i=1}^{n} X_i - E(X_i)\right) \cdot \left(\sum_{j=1}^{n} X_j - E(X_j)\right)\right) \\ &= E\left[\sum_{1 \leq i,j \leq n} (X_i - E(X_i))\,(X_j - E(X_j))\right] \\ &= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \cdot \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \end{aligned}$$

b) Hier gilt $\text{Cov}(X_i, X_j) = 0$ und b) folgt aus a).

□

Definition: Sei $X = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor.

a) Falls $E(X_1), \ldots, E(X_n)$ existieren, so heißt

$$\vec{\mu}_X = (E(X_1), \ldots, E(X_n))$$

der *Erwartungswertvektor* von X.

b) Falls $E(X_1^2), \ldots, E(X_n^2)$ existieren, dann heißt

$$\text{Cov}(X) = \text{Cov}((X_1, \ldots, X_n)) = \begin{pmatrix} \text{Cov}(X_1, X_1) & \ldots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \ldots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

Kovarianzmatrix von X.

Beispiel: (2-dimensionale Normalverteilung)
Sei $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$ und $C = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$ mit $\sigma_1, \sigma_2 > 0\ ,\ 0 \leq \rho \leq 1$
Die Dichte einer 2-dimensionalen Normalverteilung ist gegeben durch

$$f_X(x_1, x_2) = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right)$$

Man schreibt auch $X = (X_1, X_2) \sim \mathcal{N}(\mu, C)$. Dann gilt:

Randverteilung:	$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$
	$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
Erwartungswertvektor:	$\vec{\mu}_x = (\mu_1, \mu_2)$
Kovarianzmatrix:	$\text{Cov}(X) = C$.

§5 Abschätzung für Erwartungswerte und Wahrscheinlichkeiten

[11]**Satz 5.1:** (Ungleichung von Cauchy-Schwarz)
Seien X und Y zwei Zufallsvariablen mit $E[X^2]$, $E[Y^2] < \infty$.
Dann gilt:

$$|E[XY]|^2 \leq E[X^2] \cdot E[Y^2]$$

Anwendung: Es gilt stets: $|\rho_{X,Y}| = \left| \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \right| \leq 1$
oder umgeformt: $(\text{Cov}(X,Y))^2 \leq \text{Var}(X)\text{Var}(Y)$

$$\begin{aligned}(\text{Cov}(X,Y))^2 &= (E[\underbrace{(X-E(X))}_{=\tilde{X}}\underbrace{(Y-E(Y))}_{=\tilde{Y}}])^2 \leq E\left[(X-E(X))^2\right] \cdot E\left[(Y-E(Y))^2\right] \\ &= \text{Var}(X) \cdot \text{Var}(Y)\end{aligned}$$

Satz 5.2: (Ungleichung von Chebyshev)
Sei X eine Zufallsvariable mit $E[|X|] < \infty$ und $\varepsilon > 0$ beliebig.
Dann gilt:

$$P(|X| \geq \varepsilon) \leq \frac{E(|X|)}{\varepsilon}$$

Bemerkung:

(i) Wende Satz 5.2 an auf $|X|^m, \varepsilon^m$ mit $m \in \mathbb{N}$

$$P(|X|^m \geq \varepsilon^m) \leq \frac{E(|X|^m)}{\varepsilon^m}$$

$$\Rightarrow P(|X| \geq \varepsilon) \leq \frac{E(|X|^m)}{\varepsilon^m} \quad \text{mit } m \in \mathbb{N}$$

Diese Version nennt man auch *Markovsche Ungleichung*.

(ii) Ist $\text{Var}(X) < \infty$, so folgt aus (i) mit $m = 2$:

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

Beweis von Satz 5.2: Wir betrachten nur den Fall, daß X eine Dichte $f_X(x)$ hat:

$$\begin{aligned}E(|X|) &= \int_{-\infty}^{\infty} |x| f_X(x)\mathrm{d}x \geq \int_{-\infty}^{-\varepsilon} |x| f_X(x)\mathrm{d}x + \int_{\varepsilon}^{\infty} |x| f_X(x)\mathrm{d}x \\ &\geq \varepsilon \left(\int_{-\infty}^{-\varepsilon} f_X(x)\mathrm{d}x + \int_{\varepsilon}^{\infty} f_X(x)\mathrm{d}x \right) \\ &= \varepsilon \left(P(X \leq -\varepsilon) + P(X \geq \varepsilon) \right) \\ &= \varepsilon P(|X| \geq \varepsilon) \qquad \square\end{aligned}$$

[11]09.01.2003

Beispiel: Wir würfeln n mal. Wie groß ist die Wahrscheinlichkeit, daß die Zahl, die im Durchschnitt gewürfelt wurde um mehr als 10% von 3,5 abweicht?
Seien $X_1, \ldots, X_n$ die Augenzahlen der n Versuche, d.h.
$P(X_i = k) = \frac{1}{6}$ für $i = 1, \ldots, n \quad , k = 1, \ldots, 6$ und die $X_1, \ldots, X_n$ sind unabhängig.
Gesucht ist $P(|\frac{1}{n} \sum_{i=1}^{n} X_i - 3,5| \geq 0,35)$!
Nach Bemerkung (ii) ist diese Wahrscheinlickeit

$$P(.) \leq \frac{\text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)}{(0,35)^2} = \frac{\frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right)}{(0,35)^2} = \frac{\frac{1}{n}\text{Var}(X_i)}{(0,35)^2} = \frac{100^2}{n \cdot 12 \cdot 35}$$

z.B. $n = 100$: $= 0,238$
Weitere Anwendung von Satz 5.2:
Var$(X) = 0 \Leftrightarrow X = \mu$ für ein $\mu \in \mathbb{R}$.
„$\Rightarrow$“ $P(|X - E(X)| \geq \varepsilon) \leq 0$, also $= 0 \quad \forall\, \varepsilon > 0$
Da ε beliebig, muss $X \equiv \mu$ sein.
„$\Leftarrow$“ Var$(X) = E(X^2) - (E(X))^2 = \mu^2 - \mu^2 = 0$

§6 Erzeugende Funktionen

Im folgenden betrachten wir nur Zufallsvariablen mit Werten in $\mathbb{N}_0$.

Definition: Sei X eine Zufallsvariable mit Werten in $\mathbb{N}_0$ und Zähldichte $p(k)$, $k \in \mathbb{N}_0$, d.h. $p(k) = P(X = k)$. Dann heißt die Funktion $g_X : [-1, 1] \to \mathbb{R}$ mit

$$g_X(z) := \sum_{k=0}^{\infty} p_k \cdot z^k = E[z^X]$$

erzeugende Funktion von X.

Bemerkungen:

(i) $g_X(z)$ ist wohldefiniert für $|z| \leq 1$, weil $\left|\sum_{k=0}^{\infty} p_k z^k\right| \leq \sum_{k=0}^{\infty} p_k \left|z^k\right| \leq \sum_{k=0}^{\infty} p_k = 1$

(ii) Insbesondere gilt: $g_X(1) = 1$

(iii) $g_X^{(n)}(z)$ ist wohldefiniert auf $(-1, 1)$, $n \in \mathbb{N}$.

(iv) Wir schreiben $g_X^{(n)}(z-)$ und meinen $\lim\limits_{t \uparrow z} g_X^{(n)}(t)$

(v) Es gilt: $P(X = k) = p(k) = \frac{g_X^{(k)}(0)}{k!}$

Satz 6.1: Besitzt die Zufallsvariable X die erzeugende Funktion $g_X(z)$ so gilt:

a) $E(X) = g_X'(1-)$, falls $E(X)$ existiert.

b) $\text{Var}(X) = g_X''(1-) + g_X'(1-) - (g_X'(1-))^2$, falls $\text{Var}(X)$ existiert.

Beweis:

a) $g_X'(z) = \sum\limits_{k=0}^{\infty} p_k \cdot k \cdot z^{k-1} \quad \to \sum\limits_{k=0}^{\infty} p_k \cdot k = E(X)$ für $z \to 1$.

b) ähnlich $\square$

Beispiel: Sei $X \sim \text{Bin}(n, p)$
Also $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ für $k = 0, 1, \ldots, n$

$$g_X(z) = \sum_{k=0}^{n} \underbrace{\binom{n}{k} p^k (1-p)^{n-k} \cdot z^k}_{\binom{n}{k}(pz)^k (1-p)^{n-k}} = (pz + 1 - p)^n \quad ,\forall |z| \leq 1$$

$$\Rightarrow E(X) = g_X'(1-) = n(pz + 1 - p)^{n-1} \cdot p\big|_{z \to 1} = n \cdot 1 \cdot p = np$$

$$\text{Und:} \quad g_X''(1-) = n(n-1)(pz + 1 - p)^{n-2} p^2\big|_{z \to 1} = n(n-1)p^2$$

$$\Rightarrow \text{Var}(X) = n(n-1)p^2 + np - (np)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$$

Satz 6.2: (Eindeutigkeitssatz für erzeugende Funktionen)
Seien X und Y zwei Zufallsvariablen mit Werten in $\mathbb{N}_0$ und Verteilungsfunktionen F_X und F_Y und erzeugenden Funktionen g_x und g_Y. Dann gilt:

$$F_X(x) = F_Y(x) \quad \forall\, x \in \mathbb{R} \Leftrightarrow g_X(z) = g_Y(z) \;\forall\, z \in [-1, 1]$$

Beweis: Identitätssatz für Potenzreihen.

Satz 6.3: Seien X und Y zwei *unabhängige* Zufallsvariablen mit Werten in $\mathbb{N}_0$.
Dann gilt für die erzeugenden Funktionen:

$$g_{X+Y}(z) = g_X(z) \cdot g_Y(z)$$

Beweis: Es gilt: $g_{X+Y}(z) = E\left[z^{X+Y}\right] = E\left[z^X \cdot z^Y\right] = E\left[z^X\right] E\left[z^Y\right] = g_X(z) \cdot g_Y(z)$ □

Beispiel: Seien $X \sim \text{Bin}(n,p)$, $Y \sim \text{Bin}(m,p)$ und X und Y sind unabhängig. Nach dem letzten Beispiel gilt: $g_X(z) = (pz+1-p)^n$ und $g_Y(z) = (pz+1-p)^m$

$$\text{Satz 6.3} \Rightarrow g_{X+Y} = g_X(z) \cdot g_Y(z) = (pz+1-p)^{n+m} \Rightarrow X+Y \sim \text{Bin}(n+m,p)$$

[12] **Satz 6.4:** (Stetigkeitssatz für erzeugende Funktionen)
Gegeben sei eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ mit Werten in $\mathbb{N}_0$ und Zähldichte $\left(p_1^{(k)}\right)_{k\in\mathbb{N}_0}, \left(p_2^{(k)}\right)_{k\in\mathbb{N}_0}, \ldots$. Dann gilt:

a) $(g_{X_n}(z))_{n\in\mathbb{N}}$ konvergiert auf $(0,1) \Leftrightarrow \lim\limits_{n\to\infty} p_n(k) = p(k)$ existiert $\forall\, k \in \mathbb{N}_0$

b) Falls $g_{X_n}(z)$ auf $(0,1)$ konvergiert, so gegen $g(z) = \sum\limits_{k=0}^{\infty} p(k) z^k$.

Beispiel: (Grenzwertsatz von Poisson, vergleiche Lemma 3.1 [Seite 13])
Sei $X \sim B(n, \frac{\lambda}{n})$, $n \in \mathbb{N}$, $n \geq \lambda > 0$.
Die erzeugende Funktion von X_n ist (siehe Beispiel [Seite 33]):

$$g_{X_n}(z) = \left(z \cdot \frac{\lambda}{n} + 1 - \frac{\lambda}{n}\right)^n = \left(1 + \frac{\lambda}{n}(z-1)\right)^n$$

Also gilt: $\lim\limits_{n\to\infty} g_{X_n}(z) = \lim\limits_{n\to\infty} \left(1 + \frac{\lambda}{n}(z-1)\right)^n = \mathrm{e}^{\lambda(z-1)}$
Nach Lemma 3.1 und Satz 6.4 müßte das jetzt die erzeugende Funktion einer Poisson-Verteilung mit Parameter λ sein.

$$g_X(z) = \sum_{k=0}^{\infty} z^k P(X=k) = \sum_{k=0}^{\infty} z^k \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = \mathrm{e}^{-\lambda} \cdot \mathrm{e}^{\lambda z} = \mathrm{e}^{\lambda(z-1)}$$

Also folgt nach Satz 6.4 $\quad p_n(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \overset{n\to\infty}{\to} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}_0$
Im folgenden sei jetzt X wieder eine beliebige Zufallsvariable mit Werten in $\mathbb{R}$. Anstatt der erzeugenden Funktion betrachten wir hier die sogenannte charakteristische Funktion.

Definition: Sei X eine beliebige Zufallsvariable mit Werten in $\mathbb{R}$. Dann heißt die Funktion $\varphi_X : \mathbb{R} \to \mathbb{C}$ mit $\varphi_X(t) := E[\mathrm{e}^{itX}]$ *charakteristische Funktion* von X.

Bemerkung:

(i) Es gilt: $\varphi_X(t) = E[\cos(tX)] + iE[\sin(tX)] \quad \forall\, t \in \mathbb{R}$.

(ii) $\varphi_X(0) = 1$ und $|\varphi_X(t)| \leq 1 \quad \forall\, t \in \mathbb{R}$

(iii) Falls X Werte in $\mathbb{N}_0$ annimmt, gilt: $\varphi_X(t) = \sum\limits_{n=0}^{\infty} \mathrm{e}^{itn} P(X=n)$.

[12] 16.01.2003

(iv) Falls X absolut stetig ist mit Dichte $f_X(x)$ gilt:

$$\varphi_X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} f_X(x)\mathrm{d}x.$$

Beispiel: $X(\omega) = \mu$.
$\varphi_X(t) = E[\mathrm{e}^{it\mu}] = \mathrm{e}^{it\mu}$
Beispiel: $X \sim \mathcal{N}(0,1)$.
X besitzt die Dichte $f_X(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$
$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{itx}\mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x = \frac{\mathrm{e}^{-\frac{-t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(x-it)^2}\mathrm{d}x = \frac{\mathrm{e}^{-\frac{-t^2}{2}}}{\sqrt{2\pi}} h(t)$
mit $h(t) = \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(x-it)^2}\mathrm{d}x$
Wir differenzieren h: $h'(t) = \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(x-it)^2}\left(-\frac{1}{2}\cdot 2(x-it)(-i)\right)\mathrm{d}x = -i\mathrm{e}^{-\frac{1}{2}(x-it)^2}\Big|_{-\infty}^{\infty} = 0$
$\Rightarrow h(t) \equiv$ const. Außerdem ist $\varphi_X(0) = \frac{1}{\sqrt{2\pi}}h(0) = 1 \Rightarrow h(t) \equiv \sqrt{2\pi}$
Insgesamt: $\varphi_X(t) = \mathrm{e}^{-\frac{t^2}{2}}$

Lemma 6.5: Für jede Zufallsvariable X und Konstanten $a, b \in \mathbb{R}$ gilt:

$$\varphi_{aX+b}(t) = \mathrm{e}^{ibt}\varphi_X(at)$$

Beweis: $\varphi_{aX+b}(t) = E\left[\mathrm{e}^{it(aX+b)}\right] = E\left[\mathrm{e}^{itb}\mathrm{e}^{itaX}\right] = \mathrm{e}^{itb}\varphi_X(at)$. $\square$
Beispiel: Ist $X \sim \mathcal{N}(0,1)$, so ist $Y := \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$ für $\mu \in \mathbb{R}$, $\sigma > 0$.

$$\Rightarrow \varphi_Y(t) = \mathrm{e}^{it\mu}\varphi_X(\sigma t) = \mathrm{e}^{it\mu}\mathrm{e}^{-\frac{(\sigma t)^2}{2}}$$

Lemma 6.6: Sei $m \in \mathbb{N}$ und X eine Zufallsvariable. Falls $E\left[|X|^m\right] < \infty$ folgt:

$$\varphi_X^{(m)}(t) \text{ existiert und } \varphi_X^{(m)}(0) = i^m E\left[X^m\right]$$

Beweis: Es gilt $\varphi_X^{(m)}(t) = E\left[(iX)^m\mathrm{e}^{itX}\right]$. Setze $t = 0 \Rightarrow$ Behauptung $\square$
Beispiel: $X \sim \mathcal{N}(\mu, \sigma^2)$
Dann ist die charakteristische Funktion gegeben durch (siehe oben):

$$\begin{aligned}
\varphi_X(t) &= \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right) \\
\Rightarrow \varphi_X'(t) &= \varphi_X(t)(i\mu - \sigma^2 t) \\
\Rightarrow \varphi_X'(0) &= 1 \cdot i\mu = i \cdot E[X] \Rightarrow E[X] = \mu. \\
\varphi_X''(t) &= \varphi_X(t)(i\mu - \sigma^2 t)^2 + \varphi_X(t)(-\sigma^2) \\
\Rightarrow \varphi_X''(0) &= (i\mu)^2 - \sigma^2 = -(\mu^2 + \sigma^2) = \underbrace{i^2}_{=-1} E\left[X^2\right] \\
\Rightarrow E\left[X^2\right] &= \mu^2 + \sigma^2 \\
\Rightarrow \text{Var}(X) &= E[X^2] - (E[X])^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2
\end{aligned}$$

Satz 6.7: (Eindeutigkeitssatz)
Seien X und Y Zufallsvariablen mit Verteilungsfunktion F_X und F_Y, sowie charakteristischen Funktionen φ_X und φ_Y. Dann gilt:

$$F_X(t) = F_Y(t) \quad \forall\, t \in \mathbb{R} \Leftrightarrow \varphi_X(t) = \varphi_Y(t) \quad \forall\, t \in \mathbb{R}.$$

Satz 6.8: (Stetigkeitssatz)
Seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit Verteilungsfunktionen $F_X, F_{X_1}, F_{X_2}, \ldots$ sowie charakteristischen Funktionen $\varphi_X, \varphi_{X_1}, \varphi_{X_2}, \ldots$. Dann sind die beiden folgenden Aussagen äquivalent:

a) $\lim\limits_{n\to\infty} \varphi_{X_n}(t) = \varphi_X(t) \quad \forall\, t \in \mathbb{R}.$

b) $\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x) \quad x \in \mathbb{R}$, in denen die Grenzfunktion F_X stetig ist.

Eine typische Anwendung von charakteristischen Funktionen ist die Bestimmung der Faltung, also der Verteilungsfunktion von $X + Y$, wobei X und Y unabhängig sind.

Satz 6.9: Sind X und Y unabhängige Zufallsvariablen, so gilt:

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) \quad \forall\, t \in \mathbb{R}.$$

Beweis: $E\left[\mathrm{e}^{it(X+Y)}\right] = E\left[\mathrm{e}^{itX}\mathrm{e}^{itY}\right] = E\left[\mathrm{e}^{itX}\right] E\left[\mathrm{e}^{itY}\right]. \qquad \square$

Beispiel: Sei $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, X und Y unabhängig.

$$\varphi_{X+Y}(t) = \exp\left(it\mu_1 - \frac{(\sigma_1 t)^2}{2}\right) \exp\left(it\mu_2 - \frac{(\sigma_2 t)^2}{2}\right) = \exp\left(it(\mu_1+\mu_2) - \frac{1}{2}t^2(\sigma_1^2+\sigma_2^2)\right)$$

$$\Rightarrow X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

§7 Grenzwertsätze

7.1 Der zentrale Grenzwertsatz

[13]Es seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit $E(X_1) = \mu < \infty$ und $\text{Var}(X_1) = \sigma^2 < \infty$. Wir betrachten jetzt die folgende Zufallsvariable:

$$T_n := \frac{X_1 + \ldots + X_n - n \cdot \mu}{\sigma\sqrt{n}} \quad , n \in \mathbb{N}$$

Die Zufallsvariablen T_n sind standardisiert im folgenden Sinne:

$$E(T_n) = \frac{1}{\sigma\sqrt{n}} \left(\underbrace{E(X_1) + \ldots + E(X_n)}_{n\cdot\mu} - n\mu \right) = 0$$

$$\text{Var}(T_n) = \frac{1}{\sigma^2 n} \left(\underbrace{\text{Var}(X_1) + \ldots + \text{Var}(X_n)}_{n\cdot\sigma^2} \right) = 1$$

Satz 7.1: (Zentraler Grenzwertsatz)
Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilte Zufallsvariablen mit $E(X_1) = \mu < \infty$ und $0 < \text{Var}(X_1) = \sigma^2 < \infty$, so gilt für $T_n := \frac{X_1+\ldots+X_n-n\cdot\mu}{\sigma\sqrt{n}}$

$$F_{T_n}(x) = P(T_n \leq x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t = \Phi(x) = F_X(x) \text{ mit } X \sim \mathcal{N}(0,1)$$

für $n \to \infty$ und $\forall\, x \in \mathbb{R}$.
Beweis: Wir betrachten zunächst den Fall $\mu = 0, \sigma^2 = 1$
Dann ist $T_n = \frac{X_1+\ldots+X_n}{\sqrt{n}} =: \frac{S_n}{\sqrt{n}}$ und es gilt:

$$\varphi_{T_n}(t) \overset{L.6.5}{=} \varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) \overset{S.6.9}{=} \left[\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

$\varphi_{X_1}(t)$ kann man jetzt in eine Taylor-Reihe um $t = 0$ entwickeln:

$$\begin{aligned} \varphi_{X_1}(t) &= \varphi_{X_1}(0) + \varphi'_{X_1}(0)t + \varphi''_{X_1}(0)\frac{t^2}{2} + R_t \quad \text{mit } \frac{R_t}{t^2} \to 0 \text{ für } t \to 0. \\ &= 1 + it\mu - \frac{t^2}{2}E[X^2] + R_t \\ \text{Also:} \quad \varphi_{T_n}(t) &= \left[1 - \frac{t^2}{n\cdot 2} \cdot 1 + R\left(\frac{t}{\sqrt{n}}\right)\right]^n \to \mathrm{e}^{-\frac{t^2}{2}} \quad \text{für } n \to \infty. \end{aligned}$$

$\mathrm{e}^{-\frac{t^2}{2}}$ ist aber die charakteristische Funktion von $X \sim \mathcal{N}(0,1)$.
Damit folgt die Behauptung aus dem Stetigkeitssatz (Satz 6.8 [Seite 36]) für charakteristische Funktionen.
Ist $\mu \neq 0$ und $\sigma^2 \neq 1$, so definiere $Y_n := \frac{X_n-\mu}{\sigma}$, $n \in \mathbb{N}$.
Hier gilt $E(Y_n) = 0$ und $\text{Var}(Y_n) = 1$. Außerdem:
$T_n = \frac{Y_1+\ldots+Y_n}{\sqrt{n}}$ und die Behauptung folgt aus dem ersten Fall. □

[13]23.01.2003

Korollar 7.2: Unter den Voraussetzungen von Satz 7.1 [Seite 39] gilt für $\alpha < \beta$:

$$P\left(\alpha \leq \frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\sigma} \leq \beta\right) \to \Phi(\beta) - \Phi(\alpha) \quad \text{für } n \to \infty.$$

Beweis: Linke Seite:

$$\begin{array}{lccccccccc} F_{T_n}(\beta) - F_{T_n}(\alpha-) = I_n, & F_{T_n}(\beta) & - & F_{T_n}(\alpha) & \leq I_n \leq & F_{T_n}(\beta) & - & F_{T_n}(\alpha-\epsilon) & \forall\, \epsilon > 0 \\ & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\ n \overset{S.7.1}{\to} \infty & \Phi(\beta) & & \Phi(\alpha) & & \Phi(\beta) & & \Phi(\alpha-\epsilon) & \end{array}$$

$\epsilon \to 0$ liefert die Behauptung. $\square$

Beispiel: $X_1, X_2, \ldots, X_{1000}$ seien unabhängig und identisch verteilte Zufallsvariablen mit $P(X_1 = 1) = \frac{1}{5}$, $P(X_1 = 3) = \frac{1}{4}$, $P(X_1 = 6) = \frac{2}{5}$, $P(X_1 = 11) = \frac{3}{20}$

Wie groß ist die Wahrscheinlichkeit, daß $S_{1000} = \sum_{j=1}^{1000} X_j$ einen Wert zwischen 4820 und 5180 annimmt?

$\mu = E(X_1) = \frac{1}{5} + \frac{3}{4} + 6 \cdot \frac{2}{5} + 11 \cdot \frac{3}{20} = 5$

$\sigma^2 = \text{Var}(X_1) = \frac{21}{5}$

$$\begin{aligned} P(4820 \leq S_{1000} \leq 5180) &= P(-180 \leq S_{1000} - 5000 \leq 180) \\ &= P\left(\frac{-180}{\sqrt{1000 \cdot \frac{21}{5}}} \leq \frac{S_{1000} - 5000}{\sqrt{1000 \cdot \frac{21}{5}}} \leq \frac{180}{\sqrt{1000 \cdot \frac{21}{5}}}\right) \\ &\approx \Phi(0,57) - \Phi(-0,57) = 0,4314 \end{aligned}$$

Beispiel: $X_1, X_2, \ldots$ seien unabhängig und identisch verteilte Zufallsvariablen mit $X_1 \sim B(1,p)$, also $P(X_1 = 1) = p$ und $P(X_1 = 0) = 1 - p$

Wir haben $\mu = E(X_1) = p$, $\sigma^2 = \text{Var}(X_1) = p(1-p)$

Sei $S_n = X_1 + \ldots + X_n$. Dann gilt nach dem zentralen Grenzwertsatz

$$P\left(\frac{S_n - np}{\sqrt{n \cdot (1-p)}} \leq x\right) \to \Phi(x) \quad \text{für } n \to \infty$$

$$\Rightarrow \quad P\left(S_n \leq np + x\sqrt{np(1-p)}\right) \to \Phi(x) \quad \text{für } n \to \infty$$

Da $S_n \sim B(n,p)$ gilt:

$$\sum_{k \leq np + x\sqrt{np(1-p)}} \binom{n}{k} p^k (1-p)^{n-k} \to \Phi(x) \quad \text{für } n \to \infty$$

Grenzwertsatz von DeMoivre-Laplace

Beispiel: Prüfen einer Hypothese

Ein Würfel wird 1000 mal geworfen. Dabei fällt 120 mal die Eins. Ist der Würfel gezinkt?

Vorgehen: Wir berechnen die Wahrscheinlichkeit dafür, daß höchstens 120 mal die Eins fällt. Ist diese Wahrscheinlichkeit sehr klein, darf man annehmen, daß der Würfel gezinkt ist.

Wir definieren folgende Zufallsvariable:

$$X_k = \begin{cases} 1 & \text{, falls im k-ten Wurf eine Eins fällt} \\ 0 & \text{, falls im k-ten Wurf keine Eins fällt} \end{cases} \qquad k = 1, \ldots, 1000$$

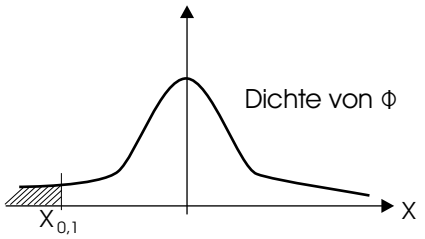
Wir nehmen an, daß die $X_1, \ldots, X_{1000}$ unabhängig und $P(X_1 = 1) = \frac{1}{6}$, $P(X_1 = 0) = \frac{5}{6}$.
$\Rightarrow E(X_1) = \frac{1}{6}$, Var$(X_1) = \frac{5}{36}$
Zentraler Grenzwertsatz: $S_{1000} = X_1, \ldots, X_{1000}$

$$P\left(\frac{S_{1000} - \frac{1000}{6}}{\sqrt{1000 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \leq \underbrace{\frac{120 - \frac{1000}{6}}{\sqrt{1000 \cdot \frac{1}{6} \cdot \frac{5}{6}}}}_{-\frac{28}{\sqrt{50}} \approx -4}\right) \approx \Phi(-4) \leq 10^{-6}$$

$$P(S_{1000} \leq 120) \leq 10^{-6}$$

Wir können also annehmen, daß der Würfel gezinkt ist.
Bemerkung:

Die Zahl $x_q \in \mathbb{R}$ die $\Phi(x_q) = q$ erfüllt mit $q \in (0, 1)$ heißt q-Quantil der Standardnormalverteilung.



Satz 7.3: (Berry Esséen)
Sind die Zufallsvariablen $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit $\mu = E(X_1) < \infty$ und $|\mu|_3 := E(|X_1 - \mu|^3) < \infty$, so gilt:

$$\sup_{x \in \mathbb{R}} |F_{T_n}(x) - \Phi(x)| \leq \frac{|\mu|_3}{\sigma^3 \sqrt{n}}$$

7.2 Gesetze der großen Zahlen

[14] **Satz 7.4:** (Schwaches Gesetz der großen Zahlen)
Sind die identisch verteilten Zufallsvariablen X_1, X_2 unabhängig und existiert ihr Erwartungswert $\mu = E(X_1)$, so gilt $\forall\, \varepsilon > 0$:

$$\lim_{n \to \infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

Beweis: Geht mit charakteristischen Funktionen und den Sätzen 6.7 und 6.8 [Seite 36].
Falls zusätzlich Var$(X_1) < \infty$, dann ist der Beweis einfacher:
Wir verwenden die Ungleichung von Chebyshev:

$$E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}\left(E(X_1) + \ldots + E(X_n)\right) = E(X_1) = \mu$$

$$\text{Var}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2}\left(\text{Var}(X_1) + \ldots + \text{Var}(X_n)\right) = \frac{1}{n}\text{Var}(X_1)$$

Also:

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\text{Var}(X_1)}{n \cdot \varepsilon^2} \to 0 \text{ für } n \to \infty. \qquad \square$$

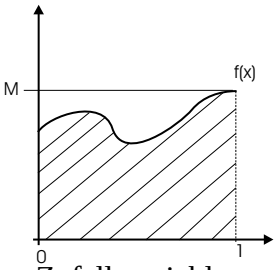
[14] 30.01.2003

Satz 7.5: (Starkes Gesetz der großen Zahlen)
Sind die identisch verteilten Zufallsvariablen X_1, X_2 unabhängig und existiert ihr Erwartungswert $\mu = E(X_1)$, so gilt:

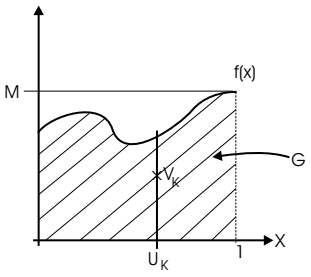
$$P\left(\left\{\omega \;\middle|\; \lim_{n\to\infty} \frac{X_1(\omega) + \ldots + X_n(\omega)}{n} = \mu\right\}\right) = 1$$

Schreibweise: $\frac{X_1+\ldots+X_n}{n} \to \mu$ für $n \to \infty$ fast sicher.

Anwendungsbeispiel: Monte Carlo Simulation
Es sei $f : [0,1] \to \mathbb{R}$ eine stetige Funktion.
$M := \max\limits_{X\in[0,1]} f(x)$ sei der maximal angenommene Wert.



Wir wollen $\int\limits_0^1 f(x)\mathrm{d}x$ numerisch näherungsweise berechnen.
Dazu sei $U_1, U_2, \ldots$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit $U_1 \sim \mathcal{U}(0,1)$.
Weiter sei $V_1, V_2, \ldots$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit $V_1 \sim \mathcal{U}(0, M)$.
Wir setzen: $I_k := \begin{cases} 1 & \text{, falls } f(U_k) > V_k \\ 0 & \text{, falls } f(U_k) \le V_k \end{cases}$
Dann gilt: $I_1, I_2, \ldots$ ist eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit



$$\begin{aligned} E[I_1] &= 1 \cdot P(f(U_1) > V_1) + 0 \cdot P(\ldots) = P((U_1, V_1) \in G) = \\ &= \int_0^1 \int_o^{f(x)} 1 \cdot \mathrm{d}y\mathrm{d}x = \int_0^1 f(x)\mathrm{d}x \end{aligned}$$

Aus dem starken Gesetz der großen Zahlen folgt also:

$$\frac{1}{n}\sum_{k=1}^{n} I_k \to \int_0^1 f(x)\mathrm{d}x \text{ für } n \to \infty \text{ fast sicher.}$$

Anwendungsbeispiel aus der Zahlentheorie
Es sei $\Omega = [0,1]$ und $\omega \in \Omega$.
Wir betrachten die Dualbruchzerlegung von ω, d.h. wir schreiben $\omega = 0, \omega_1\omega_2 \ldots$ mit $\omega_i \in \{0,1\}$, wobei $\omega = \sum\limits_{i=1}^{\infty} \frac{w_i}{2^i}$
Weiter seien die Zufallsvariablen $X_1, X_2 : \Omega \to \{0,1\}$ definiert durch

$$X_n(\omega) = X_n(0, \omega_1\omega_2\ldots) = \omega_n \text{ für } n \in \mathbb{N}$$

$$\begin{aligned} \text{und} \quad A_n &= \{\omega \in \Omega \mid X_1(\omega) = x_1, \ldots, X_n(\omega) = x_n\} \text{ für feste } x_1, \ldots, x_n \in \{0,1\}. \\ &= \left\{\omega \in \Omega \;\middle|\; \frac{x_1}{2} + \frac{x_2}{2^2} + \ldots + \frac{x_n}{2^n} \le \omega < \frac{x_1}{2} + \frac{x_2}{2^2} + \ldots + \frac{x_n}{2^n} + \frac{1}{2^n}\right\} \end{aligned}$$

Wir nehmen an, daß $P([a,b]) = b - a$, für $0 \le a < b \le 1$, also $P(A_n) = \frac{1}{2^n}$

$$
\begin{aligned}
\Rightarrow P(X_1 = 0) &= P\left(\left\{\omega \in \Omega \,\middle|\, \frac{0}{2} \leq \omega < \frac{0}{2} + \frac{1}{2}\right\}\right) = \frac{1}{2} \\
P(X_1 = 1) &= \frac{1}{2} \\
P(X_2 = 0) &= P(X_2 = 0, X_1 = 0) + P(X_2 = 0, X_1 = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}
\end{aligned}
$$

und $P(X_1 = x_1, X_2 = x_2) = \frac{1}{4} = P(X_1 = x_1)P(X_2 = x_2)$ für $x_1, x_2 \in \{0, 1\}$ etc.
Also sind die Zufallsvariablen $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit
$P(X_k = 0) = P(X_k = 1) = \frac{1}{2}$ und $E(X_k) = \frac{1}{2}$

Nach Satz 7.5 [Seite 40] gilt: $\frac{1}{n} \sum_{k=1}^{n} I_{[X_k=1]} \to \frac{1}{2}$ fast sicher für $n \to \infty$.

d.h. fast alle Zahlen des Intervalls $[0, 1]$ (bis auf eine Menge $A \subset [0, 1]$ mit $P(A) = 0$) sind in dem Sinne normal, daß der Anteil der Nullen und Einsen in ihrer Dualbruchzerlegung gegen $\frac{1}{2}$ konvergiert.

§8 Erzeugung von Zufallszahlen

Grundlagen für die Erzeugung von sogenannten Pseudo-Zufallszahlen, die Realisierungen einer bestimmten Verteilungsfunktion F_X seien sollen, ist die Erzeugung von $(0,1)$-gleichverteilten Zufallszahlen.

Erzeugung von (0,1)-gleichverteilten Zufallszahlen:

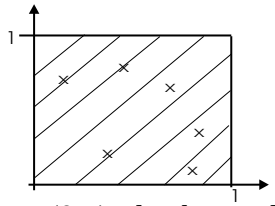
Dies geschiet zum Beispiel mit dem linearen Kongruenzgenerator.

Es sei $v_0 \in \mathbb{N},\ a, m \in \mathbb{N}, m$ eine Primzahl und $c \in \{0, 1, \ldots, m-1\}$

Dann liefert die Folge (u_n) mit

$v_n = (a \cdot v_{n-1} + c) \bmod m$

$u_n = \frac{v_n}{m}$



eine Realisierung einer Folge von unabhängigen, (0,1)-gleichverteilten Zufallsvariablen. Ein guter Generator zeichnet sich dadurch aus, daß n-Tupel von erzeugten Zahlen nicht auf Hyperebenen liegen.

„gute Werte "erhält man z.B. für $v_0 = 100000,\ m = 2^{31} - 1,\ a = 950706376$

Aus diesem (0,1)-gleichverteilten Zufallszahlen kann man nun Zufallszahlen für andere Verteilungen konstruieren.

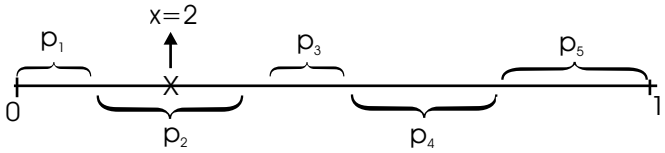
Diskrete Zufallsvariablen

Sei X eine diskrete Zufallsvariable mit $P(X = x_i) = p_i,\ i = 1, \ldots, k$

(U_n) sei eine Folge von unabhängigen, (0,1)-gleichverteilten Zufallsvariablen.

Das Ziel ist daraus Zufallsvariablen (Z_n) zu konstruieren mit $F_z = F_X$.

Einfache Methode:



$$Z_n := x_j, \text{ falls } \sum_{i=1}^{j-1} p_i \leq U_n < \sum_{i=1}^{j} p_i$$

Sollen viele Zufallszahlen erzeugt werden, ist folgender Algorithmus besser geeignet:

Alias Algorithmus

Schritt 0: o.B.d.A. $p_1 \leq p_2 \leq \ldots \leq p_k$.
Bestimme h mit $p_1 \leq p_2 \leq \ldots \leq p_h \leq \frac{1}{k} < p_{h+1} \leq \ldots \leq p_k$.
FOR $i = 1, \ldots, k$ DO
$r_i := p_i;\ a_i := 0;$
FOR $i = 1, \ldots, k$ DO
IF $r_i < \frac{1}{k}$ THEN $a_i := h + 1;\ r_{h+1} := r_{h+1} - \left(\frac{1}{k} - r_i\right);$
IF $r_{h+1} \leq \frac{1}{k}$ THEN $h := h + 1$

Schriitt 0: FOR $i = 1, 2, \ldots$ DO
Erzeuge $u_i \sim \mathcal{U}(0,1)$;
$j := \lfloor k \cdot u_i \rfloor + 1$;
$d := j - ku_i$;
IF $d \leq k \cdot r_j$ THEN $z_i := x_j$;
ELSE $z_i := x_{aj}$;

Vorteil: In Schritt 0 werden nur einmal die Werte r_i und a_i bestimmt. Bei der eigentlichen Erzeugung der Zufallszahlen in Schritt 1 ist nur ein einfacher Vergleich nötig.

Beispiel:
Sei $p_1 = 0.1,\ p_2 = 0.2,\ p_3 = 0.3,\ p_4 = 0.4,\ k = 4$
$\frac{1}{k} = 0.25 \Rightarrow h = 2$

i=1: $r_1 = p_1 = 0.1 < 0.25 \Rightarrow a_1 := 3;\ r_3 := 0.3 - (0.25 - 0.1) = 0.15$
$r_3 = 0.15 \leq 0.25 \Rightarrow h := 3$

i=2: $r_2 = p_2 = 0.2 < 0.25 \Rightarrow a_2 := 4;\ r_4 := 0.4 - (0.25 - 0.2) = 0.35$
$r_4 = 0.35 \not\leq 0.25$

i=3: $r_3 = p_3 = 0.15 < 0.25 \Rightarrow a_3 := 4;\ r_4 := 0.35 - (0.25 - 0.15) = 0.25$
$r_4 = 0.25 \leq 0.25 \Rightarrow h := 4$

i=4: $r_4 = 0.25 \not< 0.25$

Also: $\begin{array}{llll} r_1 = 0.1, & r_2 = 0.2, & r_3 = 0.15, & r_4 = 0.25 \\ a_1 = 3, & a_2 = 4, & a_3 = 4, & (a_4 = 0) \end{array}$

Poisson-Verteilung
Seien $U_1, U_2, \ldots$ unabhängig und identisch verteilte Zufallsvariablen mit $U_i \sim \mathcal{U}(0,1).\alpha = 0$.
Dann ist

$$X := \min \left\{ n \in \mathbb{N} \;\middle|\; \prod_{k=1}^{n} U_k < \mathrm{e}^{-\alpha} \right\} - 1$$

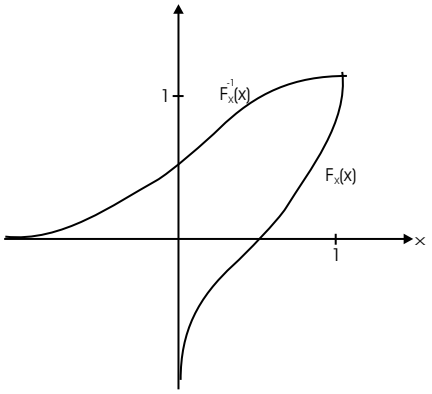
Poisson-verteilt mit Parameter α.

stetige Zufallsvariablen

Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(x)$.
Die Funktion $F_X^{-1} : [0,1] \to \mathbb{R} \cup \{-\infty, \infty\}$ mit

$$F_X^{-1} = \begin{cases} \inf\{t \in \mathbb{R} \mid F_X(t) \geq x\} & \text{für } x \in [0,1] \\ \sup\{t \in \mathbb{R} \mid F_X(t) = 0\} & \text{für } x = 0 \end{cases}$$

ist die sogenannte *Quantilfunktion*.



Ist $U \sim \mathcal{U}(0,1)$, dann ist $F_X^{-1}(U)$ verteilt wie X. (Inversionsmethode) (vgl. Aufgabe 41).
Ist F_X invertierbar, ist F_X^{-1} die übliche Inverse.
z.B. falls $X \sim \exp(\lambda)$ ist $F_X(x) = 1 - \mathrm{e}^{-\lambda x},\ x \geq 0$ und $F_X^{-1}(y) = -\frac{1}{\lambda}\log(1-y)$
Ist $U \sim \mathcal{U}(0,1)$, so ist $-\frac{1}{\lambda}\log(U) \sim \exp(\lambda)$ (vgl. Aufgabe 41)

Normalverteilung
Für die Normalverteilung gibt es eine Reihe von verschiedenen Verfahren. Eines davon ist die sogenannte Box-Muller-Methode.
Seien U_1, U_2 unabhängig und identisch verteilt mit $U_k \sim \mathcal{U}(0,1),\ k = 1, 2$. Dann sind

$$\begin{aligned} X_1 &:= \sqrt{-2\log(U_1)} \cdot \cos(2\pi U_2) \\ X_1 &:= \sqrt{-2\log(U_1)} \cdot \sin(2\pi U_2) \end{aligned}$$

unabhängig und $X_1, X_2 \sim \mathcal{N}(0,1)$ verteilt.

§9 Randomisierte Algorithmen

Hier werden im Laufe des Verfahrens zufällige Entscheidungen getroffen.

Las Vegas-Algorithmen

Liefern mit Wahrscheinlichkeit 1 das richtige Ergebnis, Laufzeit ist zufällig.

Monte Carlo-Algorithmen

Liefern mit Wahrscheinlichkeit < 1 das richtige Ergebnis. Sicherheit kann aber auf Kosten der Laufzeit beliebig groß gemacht werden.

Randomisiertes Quicksort

Aufgabe: Eine Menge von n verschiedenen Zahlen soll sortiert werden.
Üblicher Quicksort Algorithmus:

IF $|M| \in \{0, 1\}$ THEN Stop!
ELSE Wähle $x^* \in M$ (Discriminator)
Vergleiche jedes Element von $M - \{x^*\}$ mit x^*
Teile $M - \{x^*\}$ in $L := \{x \in M \mid x < x^*\}$ und $U := \{x \in M \mid x > x^*\}$
Sortiere L, U.

Der Algorithmus hat die längste Laufzeit, wenn x^* immer das kleinste oder größte Element ist.
Die Anzahl der Vergleiche beträgt hier: $\binom{n}{2} = \frac{n(n-1)}{2} = \mathrm{O}(n^2)$
Randomisierung: Wähle x^* zufällig, gleichverteilt auf M.
Der randomisierte Quicksort-Algorithmus ist also ein Las Vegas-Algorithmus.
$T_n =$ Anzahl der Vergleiche bei einem Feld der Größe n.
$E[T_n] = ?$

Satz 9.1: Die erwartete Anzahl von Vergleichen beim randomisierten Quicksort ist asymptotisch.

$$E[T_n] \sim 2n \log(n)$$

$$\left(\text{d.h. } \frac{E[T_n]}{2n \log(n)} \to \text{const für } n \to \infty,\ \mathrm{O}(n \log(n)) \right)$$

Index

σ-Algebra, **3**

absolut, **7**
    stetig, **15**, 18, 20, 35
    stetige Zufallsvariable, **14**, 24, 25
    stetiger Zufallsvektor, **18**, 19, 21, 22, 27
absolute Permutation, *siehe* Permutation
Algebra, **2**
Alias Algorithmus, **42**

Bayes´sche Formel, **10**
bedingte Wahrscheinlichkeit, **9**
Berry Esséen, **39**
Binomialverteilung, **12**, 13, 17, 18, 25, 28, 34
Borelsche σ-Algebra, **11**
Box-Muller-Methode, **22**, 43

Cauchyverteilung, **16**
charakteristische Funktion, **34**, 35–37

Dichte, 11, **14**, 18–22, 24, 25, 27, 31, 35
    Cauchyverteilung, **16**
    Exponentialverteilung, **15**
    Gleichverteilung, **16**
    Normalverteilung, **15**
      2-dimensionale, **30**
Dichtetransformation, **22**
diskrete
    Verteilungsfunktion, *siehe* Zähldichte
    Zufallsvariable, 11, **21**, 24, 25, 42
diskreter Zufallsvektor, **18**, 19, 21, 27

Eindeutigkeitssatz, *siehe* Satz
Elementarereignis, **2**
Ereignis, **2**
Erwartungswert, **24**, 26–28, 39, 40
Erwartungswertvektor, **30**
erzeugende Funktion, **33**, 34
Exponentialverteilung, **15**

Faltung, **36**
Faltungsformel, **21**

Gedächtnislosigkeit, **16**
gemeinsame
    Dichte, **21**, 22
    Verteilung, **17**
    Verteilungsfunktion, **18**
    Zähldichte, **18**
Geometrische Verteilung, **14**
Gesetz der großen Zahlen
    Schwaches, **39**
    Starkes, **40**
Gleichverteilung, **16**
Grenzwertsatz
    von DeMoivre-Laplace, **38**
    von Poisson, **34**

Hypergeometrische Verteilung, **13**

Inversionsmethode, **43**

k-tes Moment, **25**
k-tes zentriertes Moment, **25**
Kombination, **5**, 6
Kombinatorik, **2**
Korrelationskoeffizient, **28**, 29
Kovarianz, **28**
Kovarianzmatrix, **30**

Laplacescher Wahrscheinlichkeitraum, **4**

Messraum, **3**
Mittelwert der Verteilung, **24**
Monte Carlo Simulation, **40**
Multinomialverteilung, **17**, 18
Multiplikationssatz, **9**

Negative Binomialverteilung, **14**
Normalverteilung, **15**, 24, 35, 36, 43
    2-dimensionale, **30**

Permutation, **4**, 5, 7
    absolute, **7**
Poisson-Verteilung, **13**, 21, 34, 43
Pseudo-Zufallszahlen, *siehe* Zufallszahlen

q-Quantil, **39**
Quantilfunktion, **43**
Quicksort, **44**

Randdichte, **18**, 20
Randomisierte Algorithmen, **44**
  Las Vegas, **44**
  Monte Carlo, **44**
Randverteilung, **18**, 30

Satz
  Eindeutigkeitssatz, **36**
    für erzeugende Funktionen, **33**
  von der totalen Wahrscheinlichkeit, **9**
Siebformel, **4**
Standardabweichung, **25**, 26
Standardnormalverteilung, **15**, 35, 39, 43
Stetigkeitssatz, **36**, 37
  für erzeugende Funktionen, **34**

Taylor-Reihe, **37**

unabhängige
  Ereignisse, **8**
  Zufallsvariablen, **19**, 36
Ungleichung
  Markovsche, **31**
  von Cauchy-Schwarz, **31**
  von Chebyshev, **31**, 39
unkorreliert, **28**, 29

Varianz, **25**, 26
Verteilungsfunktion, **11**, 12–14, 17, 18, 33, 36, 42, 43
  Exponentialverteilung, **16**
  Gleichverteilung, **16**
  Normalverteilung, **15**

Wahrscheinlichkeits-
  maß, **3**, 4
  raum, 2, **3**, 4, 8, 9, 11, 16

Zähldichte, **12**, 21, 24, 25, 27, 33, 34
  Binomialverteilung, **13**
    Negative, **14**
  Geometrische Verteilung, **14**
  Hypergeometrische Verteilung, **13**
  Multinomialverteilung, **17**
  Poisson-Verteilung, **13**
Zentraler Grenzwertsatz, **37**, 38
Ziehen
  mit Zurücklegen
    mit Reihenfolge, **5**
    ohne Reihenfolge, **6**
  ohne Zurücklegen
    mit Reihenfolge, **5**
    ohne Reihenfolge, **5**
Zufallsvariable, **11**, 14, 16, 19–21, 24–29, 31, 33–43
Zufallsvektor, **16**, 17, 18, 20, 30
Zufallszahlen, **42**
  $(0, 1)$-gleichverteilte, **42**
  Pseudo-, **42**