

Vorlesung

Einführung in die Wahrscheinlichkeit

Prof. C. Mazza
Wintersemester 2007/2008

Literatur

W. Feller, An introduction to probability theory and some of its applications I (Wiley 1968).

K.L. Chung, Elementary probability theory with stochastic processes (Springer 1974).

J-Y. Ouvrard, Probabilités 1, Capes et Agrégation (Cassini 1998)

Contents

1	Der Begriff der Wahrscheinlichkeit, Wahrscheinlichkeitsräume, Beispiele	3
1.1	Verschiedene Wahrscheinlichkeitsbegriffe	3
1.2	Zufallsexperimente, Wahrscheinlichkeitsräume	4
1.3	Abzählbare Wahrscheinlichkeitsräume	4
1.4	Allgemeine Wahrscheinlichkeitsräume	7
2	Bedingte Wahrscheinlichkeiten, unabhängige Ereignisse	8
2.1	Die bedingte relative Häufigkeit:	8
2.2	Bedingte Wahrscheinlichkeit:	9
2.3	Unabhängigkeit:	9
3	Diskrete Zufallsgrößen	12
3.1	Die Verteilung einer Zufallsgröße	12
3.2	Einige Eigenschaften der Erwartung	13
3.3	Unabhängige reelle Zufallsgrößen	14
3.4	Moment, Varianz aund Kovarianz	15
3.5	Die Faltung von Wahrscheinlichkeiten	17

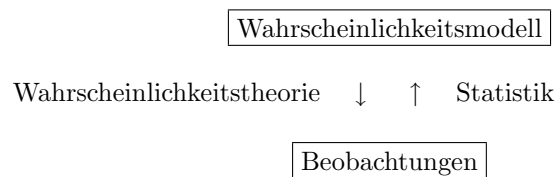
3.6	Liste einiger wichtigen (diskreten) Verteilungen	18
3.7	Die Verteilungsfunktion einer Zufallsgrösse	20
3.8	Erzeugende Funktionen	20
3.9	Beispiele von abhängigen Zufallsgrössen	20
4	Zufallsgrössen mit Dichten	23
4.1	Unabhängige Zufallsgrössen	26
4.2	Die Verteilungsfunktion einer Zufallsgrösse	27
4.3	Die Faltung von Dichten	28
4.4	Lineare Abbildungen von Zufallsvektoren	29
4.5	Funktionen von reellen Zufallsgrössen	30
4.6	Zwei weitere wichtige Dichten: Die Student und die Exponential Verteilungen	30
5	Die Gesetze der grossen Zahlen	32
5.1	Die Ungleichung von Tschebyscheff	33
5.2	Das schwache Gesetz der grossen Zahlen	34
5.3	Das starke Gesetz der grossen Zahlen	34
5.4	Anwendung der Gesetze der grossen Zahlen	34
5.5	Markovsche Ungleichung	35
6	Der zentrale Grenzwertsatz	36

Vorbemerkungen

Wahrscheinlichkeit und Statistik haben zwei gemeinsame Wurzeln, die früh zusammengewachsen sind:

1. Glücksspiele
2. Elementare beschreibende Statistik (Statistik = “Zusammenstellung von numerischen Daten für die Zwecke des Staates”)

Heutige Unterscheidung:



Ohne wahrscheinlichkeitstheoretische Grundlegung kann man die heutige Statistik *nicht* verstehen. Deshalb wird die Statistik im Sommersemester behandelt.

1 Der Begriff der Wahrscheinlichkeit, Wahrscheinlichkeitsräume, Beispiele

1.1 Verschiedene Wahrscheinlichkeitsbegriffe

Schwierigkeit: es gibt mindestens vier, nur teilweise miteinander verträgliche Wahrscheinlichkeitsbegriffe:

- a) Wahrscheinlichkeit = Mass des persönlichen Glaubens.

Das entspricht dem umgangssprachlichen Wahrscheinlichkeitsbegriff; mathematisch formalisiert von L.J. Savage (Foundations of Statistics, Wiley, 1954). Kritik: die Wahrscheinlichkeitstheorie wird damit zu einer psychologischen Theorie (wie verknüpfen wir unseren a priori Glauben mit den Beobachtungen zu einem a posteriori Glauben), und unser Geist scheint diese Verknüpfung nicht nach der sogenannten Bayes'schen Formel (s. Kapitel II) vorzunehmen, wie es die Subjektivisten von einer "idealen" Person fordern.

- b) Wahrscheinlichkeit = $\frac{\text{Anzahl günstige Fälle}}{\text{Anzahl mögliche Fälle}}$.

Das ist die klassische Definition; die Wahrscheinlichkeit wird hier durch eine Symmetriebetrachtung gefunden.

Bemerkung die Wahrscheinlichkeit in 4 Würfeln mit einem Würfel mindestens einmal eine Sechs zu werfen, ist

$$\frac{\text{günstige Fälle}}{\text{mögliche Fälle}} = 1 - \frac{\text{ungünstige Fälle}}{\text{mögliche Fälle}} = 1 - \frac{5^4}{6^4}.$$

Empirischer Hintergrund: das Resultat eines einzelnen Wurfes ist zwar nicht vorhersagbar, auf die Länge treten aber alle sechs Möglichkeiten etwa gleichhäufig auf. Nachträglich versucht man das durch eine Symmetriebetrachtung zu begründen.

Beispiel n Würfe einer symmetrischen Münze. Gesucht ist die Wahrscheinlichkeit p_k , dass man k -mal "Kopf" erhält. Man hat

$$p_k = \binom{n}{k} \left(\frac{1}{2}\right)^n, \quad k = 0, 1, \dots, n.$$

Es gibt 2^n mögliche Ausgänge und $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ günstige Fälle!

Kritik: die klassische Definition erleidet Schiffbruch, sobald man gefälschte Würfel oder Münzen betrachtet.

- c) Wahrscheinlichkeit = Grenzwert der relativen Häufigkeit.

Diese Definition wird durch die bereits erwähnte beachtliche Stabilität der relativen Häufigkeit suggeriert. Es ist schwierig, daraus eine mathematische Definition zu machen:

wie ist eine “zufällige” Folge ins Unendliche fortzusetzen? Der Ansatz von von Mises (Wahrscheinlichkeit, Statistik und Wahrheit, Springer, Wien, 1936) ist nicht ganz adäquat, wurde aber vor wenigen Jahren in Ordnung gebracht (P. Martin Löf: Definition of random sequences. Information and Control **6** (1966), 602–619).

d) Wahrscheinlichkeit = implizit durch ein Axiomensystem definiert.

Dieser Ansatz ist sehr handlich und hat sich allgemein eingebürgert, erschöpft aber nicht alle Aspekte des Wahrscheinlichkeitsbegriffes (z.B. kann er nicht zwischen zufälligen und unzufälligen Folgen von 0 und 1 unterscheiden!).

In dieser Vorlesung werden wir die Wahrscheinlichkeit durch ein **Axiomensystem** definieren.

1.2 Zufallsexperimente, Wahrscheinlichkeitsräume

Empirische Tatsache: es gibt Experimente (z.B. viermaliges Werfen eines Würfels), welche unter den gleichen Bedingungen mehrfach wiederholt werden können, aber nicht immer das gleiche Resultat liefern. Bei oftmaliger Wiederholung stabilisiert sich jedoch die relative Häufigkeit der verschiedenen möglichen Ergebnisse $\omega_1, \omega_2, \dots, \omega_N$: wenn ω_i bei n -maliger Wiederholung n_i -mal aufgetreten ist, scheint $\frac{n_i}{n}$ für $n \rightarrow \infty$ einem Grenzwert p_i zuzustreben.

Wir werden das folgende Zufallsexperiment später genauer analysieren.

Zufallsexperiment: n -maliges Werfen einer Münze.

Mögliche, unterscheidbare Ergebnisse (“Elementarereignisse”): jede Folge ω_i der Länge n von “Kopf” (0) und “Zahl” (1) ist ein mögliches Ereignis, es gibt also $N = 2^n$ mögliche Ergebnisse.

Wahrscheinlichkeiten: bei einer “idealen” Münze hat jedes mögliche Ergebnis ω (nach der klassischen Definition) die gleiche Wahrscheinlichkeit 2^{-n} ; bei einer “gefälschten” Münze werden die Wahrscheinlichkeiten verschieden sein.

Beachte: dieses Zufallsexperiment kann auch als n -malige Wiederholung eines Zufallsexperimentes mit nur zwei möglichen Ergebnissen aufgefasst werden.

Ein anderes Beispiel eines **Zufallsexperimentes**: Man wirft eine “ideale” Münze so lange bis man “Zahl” bekommt.

Mögliche Ergebnisse: alle Folgen ω_i der Form $\underbrace{(0, 0, \dots, 0)}_{(i-1)\text{-mal}}, 1, i = 1, 2, \dots$

Die Menge aller möglichen Ausgänge ist hier unendlich, aber abzählbar.

Wahrscheinlichkeiten: $p_i :=$ Wahrscheinlichkeit von $\omega_i = \left(\frac{1}{2}\right)^i, i = 1, 2, \dots$

Beachte: $\sum_{i=1}^{\infty} p_i = 1.$

1.3 Abzählbare Wahrscheinlichkeitsräume

Ein abzählbarer Wahrscheinlichkeitsraum besteht aus einer abzählbaren Menge $\Omega = \{\omega_1, \omega_2, \dots\}$; jedem Element (“Elementarereignis”) ω_i ist eine reelle Zahl $p_i \geq 0$ zugeordnet (die “Wahrscheinlichkeit”) p_i .

lichkeit" von ω_i), derart dass $\sum_{i=1}^{\infty} p_i = 1$.

Die Teilmengen $A \subseteq \Omega$ heissen **zusammengesetzte Ereignisse** oder kurz **Ereignisse**; die Wahrscheinlichkeit $P(A)$ eines Ereignisses ist definiert durch $P(A) = \sum_{i:\omega_i \in A} p_i$.

Es gilt:

- 1) $P(\emptyset) = 0$,
- 2) $P(\Omega) = 1$,
- 3) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ falls $A_i \cap A_j = \emptyset$ für $i \neq j$. (\cup ist die Vereinigung, \cap der Durchschnitt)

Eine auf der Menge \mathcal{A} aller Teilmengen von Ω definierte Funktion P , die die Eigenschaften 1), 2), 3) besitzt, wird Wahrscheinlichkeitsmass, Wahrscheinlichkeitsverteilung oder kurz Wahrscheinlichkeit genannt; das **Tripel** (Ω, \mathcal{A}, P) heisst **abzählbarer Wahrscheinlichkeitsraum**.

Wir stellen uns auf den axiomatischen Standpunkt: die p_k sind **beliebige vorgegebene** Zahlen.

Beabsichtigte Interpretation

- i) Bei oftmaliger Wiederholung des Experimentes tritt das Ereignis A mit einer relativen Häufigkeit nahe bei $P(A)$ auf.
- ii) Wenn $P(A)$ nahe bei 1 (resp. bei 0) liegt, trifft A bei einmaliger Durchführung des Experimentes praktisch sicher ein (resp. nicht ein).

Die Wahrscheinlichkeit wird also auch in dieser sogenannten "Häufigkeitsinterpretation" letzten Endes durch den subjektiven Glauben interpretiert, aber **nur qualitativ**, nicht quantitativ.

Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein abzählbarer Wahrscheinlichkeitsraum.

Die Folge A_1, A_2, \dots von Ereignissen heisst monoton wachsend (resp. fallend), falls $A_i \subseteq A_{i+1}, \forall_i (A_{i+1} \subseteq A_i, \forall_i)$ gilt.

Satz 1.1. A_1, A_2, \dots sei eine Folge von Ereignissen.

Behauptungen

1. $P(A_1^c) = 1 - P(A_1)$ (A_1^c bedeutet das Komplement von A)
2. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

$$3. P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) \\ - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

4.

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{\substack{i_1 < i_2 \\ i_1 < i_2}} P(A_{i_1} \cap A_{i_2}) + \\ \sum_{\substack{i_1 < i_2 < i_3 \\ i_1 < i_2 < i_3}} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n)$$

$$5. A_i \uparrow \implies \lim_{i \rightarrow \infty} P(A_i) = P\left(\bigcup_{i=1}^{\infty} A_i\right),$$

$$A_i \downarrow \implies \lim_{i \rightarrow \infty} P(A_i) = P\left(\bigcap_{i=1}^{\infty} A_i\right).$$

Beweis.

1. Man hat $A_1 \cup A_1^c = \Omega$ und somit $P(A_1) + P(A_1^c) = P(\Omega) = 1$.

2. Wegen $A_1 \cup A_2 = (A_1 - (A_1 \cap A_2)) \cup (A_2 - (A_1 \cap A_2)) \cup (A_1 \cap A_2)$ gilt
 $P(A_1 \cup A_2) = P(A_1) - P(A_1 \cap A_2) + P(A_2) - P(A_1 \cap A_2) + P(A_1 \cap A_2)$
 $= P(A_1) + P(A_2) - P(A_1 \cap A_2)$
 $(A - B := A \cap B^c)$

3. siehe 4.

4. Der Beweis geschieht durch Induktion über n . Die Behauptung ist richtig für $n = 2$. Nehmen wir an, sie sei bis $n - 1$ bewiesen. Dann ist

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right) \stackrel{2.}{=} P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right).$$

Nach Voraussetzung gilt

$$P\left(\bigcup_{i=1}^{n-1} A_i\right) = \sum_{i=1}^{n-1} P(A_i) - \sum_{\substack{i_1, i_2=1 \\ i_1 < i_2}}^{n-1} P(A_{i_1} \cap A_{i_2}) + \dots + (-1)^n P(A_1 \cap \dots \cap A_{n-1})$$

und

$$P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right) = P\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right) =$$

$$\sum_{i=1}^{n-1} P(A_i \cap A_n) - \sum_{\substack{i_1, i_2=1 \\ i_1 < i_2}}^{n-1} P(A_{i_1} \cap A_{i_2} \cap A_n) + \dots + (-1)^n P(A_1 \cap A_2 \dots \cap A_n).$$

Daraus folgt

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{\substack{i_1, i_2=1 \\ i_1 < i_2}}^n P(A_{i_1} \cap A_{i_2}) + \dots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right).$$

5. Setzen wir (im Falle, wo $A_i \uparrow$) $A'_i := A_i - A_{i-1}$, $i = 2, 3, \dots$, $A'_1 := A_1$. Dann gilt $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} A'_i$ und somit

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A'_i\right) = \sum_{i=1}^{\infty} P(A'_i),$$

denn die Ereignisse $\{A'_j\}$ sind paarweise disjunkt. Ferner gilt

$$\begin{aligned} \sum_{i=1}^{\infty} P(A'_i) &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A'_i) = \lim_{n \rightarrow \infty} \{P(A'_1) + \dots + P(A'_n)\} \\ &= \lim_{n \rightarrow \infty} \left\{ P(A_1) + (P(A_2) - P(A_1)) + \dots + (P(A_n) - P(A_{n-1})) \right\} \\ &= \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

Im Falle, wo $A_i \downarrow$ hat man $A_i^c \uparrow$. Deswegen ist

$$P\left(\bigcup_{i=1}^{\infty} A_i^c\right) = 1 - P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n^c) = \lim_{n \rightarrow \infty} (1 - P(A_n))$$

und somit

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n).$$

1.4 Allgemeine Wahrscheinlichkeitsräume

Ein Wahrscheinlichkeitsraum ist ein Tripel (Ω, \mathcal{A}, P) , bestehend aus einer beliebigen Menge Ω , einer Menge \mathcal{A} von Teilmengen (“Ereignisse”) von Ω und einer reellwertigen Funktion P auf \mathcal{A} , derart dass

- A0
1. $\Omega \in \mathcal{A}$,
 2. $A \in \mathcal{A} \implies A^c \in \mathcal{A}$,
 3. $A_i \in \mathcal{A}, i = 1, 2, \dots \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

(Eine solche Menge \mathcal{A} heisst σ -Algebra von Teilmengen.)

- A1
1. $0 \leq P(A) \leq 1, P(\Omega) = 1$,
 2. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ falls $A_i \cap A_j = \emptyset$ für $i \neq j$.

(Axiome von Kolmogoroff)

Es ist einfach zu sehen, dass der vorher bewiesene **Satz** auch im **allgemeinen Fall** gültig ist.

Bemerkung Betrachten wir das folgende Experiment: Ein Punkt wird im Intervall $[0, 1]$ “zufällig” ausgewählt.

Mögliche Ergebnisse: $\Omega := [0, 1]$.

Wahrscheinlichkeiten: Hier muss man $P(\{\omega\}) = 0$ setzen (warum?), und es ist nicht mehr möglich, die Wahrscheinlichkeit irgendwelcher Teilmengen A von Ω als die Summe der Wahrscheinlichkeiten ihrer Elemente zu definieren. Man kann aber zeigen, dass es eine einzige Funktion P auf der kleinsten σ -Algebra gibt, welche die Intervalle I enthält, so dass (A_1) 1, 2 mit $P(I) = \text{Länge von } I$ für alle Intervalle I erfüllt sind.

Beispiele von Wahrscheinlichkeiten, die durch **Symmetriebetrachtungen** ausgerechnet werden

Beispiel 3 Aus einem Kartenspiel (36 Karten) greift man auf gut Glück 3 Karten heraus. Gesucht ist die Wahrscheinlichkeit P dafür, dass sich unter ihnen genau ein As befindet.

Wir haben

$$P = \frac{\text{günstige Fälle}}{\text{mögliche Fälle}} = \frac{\binom{4}{1} \binom{32}{2}}{\binom{36}{3}} = \frac{496}{1785} \approx 0,2778.$$

Beispiel 4 Wir betrachten dasselbe Zufallsexperiment wie im dritten Beispiel. Gesucht ist die Wahrscheinlichkeit Q dafür, dass unter ihnen wenigstens ein As vorkommt (Ereignis A).

Wir haben

$$P(A) = 1 - P(A^c) = 1 - \frac{\binom{32}{3}}{\binom{36}{3}} \approx 0,3053.$$

Bemerkung: "auf gut Glück" bedeutet, dass alle möglichen Ausgänge gleichwahrscheinlich sind.

Beispiel 5 Eine Urne enthält n weisse und n rote Kugeln. Der Reihe nach zieht man zufällig eine Kugel und dies ohne Zurücklegen. Wie gross ist die Wahrscheinlichkeit P , dass im Laufe der Ziehung *nie* mehr rote Kugeln als weisse Kugel gezogen worden sind?

Antwort: $P = \frac{1}{n+1}$; der Beweis wird in der Vorlesung durchgeführt.

2 Bedingte Wahrscheinlichkeiten, unabhängige Ereignisse

2.1 Die bedingte relative Häufigkeit:

Wir betrachten ein Zufallsexperiment (z.B. einen Wurf mit einem symmetrischen Würfel). A und B seien zwei Ereignisse. Tritt bei n Wiederholungen des Experimentes genau n_B -mal das Ereignis B ein, und findet bei diesen n_B Versuchen $n_{A \cap B}$ -mal zusammen mit B auch das Ereignis A statt, so wollen wir den Quotienten

$$h_{A|B} = \frac{n_{A \cap B}}{n_B} \left(= \frac{n_{A \cap B}}{n} / \frac{n_B}{n} \right) \quad \text{die **bedingte relative Häufigkeit** nennen.}$$

Die bedingte relative Häufigkeit des Ereignisses A unter der Bedingung B in einer Versuchsfolge ist also gleich der relativen Häufigkeit von A in einer Teilfolge dieser Versuchsfolge, die aus denjenigen Versuchen der ursprünglichen Folge besteht, bei welchen B stattgefunden hat.

2.2 Bedingte Wahrscheinlichkeit:

Interpretiert man die Wahrscheinlichkeit eines Ereignisses als relative Häufigkeit, ist es dann sinnvoll, die **bedingte Wahrscheinlichkeit** $P(A|B)$ von A , gegeben B , wie folgt zu definieren

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \quad \text{falls } P(B) > 0 \text{ ist.}$$

(Hier wird vorausgesetzt, dass ein allgemeiner Wahrscheinlichkeitsraum vorgegeben ist.)

2.3 Unabhängigkeit:

Zwei Ereignisse A, B heissen unabhängig, wenn

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

gilt.

Bemerkung: Im Falle, wo $P(B) > 0$ ist, sind A und B unabhängig dann und nur dann, wenn $P(A|B) = P(A)$ ist.

Beachte: Die Definition von Unabhängigkeit ist symmetrisch. Die Frage, ob die kausale Unabhängigkeit durch stochastische Unabhängigkeit (wie oben definiert) formalisiert werden kann, kann nur empirisch entschieden werden.

Satz 2.1 (Der ‘‘Satz von der totalen Wahrscheinlichkeit’’ und die Formel von Bayes).
 (Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum.
 Seien B_1, \dots, B_k, A beliebige Ereignisse mit

- a) $P(B_i) > 0, \forall i$ und $P(A) > 0$,
- b) $B_i \cap B_j = \emptyset$ für $i \neq j$ und
- c) $\bigcup_{i=1}^k B_i = \Omega$.

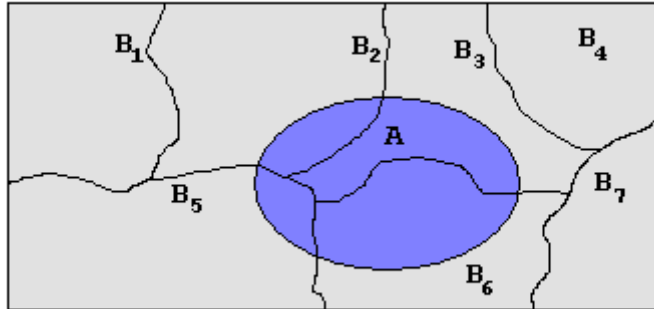
Dann gilt

$$- P(A) = \sum_{j=1}^k P(A|B_j)P(B_j) \quad (\text{‘‘Satz von der totalen Wahrscheinlichkeit’’}).$$

- Die unmittelbar daraus folgende Beziehung

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^k P(A|B_j)P(B_j)}$$

wird **Formel von Bayes** genannt.



Diese Formel hat eine fundamentale Bedeutung in der subjektiven Wahrscheinlichkeitsauffassung: sei $P(B_i)$ das Mass unseres *a priori* Glaubens an die Richtigkeit der Hypothese B_i ; wir kennen ausserdem die bedingten Wahrscheinlichkeiten $P(A|B_i)$ für das Eintreffen von A unter den verschiedenen Hypothesen. Wenn nun das Experiment tatsächlich das Resultat A ergeben hat, modifiziert eine “ideale” Person ihren *a priori* Glauben zum *a posteriori* Glauben $P(B_i|A)$ gemäss der Bayes’schen Formel.

Beispiel 1 (vgl. Kapitel I, Beispiel 2)

Zufallsexperiment: n Würfe mit einer idealen Münze.

A_k : der k -te Wurf ergibt “Zahl”. Man hat

$$P(A_k) = \frac{2^{n-1}}{2^n} = \frac{1}{2}, \quad P(A_k \cap A_\ell) = \frac{2^{n-2}}{2^n} = \frac{1}{4} \quad \text{für } k \neq \ell$$

\implies für $k \neq \ell$ sind A_k und A_ℓ unabhängig.

Beispiel 2

Ich habe einen Sack voll Münzen. Die Hälfte davon fällt mit Wahrscheinlichkeit $p = 0,9$ “Kopf”, die andere Hälfte mit Wahrscheinlichkeit $p = 0,1$. Ich ziehe auf Geratewohl eine Münze aus dem Sack und werfe sie zweimal. Sei K_i das Ereignis: “Kopf” im i -ten Wurf. Dann gilt:

$$P(K_1) = \underbrace{P(K_1|p=0.9)}_{0.9} \underbrace{P(p=0.9)}_{0.5} + \underbrace{P(K_1|p=0.1)}_{0.1} \underbrace{P(p=0.1)}_{0.5} = 0.5$$

$$P(K_2) = 0,5,$$

$$P(K_1 \cap K_2) = (0.9)^2 \cdot 0.5 + (0.1)^2 \cdot 0.5 = 0.41,$$

$$P(K_2|K_1) = \frac{0.41}{0.5} = 0.82.$$

Angenommen, ich habe zweimal “Kopf” geworfen. Wie gross ist die Wahrscheinlichkeit, dass meine Münze zur Klasse $p = 0.9$ gehört? (a posteriori Glauben!)

$$P(“p = 0.9”|K_1 \cap K_2) = \frac{P(“p = 0.9” \cap (K_1 \cap K_2))}{P(K_1 \cap K_2)} = \frac{0.5 \cdot 0.9 \cdot 0.9}{0,41} = 0.988$$

(Formel von Bayes mit $A = K_1 \cap K_2$, $B_1 \sim “p = 0.1”$ und $B_2 \sim “p = 0.9”$)

Unabhängige Ereignisse

Definition Eine Familie A_1, A_2, \dots, A_n heisst *unabhängig*, falls

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \quad \text{für alle Teilmengen } J \text{ von } \{1, 2, \dots, n\}.$$

Zum Beispiel, die Familie A_1, A_2, A_3 ist unabhängig, falls $P(A_1 \cap A_2) = P(A_1)P(A_2)$, $P(A_2 \cap A_3) = P(A_2)P(A_3)$, $P(A_1 \cap A_3) = P(A_1)P(A_3)$ und $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$ gilt.

Definition Die Ereignisse A_1, A_2, \dots, A_n heissen **paarweise unabhängig**, falls $P(A_i \cap A_j) = P(A_i)P(A_j)$ für $i \neq j$ gilt.

Beachte: paarweise Unabhängigkeit impliziert *nicht* die Unabhängigkeit der Familie.

Beispiel: Wir betrachten 2 Würfe mit einem Würfel und definieren drei Ereignisse wie folgt
 $A_1 \sim$ 1. Wurf zeigt gerade Augenzahl,
 $A_2 \sim$ 2. Wurf zeigt gerade Augenzahl,
 $A_3 \sim$ beide Würfe haben die gleiche Parität.

In diesem Falle sind die Ereignisse A_1, A_2, A_3 paarweise unabhängig, aber die Familie ist nicht unabhängig.

Beispiel 3 Rotgrün-Blindheit (R): Eine meist angeborene Störung des Farbensinnes; Farben zwischen Rot und Grün erscheinen als verschieden helles Gelb.

Untersuchungen haben ergeben: Bei den Männern (M) tritt R viel häufiger auf als bei den Frauen (F). Man kann nämlich annehmen, dass $P(R|M) = 8\%$ und $P(R|F) = 0,4\%$ gilt.

Wir wollen jetzt die bedingte Wahrscheinlichkeit $P(M|R)$ des "Ereignisses" M , gegeben R ausrechnen. Um die Sache zu vereinfachen, setzen wir $P(M) = P(F) = 1/2$. Nach dem "Satz von der totalen Wahrscheinlichkeit" und der Formel von Bayes erhalten wir

$$P(R) = P(R|M)P(M) + P(R|F)P(F) = 0,08 \cdot 0,5 + 0,004 \cdot 0,5 = 0,042$$

und somit

$$P(M|R) = \frac{P(R|M) \cdot P(M)}{P(R)} = \frac{0,08 \cdot 0,5}{0,042} = 0,95.$$

Sei (Ω, \mathcal{A}, P) ein *Wahrscheinlichkeitsraum*.

Definition Die Mengensysteme $\mathcal{A}_1, \dots, \mathcal{A}_k$ sind (stochastisch) unabhängig, falls für alle k -Tupel $A_1 \in \mathcal{A}_1, \dots, A_k \in \mathcal{A}_k$, $P(A_1 \cap A_2 \cap \dots \cap A_k) = \prod_{i=1}^k P(A_i)$.

Definition Eine Familie $(\mathcal{A}_t)_{t \in T}$ von Mengensystemen heisst unabhängig, falls die Mengensysteme $(\mathcal{A}_t)_{t \in J}$, für alle **endlichen Teilmengen** J von T , unabhängig sind.

3 Diskrete Zufallsgrößen

(Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum und E eine **abzählbare** Menge.

Definition Eine **diskrete** Zufallsgröße mit Werten in E ist eine Abbildung X von Ω in E , so dass

$$X^{-1}(\{e\}) := \{\omega \in \Omega : X(\omega) = e\} \in \mathcal{A}, \forall e \in E.$$

X ist eine reelle Zufallsgröße, falls $E \subseteq \mathbb{R}$ und ein Zufallsvektor im Falle, wo $E \subseteq \mathbb{R}^k$ ($k > 1$).

Beispiel 1 (n -maliges Werfen einer symmetrischen Münze)

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \{0, 1\}, \forall i\},$$

$$\mathcal{A} = \mathcal{P}(\Omega), \quad P(\{\omega\}) = \frac{1}{2^n} \forall \omega \in \Omega.$$

$$X(\omega) := \sum_{i=1}^n \omega_i$$

In diesem Fall ist $E = \{0, 1, 2, \dots, n\}$ und $P(X^{-1}(k)) = \binom{n}{k} \frac{1}{2^n}$, $k = 0, \dots, n$ (siehe Beispiel 2, I, § 1).

Beispiel 2 (n -maliges Werfen einer Münze: **die Binomialverteilung** $B(n, p)$)

A_i sei das Ereignis "Zahl" beim i -ten Wurf. Wir setzen voraus, dass die Familie A_1, A_2, \dots, A_n unabhängig ist. X sei wie im Beispiel 1 definiert. Da die Münze nicht unbedingt symmetrisch ist, gilt $P(X^{-1}(k)) = \binom{n}{k} p^k (1-p)^{n-k}$, wobei $p = P(A_i)$ mit $0 < p < 1$.

3.1 Die Verteilung einer Zufallsgröße

Falls X Werte in $E = \{e_1, e_2, \dots\}$ annimmt, definiert man $P_X(\{e_i\}) := P(X^{-1}(e_i))$ für $i = 1, 2, \dots$. Für eine Teilmenge A von E setzt man $P_X(A) := \sum_{e_i \in A} P_X(\{e_i\})$. Die von X induzierte Wahrscheinlichkeit P_X ist die **Verteilung** der Zufallsgröße.

Im Beispiel 2 hat man $P_X(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$ mit $E = \{0, 1, 2, \dots, n\}$. Diese Verteilung, die von zwei Parametern abhängt, spielt eine wichtige Rolle in der Wahrscheinlichkeitstheorie. Sie heisst **Binomialverteilung** ($B(n, p)$).

Die Erwartung Sei X eine reelle Zufallsgröße mit Werten in $E = \{x_1, x_2, \dots\} (\subseteq \mathbb{R})$. Die **Erwartung** von X ist definiert als

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i P(X^{-1}(x_i)) = \sum_{i=1}^{\infty} x_i P_X(\{x_i\}),$$

falls $\sum_{i=1}^{\infty} |x_i| P_X(\{x_i\}) < \infty$.

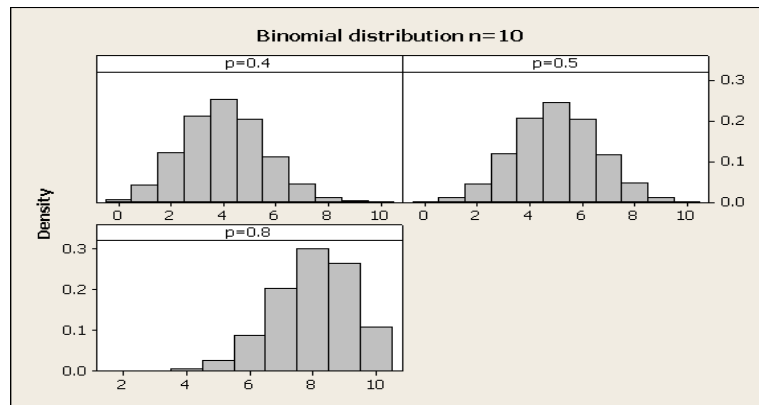


Figure 1: Die Binomialverteilung

Beispiel: Falls X eine $B(n, p)$ -Verteilung besitzt, gilt $\mathbb{E}(X) = np$:
Nach Definition ist

$$\begin{aligned}
 \mathbb{E}(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \quad (E = \{0, 1, 2, \dots, n\}) \\
 &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^n pn \cdot \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
 &= p \cdot n \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1-k)} = n \cdot p.
 \end{aligned}$$

3.2 Einige Eigenschaften der Erwartung

Satz 3.1. X, Y seien reelle Zufallsgrößen, so dass $\mathbb{E}(X)$ und $\mathbb{E}(Y)$ definiert sind. Dann gilt:

1. $X \geq 0 \implies \mathbb{E}(X) \geq 0$,
2. $\mathbb{E}(cX) = c\mathbb{E}(X)$, $\forall c \in \mathbb{R}$,
3. $X \equiv 1 \implies \mathbb{E}(X) = 1$,
4. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Beweis

Die Behauptungen 1., 2. und 3. folgen unmittelbar aus der Definition der Erwartung. Um 4. zu beweisen, zeigt man zunächst, dass $\mathbb{E}(X + Y)$ wohl definiert ist: $E = \{x_1, x_2, \dots\}$ ($F = \{y_1, y_2, \dots\}$) sei der Wertebereich von X (Y). Dann nimmt die Zufallsgröße $Z := X + Y$

Werte in $G = \{x_i + y_j : i, j = 1, 2, \dots\}$ an. Also gilt

$$\begin{aligned} & \sum_{i,j} |x_i + y_j| P\left(X^{-1}(x_i) \cap Y^{-1}(y_j)\right) \\ & \leq \sum_{i,j} |x_i| P\left(X^{-1}(x_i) \cap Y^{-1}(y_j)\right) + \sum_{i,j} |y_j| P\left(X^{-1}(x_i) \cap Y^{-1}(y_j)\right) \\ & = \sum_{i=1}^{\infty} |x_i| \sum_{j=1}^{\infty} P\left(X^{-1}(x_i) \cap Y^{-1}(y_j)\right) + \sum_{j=1}^{\infty} |y_j| \sum_{i=1}^{\infty} P\left(X^{-1}(x_i) \cap Y^{-1}(y_j)\right) \\ & = \sum_{i=1}^{\infty} |x_i| P\left(X^{-1}(x_i)\right) + \sum_{j=1}^{\infty} |y_j| P\left(Y^{-1}(y_j)\right) < \infty \end{aligned}$$

und somit existiert die Erwartung von $X + Y$.

Lässt man nun in den oberen Zeilen überall den Absolutbetrag weg, sieht man sofort, dass

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Bemerkung 1 Im Beweis hat man natürlich vorausgesetzt, dass $x_i \neq x_j$ und $y_i \neq y_j$ für $i \neq j$. Für die Zahlen $\{x_i + y_j\}$ braucht es nicht der Fall zu sein!

Bemerkung 2 Aus 4. folgt: Falls $\mathbb{E}(X_i)$ für $i = 1, 2, \dots, n$, existiert, dann existiert $\mathbb{E}(X_1 + X_2 + \dots + X_n)$ und

$$\mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n).$$

Mit Hilfe der Linearität der Erwartung lässt sich die letztere für die Binomialverteilung einfach ausrechnen: X (wie im Beispiel 2) kann man als Summe schreiben: $X = \sum_{i=1}^n Y_i$, wobei Y_i die Werte 1 (mit Wahrscheinlichkeit p) und 0 (mit Wahrscheinlichkeit $1 - p$) annimmt. $\mathbb{E}(Y_i) = 1 \cdot p + 0 \cdot (1 - p) = p \implies \mathbb{E}(X) = np$.

3.3 Unabhängige reelle Zufallsgrößen

Sei X eine Zufallsgröße mit Werten in $E = \{x_1, x_2, \dots\} \subseteq \mathbb{R}$ und $\mathcal{A}_X \subseteq \mathcal{A}$ das System aller Teilmengen von Ω , die mit Hilfe von X beschrieben werden können, d.h. alle Ereignisse der Form $X^{-1}(B)$ mit $B \subseteq E$.

Definition Die Zufallsgrößen X_1, \dots, X_k heißen (stochastisch) unabhängig, wenn die Mengensysteme $\mathcal{A}_{X_1}, \dots, \mathcal{A}_{X_k}$ unabhängig sind.

Beispiel: 2 Würfe mit einem Würfel.

Betrachten wir die Zufallsgrößen

$$X_i := \begin{cases} 1 & \text{falls beim } i\text{-ten Wurf die Augenzahl gerade ist} \\ 0 & \text{sonst} \end{cases}$$

$i = 1, 2$. Die Zufallsgrößen X_1, X_2 sind unabhängig.

Satz 3.2. Seien X, Y unabhängig. Falls $\mathbb{E}(X), \mathbb{E}(Y)$ existieren, gilt

$$\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y).$$

Beweis Seien x_1, x_2, \dots und y_1, y_2, \dots die Werte von X und Y . Dann ist

$$\sum_{i,j} |x_i y_j| P(X^{-1}(x_i) \cap Y^{-1}(y_j)) = \sum_{i,j} |x_i| |y_j| P(X^{-1}(x_i)) \cdot P(Y^{-1}(y_j))$$

wegen der Unabhängigkeit. Somit ist die summe endlich, d.h. die Erwartung von $X \cdot Y$ existiert. Weiter gilt

$$\begin{aligned} \mathbb{E}(X \cdot Y) &= \sum_{i,j} x_i y_j P(X^{-1}(x_i) \cap Y^{-1}(y_j)) \\ &= \sum_i x_i P(X^{-1}(x_i)) \cdot \sum_j y_j P(Y^{-1}(y_j)) = \mathbb{E}(X) \cdot \mathbb{E}(Y). \end{aligned}$$

3.4 Moment, Varianz aund Kovarianz

Sei X eine Zufallsgrösse und g eine reelle Funktion, die (mindestens) auf dem Wertebereich von X definiert ist. Dann ist $g(X)$ auch eine Zufallsgrösse. Falls $g(x) = x^k$, $\forall x \in \mathbb{R}$, dann heisst $E(g(X)) = \mathbb{E}(X^k)$ das **k -te Moment von X** (vorausgesetzt, dass $\mathbb{E}(|X|^k) < \infty$) und $E(g(X - \mathbb{E}(X))) = E((X - \mathbb{E}(X))^k)$ das **k -te zentrale Moment**.

Beachte: Wenn $k \leq m$ und $\mathbb{E}(|X|^m) < \infty$, dann ist auch $\mathbb{E}(|X|^k)$ **endlich**.

Beweis: Für $k \leq m$ ist $|X|^k \leq 1 + |X|^m$. Nach Satz 3.1 gilt dann $\mathbb{E}(|X|^k) \leq \mathbb{E}(1) + \mathbb{E}(|X|^m) < \infty$.

Wichtig ist das zweite zentrale Moment, die **Varianz**

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

Beachte: Für alle reellen Zahlen a, b gilt

$$\sigma^2(aX + b) = a^2 \sigma^2(X).$$

$\sigma(X)$ heisst die **Streuung** von X .

Interpretationen: Die Erwartung sagt etwas über die "**Lage**" der Zufallsgrösse, während die Streuung (Varianz) dazu dient, die **Abweichung** von der Erwartung zu charakterisieren.

Satz 3.3. (*Schwarz'sche Ungleichung*)

X, Y seien zwei reelle Zufallsgrössen.

Behauptung $\mathbb{E}(|XY|) \leq \left(\mathbb{E}(X^2) \cdot \mathbb{E}(Y^2) \right)^{1/2}$.

Beweis: $\forall \lambda \in \mathbb{R}$, gilt $\mathbb{E}(|X| + \lambda|Y|)^2 \geq 0$. Durch Satz 3.1 hat man aber

$$P(\lambda) := \mathbb{E}(|X| + \lambda|Y|)^2 = \mathbb{E}(X^2) + \lambda^2 \mathbb{E}(Y^2) + 2\lambda \mathbb{E}(|XY|)$$

und somit $\mathbb{E}(|XY|)^2 \leq \mathbb{E}(X^2) \mathbb{E}(Y^2)$.

X und Y seien zwei reelle Zufallsgrössen mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$ und $\text{Var}(X) > 0$, $\text{Var}(Y) > 0$.

Definitionen Die **Kovarianz** und die **Korrelation** zwischen X und Y sind definiert als

1. $\text{Cov}(X, Y) := E[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$,
2. $\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}$.

Bemerkung Falls X und Y unabhängig sind, gilt $\text{Cov}(X, Y) = \rho(X, Y) = 0$. Aus der Linearität der Erwartung folgt, dass $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. Nach Satz 3.2 ist $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ und deswegen $\text{Cov}(X, Y) = 0$.

Definition Zwei Zufallsgrößen X und Y sind fast sicher gleich ($X \stackrel{\text{f.s.}}{=} Y$), falls $P(\{\omega : X(\omega) \neq Y(\omega)\}) = 0$.

Satz 3.4. X, Y seien zwei Zufallsgrößen mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$, $\sigma(X) > 0$, $\sigma(Y) > 0$.

Behauptungen

1. $\rho^2(X, Y) \leq 1$,
2. $\rho(X, Y) = 1 \iff \exists a > 0, b \in \mathbb{R}$, so dass $Y \stackrel{\text{f.s.}}{=} aX + b$,
3. $\rho(X, Y) = -1 \iff \exists a < 0, b \in \mathbb{R}$, so dass $Y \stackrel{\text{f.s.}}{=} aX + b$.

Beweis

1. Die Ungleichung ist nichts anderes als die Ungleichung von Schwarz (Satz 3.3, wenn man in der letzteren X durch $X - \mathbb{E}(X)$ und Y durch $Y - \mathbb{E}(Y)$ ersetzt.

2. “ \Leftarrow ”: $\sigma^2(aX + b) = a^2\sigma^2(X)$ und $\text{Cov}(X, aX + b) = a\sigma^2(X)$. Also gilt

$$\rho(X, Y) = \frac{a\sigma^2(X)}{\sqrt{a^2\sigma^2(X)\sigma^2(X)}} = 1.$$

2. “ \Rightarrow ”: Man definiert $X' := \frac{X - \mathbb{E}(X)}{\sigma(X)}$ und $Y' := \frac{Y - \mathbb{E}(Y)}{\sigma(Y)}$. Die Korrelation lässt sich dann schreiben als $\rho(X, Y) = \mathbb{E}(X' \cdot Y')$. Nach Voraussetzung gilt also $\mathbb{E}(Y' - X')^2 = \mathbb{E}(Y')^2 + \mathbb{E}(X')^2 - 2\mathbb{E}(X' \cdot Y') = 0$ und deswegen ist $Y' - X' \stackrel{\text{f.s.}}{=} 0$, d.h.

$$Y \stackrel{\text{f.s.}}{=} \mathbb{E}(Y) + \sigma(Y) \frac{(X - \mathbb{E}(X))}{\sigma(X)} = aX + b$$

mit $a = \frac{\sigma(Y)}{\sigma(X)}$ und $b = \mathbb{E}(Y) - \frac{\sigma(Y)}{\sigma(X)} \cdot \mathbb{E}(X)$.

3. “ \Leftarrow ”: Wie oben zeigt man, dass $\rho(X, Y) = \frac{a\sigma^2(X)}{\sqrt{a^2\sigma^4(X)}}$. Also gilt

$$\rho(X, Y) = \frac{a}{|a|} = -1.$$

3. “ \implies ”: Selbe Überlegung wie oben. Man arbeitet aber mit der Summe $Y' + X'$ und zeigt, dass in diesem Falle

$$Y' + X' \stackrel{\text{f.s.}}{=} 0.$$

Satz 3.5. X_1, X_2, \dots, X_n seien unabhängige Zufallsgrößen mit $\mathbb{E}(X_i^2) < \infty$, $i = 1, 2, \dots, n$.

Behauptung $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$.

Beweis: $\text{Var}\left(\sum_{i=1}^n X_i\right) = E\left[\sum_{i=1}^n (X_i - \mathbb{E}(X_i))\right]^2 = E\left[\sum_{i,j=1}^n (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right] = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i,j=1; i \neq j}^n \text{Cov}(X_i, X_j)$. Wegen der Unabhängigkeit ist aber die Kovarianz zwischen X_i und X_j ($i \neq j$) null.

Mit Hilfe von Satz 3.5 lässt sich z.B. die Varianz der Binomialverteilung leicht ausrechnen:

Sei X $B(n, p)$ -verteilt. Dann gilt $X = \sum_{i=1}^n Y_i$, wobei Y_1, \dots, Y_n unabhängig und identisch verteilt sind (siehe Satz 3.1, Bemerkung 2).

$\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - (\mathbb{E}(Y_i))^2 = p - p^2 = p(1 - p)$ und deswegen ist $\text{Var}(X) = np(1 - p)$.

3.5 Die Faltung von Wahrscheinlichkeiten

Frage: Gegeben n unabhängige reelle Zufallsgrößen X_1, X_2, \dots, X_n mit bekannten Verteilungen $P_{X_1}, P_{X_2}, \dots, P_{X_n}$. Wie sieht die Verteilung P_X der Summe $X = \sum_{i=1}^n X_i$, die **sogenannte**

Faltung von $P_{X_1}, P_{X_2}, \dots, P_{X_n}$ aus?

Im allgemeinen (n beliebig) ist es unmöglich, die Faltung P_X auf einfache Weise auszudrücken. Deshalb betrachten wir zunächst den Spezialfall $n = 2$.

Satz 3.6. X, Y seien zwei reelle unabhängige Zufallsgrößen mit Verteilungen P_X, P_Y und Wertebereichen $E_1 := \{x_1, x_2, \dots\}$ und $E_2 := \{y_1, y_2, \dots\}$.

Sei $E := \{z_1, z_2, \dots\}$ der Wertebereich von $Z := X + Y$. (Beachte, dass $E = \{x + y : x \in E_1, y \in E_2\}$.)

Behauptung

$$P_Z(\{z_i\}) = \sum_{j=1}^{\infty} P_Y(\{z_i - x_j\})P_X(\{x_j\}) = \sum_{j=1}^{\infty} P_X(\{z_i - y_j\})P_Y(\{y_j\}).$$

Beweis

$$\begin{aligned}
 P_Z(\{z_i\}) &= P(\{\omega : Z(\omega) = z_i\}) = \sum_{j=1}^{\infty} P(\{\omega : Z(\omega) = z_i\} \cap \{\omega : X(\omega) = x_j\}) \\
 &= \sum_{j=1}^{\infty} P(\{Z = z_i\} \cap \{X = x_j\}) = \sum_{j=1}^{\infty} P(\{Z = z_i\} \mid \{X = x_j\})P(\{X = x_j\}) \\
 &= \sum_{j=1}^{\infty} P(\{Y = z_i - x_j\} \mid \{X = x_j\})P_X(\{x_j\}) \\
 &= \sum_{j=1}^{\infty} P(\{Y = z_i - x_j\})P_X(\{x_j\}) \quad \text{wegen der Unabhängigkeit.}
 \end{aligned}$$

Also gilt

$$P_Z(\{z_i\}) = \sum_{j=1}^{\infty} P_Y(\{z_i - x_j\})P_X(\{x_j\}).$$

3.6 Liste einiger wichtigen (diskreten) Verteilungen

X sei eine reelle Zufallsgrösse.

1. X besitzt eine Binomialverteilung $(B(n, p))$, falls
 - (a) X nimmt Werte in $E := \{0, 1, 2, \dots, n\}$ an,
 - (b) $P_X(\{i\}) = \binom{n}{i} p^i (1-p)^{n-i}$, $i \in E$.
(siehe III, § 1, Beispiel 2)
2. M, N, n seien positive ganze Zahlen mit $n \leq N$, $M < N$. X besitzt eine hypergeometrische Verteilung mit Parametern M, N, n , falls
 - (a) X nimmt Werte in $E := \{k : k \in \mathbb{N}, k \leq M, n-k \leq N-M\}$ an,
 - (b) $P_X(\{k\}) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$, $k \in E$.
3. Die Poisson Verteilung mit Parameter $\lambda (> 0)$.
 X besitzt eine Poisson Verteilung $\mathcal{P}(\lambda)$, falls
 - (a) X Werte in $E := \{0, 1, 2, 3, \dots\}$ annimmt,
 - (b) $P_X(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in E$.

Herleitung der Poisson Verteilung als Grenzwert von Binomialverteilungen

Wir betrachten eine gewisse Menge eines radioaktiven Elementes und ein Zeitintervall $[0, T]$.

X sei die Anzahl der radioaktiven Zerfälle im Intervall $[0, T]$. X ist eine Zufallsgrösse (empirische Tatsache) und gesucht ist eine Approximation für die Verteilung von X :

Wir dividieren das Intervall $[0, T]$ in n Teilintervalle $\{\Delta_i\}$ der gleichen Länge $\frac{T}{n}$. Für grosse Werte von n darf man annehmen, dass in jedem Intervall Δ_i ($i = 1, \dots, n$) höchstens ein Zerfall stattfindet. Ferner machen wir die folgenden Voraussetzungen:

1. Bezeichnet A_k das Ereignis, dass im Zeitintervall Δ_k ein Zerfall stattfindet, so ist die Familie A_1, A_2, \dots, A_n unabhängig.
2. \exists eine Konstante λ (die von der Substanz abhängt), so dass

$$P(A_i) = \lambda \cdot \text{Länge von } \Delta_i = \lambda \cdot \frac{T}{n}, i = 1, 2, \dots, n.$$

Unter diesen Voraussetzungen gilt:

$$P(X = k) = P_X(\{k\}) = \binom{n}{k} \left(\lambda \frac{T}{n}\right)^k \left(1 - \lambda \frac{T}{n}\right)^{n-k}, k = 0, 1, \dots, n,$$

d.h. X besitzt eine $B(n, \lambda \frac{T}{n})$ -Verteilung.

Für ein festes k lassen wir nun n gegen ∞ streben. Wir bekommen dann

$$\lim_{n \rightarrow \infty} P(X = k) = \lim_{n \rightarrow \infty} P_X(\{k\}) = e^{-\lambda T} \frac{(\lambda T)^k}{k!}.$$

Die Grenzverteilung ist also eine Poisson Verteilung mit Parameter λT .

Bemerkung: X sei $\mathcal{P}(\lambda)$ -verteilt. Dann gilt $\mathbb{E}(X) = \lambda$: Nach Definition der Erwartung ist

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda.$$

4. Die Multinomialverteilung mit Parametern n, p_1, p_2, \dots, p_k .

Diese Verteilung ist eine natürliche Verallgemeinerung der Binomialverteilung: Ein zufälliges Experiment mit mehreren möglichen Resultaten A_1, \dots, A_k wird n -mal unabhängig wiederholt. Die Wahrscheinlichkeiten $P(A_j) =: p_j$ ($j = 1, \dots, k$) der möglichen Resultate genügen dann der Bedingung $p_1 + p_2 + \dots + p_k = 1$. Wiederholt man den Versuch n -mal und bedeutet B_{n_1, n_2, \dots, n_k} das Ereignis, dass unter den n Ergebnissen n_1 -mal A_1 , n_2 -mal A_2 , \dots n_k -mal A_k auftreten, wobei $n_1 + n_2 + \dots + n_k = n$ gilt, so ist

$$P(B_{n_1, n_2, \dots, n_k}) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}.$$

Beispiel: n -maliges Werfen eines (nicht unbedingt symmetrischen) Würfels:

$$P(B_{n_1, n_2, \dots, n_6}) = \frac{n!}{n_1! \dots n_6!} p_1^{n_1} \dots p_6^{n_6}, \text{ wobei } p_i := P(\{i\}), i = 1, \dots, 6.$$

Satz 3.7. X, Y seien zwei unabhängige Zufallsgrößen mit Verteilungen $\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)$.

Behauptung Die Verteilung der Summe $Z := X + Y$, d.h. die Faltung von $\mathcal{P}(\lambda_1)$ und $\mathcal{P}(\lambda_2)$ ist die Poisson Verteilung $\mathcal{P}(\lambda_1 + \lambda_2)$.

Beweis. Nach Satz 3.6 gilt

$$\begin{aligned} P_Z(\{k\}) &= \sum_{j=0}^{\infty} P_Y(\{k-j\}) P_X(\{j\}) = \sum_{j=0}^k P_Y(\{k-j\}) P_X(\{j\}) \\ &= \sum_{j=0}^k e^{-\lambda_2} \frac{\lambda_2^{k-j}}{(k-j)!} e^{-\lambda_1} \frac{\lambda_1^j}{j!} = \frac{1}{k!} e^{-(\lambda_1 + \lambda_2)} \sum_{j=0}^k \frac{k!}{(k-j)! j!} \lambda_1^j \lambda_2^{k-j} \\ &= \frac{1}{k!} e^{-(\lambda_1 + \lambda_2)} \sum_{j=0}^k \binom{k}{j} \lambda_1^j \lambda_2^{k-j} = \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \cdot (\lambda_1 + \lambda_2)^k. \end{aligned}$$

3.7 Die Verteilungsfunktion einer Zufallsgrösse

X sei eine Zufallsgrösse. Die Verteilungsfunktion von X ist definiert als $F(u) := P(X \leq u)$. F erfüllt:

- 1) F ist monoton wachsend,
- 2) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$,
- 3) F ist von rechts stetig, d.h. $F(u+0) := \lim_{u_n \rightarrow \text{arrow} u} F(u_n) = F(u)$, denn

$$F(u_n) = P(X \leq u_n) = P_X((-\infty, u_n]) \longrightarrow P_X((-\infty, u]) = F(u),$$

da $(-\infty, u_n] \longrightarrow (-\infty, u]$.

Beachte: X nimmt höchstens abzählbar viele Werte x_1, x_2, \dots an. Deswegen ist in diesem Falle F stückweise konstant mit höchstens abzählbar vielen Sprüngen der Höhe $F(x_k) - F(x_k - 0)$ an den Stellen x_k , $k = 1, 2, \dots$. *Bemerkung:* Jeder Verteilung entspricht eine Verteilungsfunktion

3.8 Erzeugende Funktionen

Z sei eine Zufallsgrösse mit Werten in $\mathbb{Z}^+ := \{0, 1, 2, \dots\}$. Setzt man $p_k := P(Z = k)$, $k = 1, 2, \dots$, so ist die erzeugende Funktion g (oder g_Z) von Z definiert als $g(t) = \sum_{n=0}^{\infty} p_n t^n = \mathbb{E}(t^Z)$. Da $\sum_{n=0}^{\infty} p_n = 1$ ist, konvergiert die Reihe mindestens für alle t mit $|t| \leq 1$.

1. $p_n = \frac{g^{(n)}(0)}{n!}$, wobei $g^{(n)}(t)$ die n -te Ableitung von g an der Stelle t ist.
2. Für $0 \leq t \leq 1$ ist g stetig, monoton wachsend und konvex und es ist

$$g(0) = p_0 \quad g(1) = 1.$$

3. $E(Z(Z-1)\dots(Z-k+1)) = g^{(k)}(1-)$, wobei $g^{(k)}(1-) = \lim_{t \uparrow 1} g^{(k)}(t)$.

Mittels 3 lassen sich Momente von Z oft leichter berechnen als direkt aus der Verteilung. Man geht rekursiv vor:

$$\mathbb{E}(Z) = g^{(1)}(1-), \quad \mathbb{E}(Z^2) = E(Z(Z-1)) + \mathbb{E}(Z) = g^{(2)}(1-) + g^{(1)}(1-), \text{ usw..}$$

3.9 Beispiele von abhängigen Zufallsgrössen

Bei Folgen von Zufallsgrössen war bis jetzt immer die Unabhängigkeit vorausgesetzt. Z.B. war das der Fall in den Kapiteln IV und V. Nachstehend sind drei Beispiele angegeben, wo diese Voraussetzung nicht erfüllt ist. Das dritte Beispiel wird am Ende dieses Kapitels näher untersucht.

Beispiel 1 Sei X_1, X_2, \dots eine Folge unabhängiger und identisch verteilter Zufallsgrößen. Definiert man $Z_n := \sum_{i=1}^n X_i$ für $n = 1, 2, \dots$, so sind die Zufallsgrößen $\{Z_n\}$ nicht mehr unabhängig. Die schwachen Gesetze der grossen Zahlen und insbesondere der Zentralgrenzwertsatz geben uns Informationen über das Verhalten von Z_n im Falle, wo n gegen unendlich strebt.

Beispiel 2 (einfaches Warteschlangen-Modell)

Seien $0, 1, 2, \dots$ die Zeitpunkte, an denen ein Skilift, der pro Zeiteinheit eine Person befördern kann, abfährt. Zwischen den Zeitpunkten n und $n + 1$ kommen Y_n neue Skifahrer an. Die Y_n seien unabhängig. Die Länge Z_n der Warteschlange unmittelbar vor der Abfahrt zur Zeit n bestimmt sich rekursiv durch

$$Z_n = \max(0, Z_{n-1} - 1) + Y_{n-1} \quad (n \geq 1).$$

$Z_0 = i_0$ sei eine bekannte Zahl.

Beispiel 3 Galton studierte 1873 das Phänomen des Aussterbens berühmter Familiennamen. Es stellte sich die Frage nach der Wahrscheinlichkeit des Aussterbens der männlichen Linie der Nachkommenschaft eines Mannes, wenn dieser und jeder seiner Söhne, Enkel usw. unabhängig voneinander mit Wahrscheinlichkeit p_k genau k Söhne hat: Sei $Z_0 = 1$. Ist Z_n die Anzahl der männlichen Nachkommen (in männlicher Linie) in der n -ten Nachkommengeneration, und hat der j -te dieser Nachkommen $X_{n+1}^{(j)}$ Söhne, so ist $Z_{n+1} = \sum_{j=1}^{Z_n} X_{n+1}^{(j)}$.

Diese Familie $\{Z_n\}$ ist ein sogenannter *Verzweigungsprozess*.

Beachte: in diesem Falle sind die Zufallsgrößen Z_0, Z_1, Z_2, \dots *nicht* unabhängig.

Um die Frage von Galton zu beantworten, müssen wir die Folge $q_n := P(Z_n = 0)$, $n = 1, 2, \dots$ untersuchen, denn $q := \lim_{n \rightarrow \infty} q_n$ ist die gesuchte Aussterbewahrscheinlichkeit.

Heute interessiert man sich für Verzweigungsprozesse, von denen die obigen Prozesse den einfachsten Fall darstellen; natürlich nicht wegen der Familiennamen, sondern weil ähnliche Verzweigungen auch in anderen Situationen auftreten. Z.B. macht ein Neutron bei der Kernspaltung eine zufällige Zahl weiterer Neutronen frei.

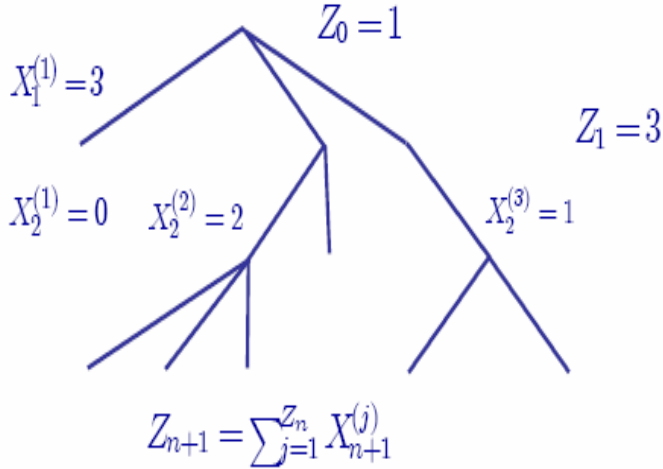
In den obigen Beispielen nehmen die Zufallsgrößen $\{Z_n\}$ Werte in $\mathbb{Z}^+ := \{0, 1, 2, \dots\}$ an. Alle Prozesse haben eine gemeinsame Eigenschaft, nämlich: für alle n und alle $i_0, i_1, \dots, i_n \in E$ gilt

$$P(Z_n = i_n \mid Z_{n-1} = i_{n-1}, \dots, Z_0 = i_0) = P(Z_n = i_n \mid Z_{n-1} = i_{n-1}).$$

Dies ist die sogenannte *Markoffsche Eigenschaft*. Die Prozesse sind dann Markoffsche Ketten (siehe z.B. Karlin: A first course in stochastic processes, Academic Press (1969); Karlin-Taylor: A second course in stochastic processes, Academic Press (1981)).

Gesucht ist die Aussterbewahrscheinlichkeit q . Da $Z_n = 0$, $Z_m = 0$ für alle $m \geq n$ impliziert, gilt $q = \lim_{n \rightarrow \infty} P(Z_n = 0) = \lim_{n \rightarrow \infty} q_n$. Die Zufallsgrößen $\{X_n^{(j)}\}$ haben alle die gleiche Verteilung, also auch die gleiche erzeugende Funktion

$$g(t) = \sum_{k=0}^{\infty} p_k t^k.$$



Bezeichnet h_n die erzeugende Funktion von Z_n , so ist wegen $P(Z_0 = 1) = 1$ natürlich $h_0(t) = t$. Ausserdem gilt

$$\begin{aligned}
h_{n+1}(t) &= h_n(g(t)) : h_{n+1}(t) = \sum_{j=0}^{\infty} P(Z_{n+1} = j) t^j = \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} P(Z_{n+1} = j, Z_n = m) t^j \\
&= \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} P\left(\sum_{\ell=1}^m X_{n+1}^{(\ell)} = j, Z_n = m\right) t^j = \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} P\left(\sum_{\ell=1}^m X_{n+1}^{(\ell)} = j\right) \cdot P(Z_n = m) t^j \\
&\quad (\text{wegen der Unabhängigkeit von } Z_n \text{ und } \{X_{n+1}^{(1)}, \dots, X_{n+1}^{(m)}\}) \\
&= \sum_{m=0}^{\infty} P(Z_n = m) \left(\sum_{j=0}^{\infty} P\left(\sum_{\ell=1}^m X_{n+1}^{(\ell)} = j\right) t^j \right) \\
&= \sum_{m=0}^{\infty} P(Z_n = m) E\left(t^{\sum_{\ell=1}^m X_{n+1}^{(\ell)}}\right) = \sum_{m=0}^{\infty} P(Z_n = m) \prod_{\ell=1}^m E(t^{X_{n+1}^{(\ell)}}) \\
&\quad (\text{wegen der Unabhängigkeit der Zufallsgrössen } X_{n+1}^{(1)}, \dots, X_{n+1}^{(m)}) \\
&= \sum_{m=0}^{\infty} P(Z_n = m) (g(t))^m \quad (\text{die Zufallsgrössen } X_{n+1}^{(1)}, \dots, X_{n+1}^{(m)} \\
&\quad \text{sind identisch verteilt mit erzeugender Funktion } g!) \\
&= h_n(g(t)).
\end{aligned}$$

Also ist $h_1(t) = g(t)$, $h_2(t) = (g \circ g)(t)$ und allgemein $h_n(t) = (g \circ g \circ \dots \circ g)(t)$ die Funktion, die man durch n -fache iterierte Anwendung der Abbildung g erhält. Da $q_n = P(Z_n = 0) = h_n(0)$, gilt also $q = \lim_{n \rightarrow \infty} h_n(0)$. Damit haben wir bei gegebenem g nur noch ein rein analytisches Problem zu lösen.

Satz 3.8. Die Aussterbewahrscheinlichkeit q ist die kleinste nicht-negative Lösung der Gleichung

$$g(t) = t.$$

Ist $g^{(1)}(1) \leq 1$ und $p_1 < 1$, so ist $q = 1$; ist $g^{(1)}(1) > 1$, so ist $q < 1$.

($g^{(1)}(1)$ ist die erwartete Zahl der männlichen Nachkommen jedes Mitgliedes der Nachkommenschaft. Der Prozess stirbt also — abgesehen vom Fall $p_1 = 1$ — mit Wahrscheinlichkeit 1 aus, wenn im Mittel höchstens 1 männlicher Nachkomme geboren wird, und sonst nur mit Wahrscheinlichkeit < 1 .)

Beweis Es gilt, wegen der Stetigkeit von g ,

$$g(q) = g(\lim h_n(0)) = \lim g(h_n(0)) = \lim h_{n+1}(0) = q.$$

q ist demnach Lösung der Gleichung $g(t) = t$. Ist $u \geq 0$ eine weitere Lösung, so ist $u = g(u) \geq g(0) = h_1(0)$, und durch Induktion folgt aus $u \geq h_n(0)$ dann $u = g(u) \geq g(h_n(0)) = h_{n+1}(0)$. Durch den Grenzübergang $n \rightarrow \infty$ ergibt sich $u \geq q$. Damit ist die erste Teilaussage bewiesen.

Ist $p_0 + p_1 = 1$, so kann in jeder Generation maximal ein männlicher Nachfahre existieren. Aus $P(Z_{n+1} = 1) = P(Z_n = 1) P(X_{n+1}^{(1)} = 1) = p_1 P(Z_n = 1)$ folgt induktiv $P(Z_n = 1) = p_1^n$. Damit gilt $q = \lim(1 - p_1^n)$. In diesem Fall ist $g^{(1)}(1) = p_1 \leq 1$. Ist $p_1 < 1$, so ist $q = 1$.

Sei also nun $p_0 + p_1 < 1$. Dann ist mindestens eines der p_k mit $k \geq 2$ positiv. $g^{(1)}(t) = \sum_{k=1}^{\infty} k p_k t^{k-1}$ ist dann auf $[0, 1)$ strikt monoton und $g(t)$ dort strikt konvex. Wir betrachten zwei Fälle:

- a) Ist $g^{(1)} \leq 1$, so ist $g'(t) < 1$ für $0 \leq t < 1$. Nach dem Mittelwertsatz muss $g(t) > t$ für $t \in (0, 1)$ sein. Also ist 1 die einzige Lösung von $g(t) = t$ und damit $q = 1$.
- b) Ist $g^{(1)}(1) > 1$, so ist $g^{(1)}(t) > 1$ für hinreichend nahe bei 1 liegende $t < 1$. In diesem Bereich ist $g(t) < t$. Da q die kleinste Lösung ist, gilt dann $0 < q < 1$, falls $p_0 > 0$. Ist $p_0 = 0$, so ist $g(0) = 0$ und also $q = 0$. \square

Numerisches Beispiel : Hier kann die zufällige Anzahl der Kinder die Werten 0, 1, und 2 mit Wahrscheinlichkeiten 0.25, 0.25 beziehungsweise 0.5 annehmen. Dann ist g durch

$$g(t) = \frac{1}{4} + \frac{1}{4}t + \frac{1}{2}t^2,$$

gegeben und die Lösung der Gleichung $t = g(t)$ ist $t = 0.5$, die Aussterbewahrscheinlichkeit der Bevölkerung ist somit 0.5 !

4 Zufallsgrößen mit Dichten

Definition (Dichte)

Eine reellwertige Funktion f heisst Dichte auf \mathbb{R}^k , falls

- a) $f \geq 0$ und
- b) $\int_{\mathbb{R}^k} f dx = 1$.

Definition (Zufallsgrößen mit Dichten)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Eine Abbildung X von Ω in \mathbb{R}^k ist eine Zufallsgröße (Zufallsvektor) mit Dichte f , falls

- a) $X^{-1}(I_1 \times I_2 \times \dots \times I_k) \in \mathcal{A}$ für jede mögliche Wahl von Intervallen I_1, \dots, I_k und
 b) $P(X \in I_1 \times I_2 \times \dots \times I_k) = P_X(I_1 \times \dots \times I_k) = \int_{I_1 \times \dots \times I_k} f(x) dx$ für alle "Rechtecke"
 $I_1 \times I_2 \times \dots \times I_k$.

Beispiel 1 (Die gleichförmige Verteilung auf dem Intervall $[0, 1]$)

Die reelle Zufallsgröße X besitzt eine gleichförmige Verteilung auf $[0, 1]$, falls seine Dichte f

durch $f(x) := \begin{cases} 1 & \text{für } x \in [0, 1], \\ 0 & \text{sonst} \end{cases}$ definiert ist.

Beispiel 2 (Die Normalverteilung $N(\mu, \sigma^2)$)

Die reelle Zufallsgröße X besitzt eine Normalverteilung $N(\mu, \sigma^2)$, $\sigma > 0$, $\mu \in \mathbb{R}$, falls ihre Dichte φ_{μ, σ^2} durch

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

$x \in \mathbb{R}$ definiert ist. Die **Standard-Normalverteilung** ist definiert durch die Dichte $\varphi := \varphi_{0,1}$.

Sei X eine $N(0, 1)$ Zufallsgröße. Die reelle Zufallsgröße

$$Y = \mu + \sigma X, \quad \mu \in \mathbb{R}, \quad \sigma > 0,$$

besitzt eine Normalverteilung $N(\mu, \sigma^2)$.

Beachte: $\varphi_{0,1}$ ist eine Dichte, denn

$$\begin{aligned} \left(\int \varphi(x) dx \right)^2 &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \cdot \int \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \int_{\mathbb{R}^2} \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}} dx dy \\ &= \frac{1}{2\pi} \int_0^\infty \left(\int_0^{2\pi} e^{-\frac{r^2}{2}} r d\varphi \right) dr \quad (\text{Polarkoordinaten}) = \int_0^\infty e^{-\frac{r^2}{2}} r dr = 1. \end{aligned}$$

Definition (Erwartung)

X sei eine reelle Zufallsgröße mit Dichte f . Die Erwartung von X ist definiert als $\mathbb{E}(X) := \int_{\mathbb{R}} x f(x) dx$, falls $\int_{\mathbb{R}} |x| f(x) dx < \infty$.

Definition X sei wie oben und g sei eine auf \mathbb{R} definierte reelle Funktion. Dann definiert man $E(g(X)) := \int_{\mathbb{R}} g(x) f(x) dx$, falls $\int_{\mathbb{R}} |g(x)| f(x) dx < \infty$.

Beachte: Damit die letzte Definition einen Sinn hat, sollte man die folgende Eigenschaft beweisen:

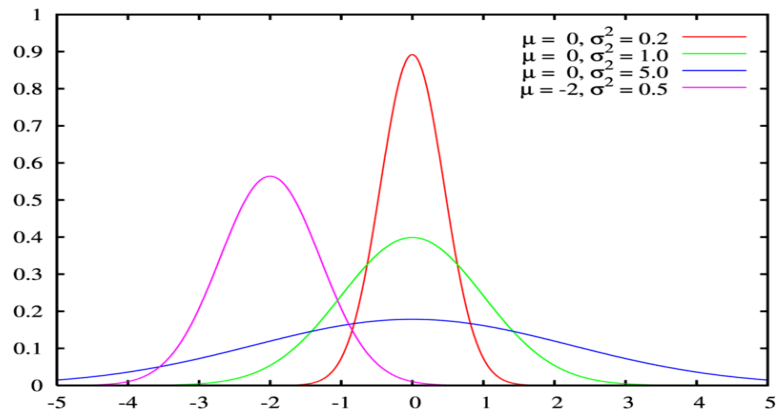


Figure 2: Normale Dichten

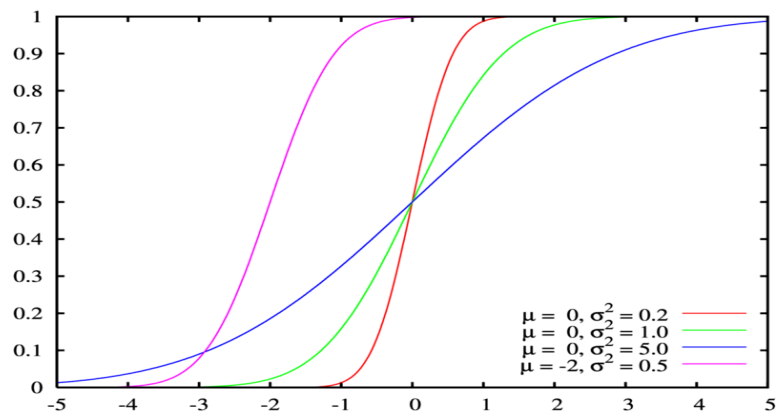


Figure 3: Gaussische Verteilungsfunktionen

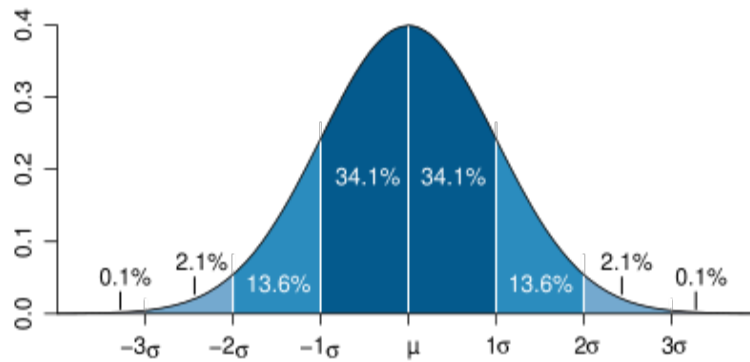


Figure 4: Die Normal $N(\mu, \sigma^2)$ Dichte

Besitzt $g(X)$ eine Dichte h , dann gilt

$$\int_{\mathbb{R}} x h(x) dx = \int g(x) f(x) dx.$$

(Ein Beweis (in einem Spezialfall) wird später angegeben.)

Definition (Varianz)

X sei eine reelle Zufallsgrösse mit Dichte f , so dass $\mathbb{E}(X^2) < \infty$. Die Varianz ist definiert als

$$\text{Var}(X) := \int (x - \mathbb{E}(X))^2 f(x) dx = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Die Streuung or Standard-Abweichung von X ist definiert als

$$\sigma(X) = \sqrt{\text{Var}(x)}.$$

Beispiel. Es ist sehr einfach zu verifizieren, dass

i) im Beispiel 1 (oben), $\mathbb{E}(X) = \frac{1}{2}$, $\text{Var}(X) = \frac{1}{12}$ und

ii) im Beispiel 2, $\mathbb{E}(X) = \mu$, $\text{Var}(X) = \sigma^2$.

Definition (Kovarianz, Korrelation)

Der Zufallsvektor $X = (X_1, X_2)$ mit Werten in \mathbb{R}^2 besitze die Dichte f . Die Kovarianz zwischen X_1 und X_2 ist definiert als

$$\text{Cov}(X_1, X_2) := \int_{\mathbb{R}^2} (x_1 - \mathbb{E}(X_1))(x_2 - \mathbb{E}(X_2)) f(x_1, x_2) dx_1 dx_2$$

und die Korrelation als

$$\rho(X_1, X_2) := \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \cdot \text{Var}(X_2)}}.$$

Beachte: Die Kovarianz ist nur dann definiert, wenn $\mathbb{E}(X_1^2) < \infty$ und $\mathbb{E}(X_2^2) < \infty$. Für die Korrelation braucht man die zusätzlichen Bedingungen $\text{Var}(X_1) > 0$, $\text{Var}(X_2) > 0$.

4.1 Unabhängige Zufallsgrössen

X_1, X_2, \dots, X_n seien n reelle Zufallsgrössen.

Definition Die Zufallsgrössen sind unabhängig, falls

$$P(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \prod_{i=1}^n P(X_i \in I_i)$$

für jede mögliche Wahl von Intervallen I_1, I_2, \dots, I_n .

Satz 4.1. Sind X_1, X_2, \dots, X_n unabhängige reelle Zufallsgrößen mit Dichten f_i , $i = 1, \dots, n$, dann besitzt der Zufallsvektor $X := (X_1, X_2, \dots, X_n)$ die Dichte $f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$.

Beweis.

$$\begin{aligned} P(X \in I_1 \times I_2 \times \dots \times I_n) &= \prod_{i=1}^n P(X_i \in I_i) = \prod_{i=1}^n \int_{I_i} f(x_i) dx_i \\ &= \int_{I_1 \times I_2 \times \dots \times I_n} \left(\prod_{i=1}^n f_i(x_i) \right) dx_1 dx_2 \dots dx_n. \end{aligned}$$

Dies gilt für alle "Rechtecke" $I_1 \times I_2 \times \dots \times I_n$. Also ist $\prod_{i=1}^n f_i(x_i)$ die Dichte von X .

Bemerkung Die Sätze 1, 2, 3, 4, 5 vom Abschnitt 1 (Diskreter Fall) sind auch für Zufallsgrößen mit Dichten gültig.

4.2 Die Verteilungsfunktion einer Zufallsgröße

X sei eine reelle Zufallsgröße mit Dichte f . Die Verteilungsfunktion von X ist definiert als $F(u) := P(X \leq u) = \int_{-\infty}^u f(v) dv$.

Die Funktion F besitzt dieselben Eigenschaften wie im diskreten Fall.

Beachte: Falls die Dichte f im Punkte u stetig ist, dann gilt $F'(u) = f(u)$.

Beispiel $X_1, X_2, X_3, \dots, X_n$ seien unabhängige Zufallsgrößen mit gleichförmiger Verteilung auf dem Intervall $[0, 1]$.

Wie sieht die Dichte von $Y := \max\{X_1, X_2, \dots, X_n\}$ aus?

Wir berechnen zunächst die Verteilungsfunktion F von Y :

$$F(u) = P(Y \leq u) = P(X_1 \leq u, X_2 \leq u, \dots, X_n \leq u) = \prod_{i=1}^n P(X_i \leq u)$$

wegen der Unabhängigkeit. Also gilt $F(u) = 0$ für $u \leq 0$, $F(u) = 1$ für $u \geq 1$ und $F(u) = u^n$ für $0 < u < 1$. Die Dichte f von Y erhalten wir, indem man F ableitet. Also ist $f(u) = nu^{n-1}$ für $0 \leq u \leq 1$ und $f(u) = 0$ sonst. Wir sind jetzt in der Lage, $\mathbb{E}(Y)$ und $\text{Var}(Y)$ auszurechnen:

$$\mathbb{E}(Y) = \int_0^1 un u^{n-1} du = \int_0^1 nu^n du = \frac{n}{n+1} u^{n+1} \Big|_0^1 = \frac{n}{n+1},$$

$$\begin{aligned}\text{Var}(Y) &= \mathbb{E}(Y^2) - \left(\frac{n}{n+1}\right)^2 = \int_0^1 u^2 n u^{n-1} du - \left(\frac{n}{n+1}\right)^2 \\ &= n \int_0^1 u^{n+1} du - \left(\frac{n}{n+1}\right)^2 = \frac{n}{n+2} u^{n+2} \Big|_0^1 - \left(\frac{n}{n+1}\right)^2 = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2.\end{aligned}$$

4.3 Die Faltung von Dichten

X, Y seien zwei reelle unabhängige Zufallsgrößen mit Dichten f, g .

Definition (Faltung)

Die Faltung der Dichten f und g ist die Dichte h der Summe $Z := X + Y$.

Satz 4.2. Die Faltung h der Dichten f und g ist gegeben durch

$$h(z) = \int_{-\infty}^{\infty} f(z-x)g(x)dx = \int_{-\infty}^{\infty} g(z-x)f(x)dx, \forall z \in \mathbb{R}.$$

Beweis. Sei $Z = X + Y$. Dann gilt

$$P(Z \leq z) = P(X + Y \leq z) = \int \int_{x+y \leq z} f(x)g(y)dx dy.$$

(Nach Satz 4.1 besitzt der Zufallsvektor (X, Y) die Dichte $f(x)g(y)$.)

Das letzte Integral kann man schreiben als

$$\begin{aligned}\int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} g(y)dy \right) f(x)dx &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^z g(v-x)dv \right) f(x)dx \\ &\stackrel{\text{Fubini}}{=} \int_{-\infty}^z \left(\int_{-\infty}^{\infty} g(v-x)f(x)dx \right) dv = \int_{-\infty}^z h(v)dv.\end{aligned}$$

Also gilt $P(Z \leq z) = \int_{-\infty}^z h(v)dv, \forall z \in \mathbb{R}$ und somit ist h die Dichte der Summe.

Definition (Chi-Quadrat Verteilung)

X_1, X_2, \dots, X_n seien unabhängige Zufallsgrößen mit Standard Normal $N(0, 1)$ Dichte φ .

Die Chi-Quadrat Verteilung mit n Freiheitsgraden ist die Verteilung der Summe

$$Y_n := \sum_{i=1}^n X_i^2.$$

Satz 4.3. Die Zufallsgrösse Y_n besitzt die Dichte

$$f_n(y) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} y^{n/2-1} e^{-\frac{y}{2}} \quad \text{für } y > 0 \quad (n = 1, 2, \dots),$$

wobei $\Gamma(p) := \int_0^\infty z^{p-1} e^{-z} dz$ ($p > 0$).

Ein Beweis kann mit Hilfe von Satz 4.2 durch Induktion geführt werden.

Die Behauptung kann auch bewiesen werden, indem man mit Polarkoordinaten arbeitet:

$$\begin{aligned} F_n(y) := P(Y_n \leq y) &= \int_{x_1^2+x_2^2+\dots+x_n^2 \leq y} \varphi(x_1)\varphi(x_2)\dots\varphi(x_n) dx_1 dx_2 \dots dx_n \\ &= \int_{x_1^2+x_2^2+\dots+x_n^2 \leq y} e^{-\sum_{i=1}^n \frac{x_i^2}{2}} dx_1 dx_2 \dots dx_n = C \int_0^{\sqrt{y}} e^{-\frac{r^2}{2}} r^{n-1} dr, \end{aligned}$$

wobei C so gewählt wird, dass $P(Y_n < \infty) = 1$. Differenziert man die Verteilungsfunktion F_n , erhält man

$$f_n(y) = C e^{-\frac{y}{2}} (\sqrt{y})^{n-1} \cdot \frac{1}{2\sqrt{y}} = C e^{-\frac{y}{2}} y^{\frac{n}{2}-1} \frac{1}{2}.$$

Es muss gelten:

$$C \int_0^\infty e^{-\frac{y}{2}} y^{\frac{n}{2}-1} \frac{1}{2} dy = 1 = C \int_0^\infty e^{-z} 2^{\frac{n}{2}-1} z^{\frac{n}{2}-1} dz$$

d.h. $C = \frac{1}{\Gamma(\frac{n}{2})2^{n/2-1}}$ und somit $f_n(y) = \frac{1}{\Gamma(\frac{n}{2})2^{n/2}} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}$.

Summe von unabhängige Normale Zufallsgrösse Seien X und Y zwei unabhängige normale Zufallsgrösse $N(\mu_1, \sigma_1^2)$, resp. $N(\mu_2, \sigma_2^2)$. Dann besitzt die Zufallsgrösse $Z = X + Y$ eine normale Dichte $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4.4 Lineare Abbildungen von Zufallsvektoren

$X := (X_1, \dots, X_n)^T$ sei ein Zufallsvektor mit Dichte $f(x_1, \dots, x_n)$.

Satz 4.4. Wenn A eine reguläre ($n \times n$)-Matrix ist, dann besitzt der Vektor $Y := AX$ die Dichte $g(y) = f(A^{-1}y) \frac{1}{|\det(A)|}$ ($y := (y_1, y_2, \dots, y_n)^T$).

Beweis. Sei $R = I_1 \times I_2 \times \dots \times I_n$ ein "Rechteck" in \mathbb{R}^n . Dann gilt:

$$P(Y \in R) = P(AX \in R) = P(X \in A^{-1}R) = \int_{A^{-1}R} f(x) dx \stackrel{x:=A^{-1}y}{=} \int_R f(A^{-1}(y)) |\det(A^{-1})| dy$$

und somit ist $f(A^{-1}(y)) \frac{1}{|\det(A)|}$ die Dichte von Y .

Spezialfall: Sind die Zufallsgrössen X_1, \dots, X_n unabhängig mit Dichte φ und ist die Matrix A orthogonal, so sind die Zufallsgrössen Y_1, Y_2, \dots, Y_n auch unabhängig mit der gleichen Dichte φ .

4.5 Funktionen von reellen Zufallsgrößen

Satz 4.5. Sei X eine reelle Zufallsgröße mit Werten in einem offenen Intervall I und Dichte $f > 0$ auf I . Sei g eine eindeutige stetig differenzierbare Funktion, die auf I definiert ist.

Behauptung Falls $g'(x) \neq 0, \forall x \in I$, dann besitzt die Zufallsgröße $Y := g(X)$ die Dichte

$$h(y) = f(g^{-1}(y)) \frac{1}{|g'(g^{-1}(y))|}.$$

Beweis. Sei J ein Intervall in $g(I)$. Dann gilt:

$$\begin{aligned} P(Y \in J) &= P(g(X) \in J) = P(X \in g^{-1}(J)) \\ &= \int_{g^{-1}(J)} f(x) dx \stackrel{y:=g(x)}{=} \int_J f(g^{-1}(y)) \frac{1}{|g'(g^{-1}(y))|} dy \quad \square \end{aligned}$$

Korollar Wenn die Voraussetzungen von Satz 4.5 erfüllt sind, dann folgt unmittelbar

$$E(g(X)) := \int_I g(x) f(x) dx = \int_{g(I)} y h(y) dy =: E(Y).$$

Beispiel Sei X gleichförmig verteilt auf dem Intervall $(0, 1)$. Wir betrachten die Funktion $Y := X^2$. Nach Satz 4.5 ist dann die Dichte h von Y :

$$h(y) = \frac{1}{2\sqrt{y}} \quad \text{für } 0 < y < 1 \quad \text{und } 0 \quad \text{sonst.}$$

Bemerkung. Für das erwähnte Beispiel ist Satz 4.5 nicht direkt anwendbar. Man kann aber den Wertebereich von X so zerlegen ($\mathbb{R} = (-\infty, 0) \cup (0, \infty)$), dass auf beiden Teilmengen die Voraussetzungen des Satzes erfüllt sind.

4.6 Zwei weitere wichtige Dichten: Die Student und die Exponential Verteilungen

(Die Student-Verteilung und die Exponentialverteilung)

1. Die Student-Verteilung

Definition Die Student-Verteilung mit n Freiheitsgraden ist die Verteilung von $U_n :=$

$$\frac{X_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}} \quad \text{wobei die Zufallsgrößen } X_0, X_1, \dots, X_n \text{ unabhängig Normal } N(0, 1) \text{ sind.}$$

Satz 4.6. ?? Die Zufallsgrösse U_n besitzt die Dichte

$$h_n(z) = \frac{1}{\sqrt{\pi} \cdot \sqrt{n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{(1 + \frac{z^2}{n})^{\frac{n+1}{2}}}.$$

Beweis. $\sqrt{\sum_{i=1}^n X_i^2}$ besitzt die Dichte

$$k_n(z) = 2z f_n(z^2) = \frac{z}{\Gamma(\frac{n}{2})} \left(\frac{z^2}{2}\right)^{n/2-1} e^{-\frac{z^2}{2}}, z > 0.$$

(f_n ist die Dichte der Chi-Quadrat Verteilung mit n Freiheitsgraden.) Der Quotient $Q_n :=$

$\frac{X_0}{\sqrt{\sum_{i=1}^n X_i^2}}$ besitzt dann die Dichte

$$r_n(u) = \int_0^\infty z k_n(z) \varphi(uz) dz = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi} \Gamma(\frac{n}{2})} \frac{1}{(1+u^2)^{(n+1)/2}},$$

(wenn man die Variablentransformation $\frac{z^2}{2}(1+u^2) = v$ benützt.) U_n ist aber gleich $\sqrt{n}Q_n$ und somit folgt die Behauptung.

Bemerkung Die Student-Verteilung mit einem Freiheitsgrad besitzt die Dichte $h_1(z) = \frac{1}{\sqrt{\pi}} \frac{1}{(1+z^2)}$. Dies ist die sogenannte *Cauchy Verteilung*. Beachte:

$$\int |z| h_1(z) dz = \infty.$$

2. Die Exponentialverteilung

Definition Eine reelle Zufallsgrösse X hat eine Exponentialverteilung mit Parameter λ ($\lambda > 0$), falls X die Dichte

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

besitzt.

Herleitung der Exponentialverteilung mit Hilfe eines Beispiels aus der Physik:

Die Atome eines radioaktiven Elementes zerfallen in zufälligen Zeitpunkten. Wie die Erfahrung zeigt, hängt die Wahrscheinlichkeit dafür, dass ein zu einem gewissen Zeitpunkt t_0 noch nicht zerfallenes Atom während des folgenden Zeitintervalls der Länge t zerfällt, nur von der Länge t dieses Zeitintervalls ab, aber nicht vom Zeitpunkt t_0 . Wir bezeichnen mit X die Lebensdauer eines Atoms und F sei ihre Verteilungsfunktion. Wenn $G(t) := 1 - F(t)$, wissen wir, dass diese Funktion monoton abnimmt und dass $G(0) = 1$.

Ferner gilt:

$$P(X \geq t + s | X \geq s) = P(X \geq t) \quad \text{für alle } t, s \geq 0,$$

d.h. $G(s+t) = G(s)G(t)$, $\forall t, s \geq 0$. Damit haben wir für die Funktion $G(t)$ eine Funktionalgleichung erhalten, aus der wir diese bestimmen können. Um die Sache zu vereinfachen, nehmen wir zunächst an, dass G im Nullpunkt differenzierbar ist. Wenn wir in $G(s+t) = G(s)G(t)$, s durch $\Delta t (> 0)$ ersetzen, bekommen wir

$$\frac{G(t + \Delta t) - G(t)}{\Delta t} = G(t) \frac{(G(\Delta t) - 1)}{\Delta t}.$$

Lässt man nun Δt gegen Null streben, so folgt

$$G'(t) = G'(0)G(t).$$

$G'(0)$ muss negativ sein, denn $G'(0) \leq 0$ (G ist monoton abnehmend). Aus $G'(0) = 0$ und $G(0) = 1$ würde $G(t) \equiv 1$ folgen; es würde also kein radioaktiver Zerfall stattfinden. Man darf daher $G'(0) = -\lambda$ mit $\lambda > 0$ setzen und als Lösung erhält man, wegen $G(0) = 1$,

$$G(t) = e^{-\lambda t}, \quad \text{d.h.} \quad F(t) = 1 - e^{-\lambda t} \quad \text{und somit} \quad f(t) := F'(t) = \lambda e^{-\lambda t}.$$

Wir werden in der Vorlesung zeigen, dass man ohne die Voraussetzung der Differenzierbarkeit von G im Nullpunkt dasselbe Ergebnis erhält.

5 Die Gesetze der grossen Zahlen

Sei X_1, X_2, X_3, \dots eine Folge von reellen Zufallsgrössen, die auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definiert sind. Sei c eine Konstante.

Definition 1 Die Folge $\{X_n\}$ konvergiert in Wahrscheinlichkeit gegen c $\left(X_n \xrightarrow[n \rightarrow \infty]{P} c\right)$, falls:

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - c| > \varepsilon) = 0.$$

Definition 2 Die Folge $\{X_n\}$ konvergiert fast sicher gegen c $\left(X_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} c\right)$, falls

$$P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = c\}) = 1.$$

Satz 5.1. Die beiden folgenden Aussagen sind äquivalent:

1. $x_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} c$
2. $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P\left(\bigcup_{j=n}^{\infty} \{|X_j - c| > \varepsilon\}\right) = 0.$

Beweis. Setzen wir $A_n := \bigcup_{j=n}^{\infty} \{|X_j - c| > \varepsilon\}$. Da $A_n \downarrow A := \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \{|X_j - c| > \varepsilon\}$, gilt (nach dem Satz) $P(A) = \lim_{n \rightarrow \infty} P(A_n) = 0$. Wir haben also

$$\begin{aligned} P\left(\bigcup_{j=n}^{\infty} \{|X_j - c| > \frac{1}{k}\}\right) &\xrightarrow{n \rightarrow \infty} 0, \forall k \in \{1, 2, 3, \dots\} \iff \\ P\left(\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \{|X_j - c| > \frac{1}{k}\}\right) &= 0, \forall k \in \{1, 2, 3, \dots\} \iff \\ P\left(\bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \{|X_j - c| > \frac{1}{k}\}\right) &= 0 \iff \\ P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} \{|X_j - c| \leq \frac{1}{k}\}\right) &= 1 \iff X_n \xrightarrow[n \rightarrow \infty]{f.s.} c. \end{aligned}$$

Korollar Wenn $X_n \xrightarrow[n \rightarrow \infty]{f.s.} c$, konvergiert die Folge in Wahrscheinlichkeit gegen c .

Beweis. $\varepsilon > 0$ sei vorgegeben. Nach Satz 5.1, $\lim_{n \rightarrow \infty} P\left(\bigcup_{j=n}^{\infty} \{|X_j - c| > \varepsilon\}\right) = 0$.

Da $P(|X_n - c| > \varepsilon) \leq P\left(\bigcup_{j=n}^{\infty} \{|X_j - c| > \varepsilon\}\right)$, folgt die Behauptung.

5.1 Die Ungleichung von Tschebyscheff

Satz 5.2. Sei X eine reelle Zufallsgrösse. Dann gilt:

$$\forall \varepsilon > 0, \quad P(\{\omega : |X(\omega)| \geq \varepsilon\}) = P(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(X^2)}{\varepsilon^2}.$$

Beweis: Für $A \subseteq \Omega$ definiert man die **Indikatorfunktion** von A als $1_A(\omega) = 1$, falls $\omega \in A$ und $= 0$ sonst.

Da $1_{\{|X| \geq \varepsilon\}} \cdot \varepsilon^2 \leq X^2$, bekommt man die Tschebyscheff'sche Ungleichung, indem man auf beiden Seiten die Erwartung nimmt.

Bemerkung 1 Falls $\mathbb{E}(X^2) < \infty$, existiert die Erwartung von X . Wenn man in der Ungleichung von Tschebyscheff X durch $X - \mathbb{E}(X)$ ersetzt, bekommt man

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

Interpretation: Je kleiner die Varianz von X ist, desto "kleiner" ist die Abweichung von der Erwartung.

Bemerkung 2 X sei $B(n, p)$ -verteilt. Dann ist

$$P(|X - np| \geq n\varepsilon) = P\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) \leq \frac{\text{Var}(X)}{\varepsilon^2 n^2} = \frac{np(1-p)}{\varepsilon^2 n^2} = \frac{p(1-p)}{\varepsilon^2 \cdot n} \leq \frac{1}{4\varepsilon^2 n},$$

d.h. $\sum_{k=0; k: |\frac{k}{n} - p| \geq \varepsilon}^n \binom{n}{k} p^k (1-p)^{n-k} \leq \frac{1}{4\varepsilon^2 n}$.

5.2 Das schwache Gesetz der grossen Zahlen

Satz 5.3. X_1, X_2, \dots sei eine Folge unabhängiger und identisch verteilter Zufallsgrössen.

Behauptung Falls $\mathbb{E}(|X_1|) < \infty$, dann gilt

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{P} \mathbb{E}(X_1), \quad \text{wobei } S_n := \sum_{i=1}^n X_i.$$

Beweis: Diesen Satz beweisen wir unter der stärkeren Bedingung $\mathbb{E}(X_1^2) < \infty$. (Der allgemeine Fall ist zu kompliziert für eine Einführungsvorlesung!)

Nach der Ungleichung von Tschebyscheff hat man

$$P\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| > \varepsilon\right) \leq \frac{\text{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} \quad \forall \varepsilon > 0.$$

Weiter gilt

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \mathbb{E}(X_1) \quad \text{und} \quad \text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \text{Var}(S_n) = \frac{1}{n} \text{Var}(X_1)$$

und somit folgt die Behauptung.

5.3 Das starke Gesetz der grossen Zahlen

Satz 5.4. (ohne Beweis)

X_1, X_2, \dots , sei eine Folge von unabhängigen identisch verteilten Zufallsgrössen. S_n sei wie im Satz 5.3 definiert.

Behauptung Falls $\mathbb{E}(|X_1|) < \infty$, dann gilt

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E}(X_1).$$

5.4 Anwendung der Gesetze der grossen Zahlen

1. Als Zufallsexperiment betrachten wir das n -malige Werfen einer symmetrischen Münze, wobei n gross ist. S_n bezeichne die Anzahl von "Kopf". S_n lässt sich schreiben als $S_n = \sum_{i=1}^n X_i$, wobei die Zufallsgrössen $\{X_j\}$ i.i.d. sind, mit $X_i = 1$ ("Kopf" beim i -ten Wurf) mit Wahrscheinlichkeit $\frac{1}{2}$ und $X_i = 0$ mit Wahrscheinlichkeit $\frac{1}{2}$. Nach dem starken Gesetz der grossen Zahlen ist $\frac{S_n}{n}$ ungefähr gleich $\mathbb{E}(X_i) = \frac{1}{2}$. Diese Aussage entspricht unserer Idee von der Stabilisierung der relativen Häufigkeit.
2. Wir betrachten eine gewisse Menge eines radioaktiven Elementes. Wir haben gesehen, dass die Lebensdauer X eines Atoms eine Zufallsgrösse ist, die eine exponentielle Verteilung besitzt, d.h. ihre Verteilungsfunktion F lässt sich schreiben als $F(t) = 1 - e^{-\lambda t}$, $t \geq 0$, wobei λ eine positive Konstante ist (die sogenannte Zerfallskonstante). Nach Definition ist die Halbwertszeit T des radioaktiven Elementes diejenige

Zeitdauer, während der ein Atom mit der Wahrscheinlichkeit $\frac{1}{2}$ zerfällt. Es muss also gelten $F(t) = \frac{1}{2}$, also $e^{-\lambda T} = \frac{1}{2}$ oder $T = \frac{\ln 2}{\lambda} = \ln 2 \cdot \mathbb{E}(X)$. Die Halbwertszeit ist somit proportional zur Erwartung der Lebensdauer ($\mathbb{E}(X) = \frac{1}{\lambda}$!).

Im Zeitpunkt $t = 0$ seien N Atome vorhanden. S_t sei die Anzahl der im Zeitpunkt $t > 0$ zerfallenen Atome. Wegen der Gesetze der grossen Zahlen, d.h. wegen des Zusammenhangs zwischen relativer Häufigkeit und Wahrscheinlichkeit, ist die relative Anzahl der Zerfälle bis zur Zeit t ungefähr gleich $1 - e^{-\lambda t}$ ($N \gg 1$). Man sieht also, dass die Halbwertszeit diejenige Zeit ist, während der ungefähr die Hälfte der Masse eines radioaktiven Elementes zerfällt.

5.5 Die Markovsche Ungleichung

Satz 5.5. Sei $f : \mathbb{R} \rightarrow [0, \infty)$. Sei X eine Zufallsvariable mit $\mathbb{E}(f(X)) < \infty$. Es gilt

$$P(f(X) > \varepsilon) \leq \frac{\mathbb{E}(f(X))}{\varepsilon}, \quad \forall \varepsilon > 0.$$

Beweis: Für $A \subseteq \Omega$ definiert man die **Indikatorfunktion** von A als $1_A(\omega) = 1$, falls $\omega \in A$ und $= 0$ sonst. Da $f(X) \geq \varepsilon 1_{f(X) \geq \varepsilon}$, bekommt man die Ungleichung

$$\mathbb{E}(f(X)) \geq \mathbb{E}(\varepsilon 1_{f(X) \geq \varepsilon}) = \varepsilon P(f(X) \geq \varepsilon).$$

Bemerkung Wenn man $f(x) = x^2$ einsetzt kriegt man wieder die Ungleichung von Tschebyscheff, da

$$\begin{aligned} \mathbb{E}(f(X)) &= \mathbb{E}(X^2) \geq \varepsilon P(X^2 \geq \varepsilon) \\ &= \varepsilon P(|X| \geq \sqrt{\varepsilon}). \end{aligned}$$

Beispiel: Sei $S_n = \sum_{i=1}^n X_i$, mit X_i unabhängige Bernoulli Zufallsvariablen mit Parameter $p = 1/2$. Die Ungleichung von Tschebyscheff angewandt auf $S_n - n/2$ besagt

$$P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}. \quad (5.1)$$

Wenn $n = 1000$ und $\varepsilon = 1/10$ ergibt das

$$P(S_{1000} \notin [400, 600]) \leq \frac{1}{40}. \quad (5.2)$$

Wir werden sehen, dass die von der Ungleichung (5.1) gegebene Schätzung nicht gut ist. Sei

$$f(x) = \exp(tx).$$

Mit der Markovschen Ungleichung gilt

$$\begin{aligned} P\left(\frac{S_n}{n} - \frac{1}{2} \geq \varepsilon\right) &= P\left(S_n - \frac{n}{2} \geq n\varepsilon\right) \\ &= P\left(\exp\left(t\left(S_n - \frac{n}{2}\right)\right) \geq \exp(tn\varepsilon)\right) \\ &\leq \frac{1}{\exp(tn\varepsilon)} \mathbb{E}\left(\exp\left(t\left(S_n - \frac{n}{2}\right)\right)\right), \end{aligned}$$

und somit

$$P(S_n - \frac{n}{2} \geq n\varepsilon) \leq \inf_{t \geq 0} \frac{\mathbb{E}(\exp(t(S_n - \frac{n}{2})))}{\exp(tn\varepsilon)}.$$

Wir benutzen die Unabhängigkeit der beteiligten Zufallsvariablen, um zu zeigen dass, mit $q = 1 - p$,

$$\begin{aligned} \mathbb{E}(\exp(t(S_n - \frac{n}{2}))) &= \prod_{i=1}^n \mathbb{E}(\exp(t(X_i - \frac{1}{2}))) \\ &= \mathbb{E}(\exp(t(X - \frac{1}{2})))^n \\ &= (p \exp(t/2) + q \exp(-t/2))^n \\ &= \cosh(t/2)^n. \end{aligned}$$

Daraus schliesst man

$$P(S_n - \frac{n}{2} \geq n\varepsilon) \leq \inf_{t \geq 0} \exp(n(\ln(\cosh(t/2)) - t\varepsilon)).$$

Wir wollen jetzt diese Ungleichung optimieren, das heisst wir suchen das Minimum der Funktion $h(t) = \ln(\cosh(t/2)) - t\varepsilon$. Es kann leicht nachgeprüft werden, dass dieses Minimum im Punkte t_ε erreicht wird, wobei

$$t_\varepsilon = \ln\left(\frac{1 + 2\varepsilon}{1 - 2\varepsilon}\right).$$

Wir betrachten die Entropiefunktion

$$\begin{aligned} I(\varepsilon) &= -h(t_\varepsilon) \\ &= \frac{1}{2}(1 + 2\varepsilon) \ln(1 + 2\varepsilon) + \frac{1}{2}(1 - 2\varepsilon) \ln(1 - 2\varepsilon). \end{aligned}$$

Man kann schreiben

$$P\left(\frac{S_n}{n} - \frac{1}{2} \geq \varepsilon\right) \leq \exp(-nI(\varepsilon)). \quad (5.3)$$

Wir machen dasselbe für die Wahrscheinlichkeit $P(S_n/n - 1/2 \leq -\varepsilon)$, so dass

$$P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| > \varepsilon\right) < 2 \exp(-nI(\varepsilon)). \quad (5.4)$$

Wenn $n = 1000$ und $\varepsilon = 1/10$, $I(\varepsilon) \approx 0.02$, und (5.4) ergibt

$$P(S_{1000} \notin [400, 600]) \leq 3.6 \cdot 10^{-9} !$$

(siehe (5.2))

6 Der zentrale Grenzwertsatz

Die wichtigsten zentralen Grenzwertsätze drücken die Tatsache aus, dass die Summe einer grossen Anzahl von unabhängigen Zufallsgrössen unter allgemeinen Bedingungen angenähert normal verteilt ist: "Sei S die Summe von vielen unabhängigen kleinen Summanden und seien $\mu := \mathbb{E}(S)$ und $\sigma^2 := \text{Var}(S)$. Dann ist $\frac{S - \mu}{\sigma}$ genähert $N(0, 1)$ verteilt". Diese Sätze decken

die Gründe dafür auf, dass man in vielen Anwendungsgebieten sehr oft normalen oder fast normalen Verteilungen begegnet. Ein typisches Beispiel hierfür sind die Ungenauigkeiten bei Messungen; der gesamte Messfehler setzt sich aus vielen verschiedenen kleinen Fehlern zusammen. Durch die zentralen Grenzwertsätze wird also die Annahme gerechtfertigt, dass die Messfehler normal verteilt sind.

Satz 6.1. (Satz von de Moivre-Laplace)

X_1, X_2, \dots sei eine Folge von i.i.d. Zufallsgrößen, wobei $X_i = 1$ mit Wahrscheinlichkeit p ($0 < p < 1$) und $X_i = 0$ mit Wahrscheinlichkeit $1 - p$. S_n sei als $S_n = \sum_{i=1}^n X_i$ definiert.

Behauptung Für alle $a, b, a \leq b$, gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left(a < \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} < b \right) &= \lim_{n \rightarrow \infty} P \left(a < \frac{S_n - np}{\sqrt{np(1-p)}} < b \right) \\ &= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_a^b \varphi(x) dx = \Phi(b) - \Phi(a), \quad \text{wobei } \Phi(u) := \int_{-\infty}^u \varphi(x) dx. \end{aligned}$$

Satz 6.2. (Der klassische zentrale Grenzwertsatz)

X_1, X_2, \dots sei eine Folge von i.i.d. Zufallsgrößen mit $\mathbb{E}(X_i^2) < \infty$. Wir setzen $\mu := \mathbb{E}(X_i)$ und $\sigma^2 := \text{Var}(X_i)$.

Behauptung Für alle $a, b, a \leq b$ gilt

$$\lim_{n \rightarrow \infty} P \left(a < \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} < b \right) = \lim_{n \rightarrow \infty} P \left(a < \frac{S_n - n\mu}{\sqrt{n\sigma^2}} < b \right) = \int_a^b \varphi(x) dx.$$

Diese Sätze beweist man normalerweise mit Hilfe von Fouriertransformationen: sei X eine reelle Zufallsgröße, dann ist die charakteristische Funktion Ψ von X definiert durch

$$\Psi(t) = \mathbb{E}(e^{itX}) := E(\cos(tX)) + iE(\sin(tX)).$$

Die charakteristische Funktion der Summe von unabhängigen Zufallsgrößen ist gleich dem Produkt der charakteristischen Funktionen:

$$E(e^{it(X+Y)}) = E(e^{itX} \cdot e^{itY}) = E(e^{itX}) \cdot E(e^{itY}).$$

In dieser Vorlesung wollen wir aber eine andere "elementare" Methode benützen. Mit dieser Methode beweisen wir den

Satz 6.3. (Satz von Ljapunoff)

Voraussetzungen Für jedes n seien X_{n1}, \dots, X_{nn} unabhängige Zufallsgrößen mit Erwartung 0 und $E(|X_{ni}|^3) < \infty, \forall i$.

Setze

$$\begin{aligned} S_n &:= X_{n1} + X_{n2} + \dots + X_{nn}, \\ \sigma_{ni}^2 &:= \mathbb{E}(X_{ni}^2) = \text{Var}(X_{ni}), \quad i = 1, 2, \dots, n, \\ \sigma_n^2 &:= \mathbb{E}(S_n^2) = \text{Var}(S_n) = \sum_{i=1}^n \sigma_{ni}^2. \end{aligned}$$

Behauptung Wenn $\frac{\sum_{i=1}^n \mathbb{E}(|X_{ni}|^3)}{\sigma_n^3} \rightarrow 0$ für $n \rightarrow \infty$, dann gilt

$$P\left(\frac{S_n}{\sigma_n} < x\right) \xrightarrow{n \rightarrow \infty} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du, \forall x.$$

Bemerkung Die Prämisse der Behauptung sorgt dafür, dass die X_{ni} "klein" sind gegenüber S_n .

Beispiel Die X_{ni} haben alle die gleiche Verteilung mit $\mathbb{E}(X_{ni}^2) = \sigma^2$, $\mathbb{E}(|X_{ni}|^3) = \gamma$. Dann gilt

$$\frac{\sum_{i=1}^n \mathbb{E}(|X_{ni}|^3)}{\sigma_n^3} = \frac{n\gamma}{(n\sigma^2)^{3/2}} = \frac{\gamma}{\sigma^3 \cdot \sqrt{n}} \xrightarrow{n \rightarrow \infty} 0.$$

Satz 6.1 ist also ein Korollar von Satz 3.

Satz 6.2 ist unter der stärkeren Voraussetzung $\mathbb{E}(|X_i|^3) < \infty$ auch ein Korollar von Satz 3.

Beweis von Satz 6.3 Ohne Beschränkung der Allgemeinheit nehmen wir $\sigma_n = 1$ an. (Falls das nicht der Fall ist, ersetzt man X_{ni} durch $\frac{X_{ni}}{\sigma_n}$, $i = 1, \dots, n$.) Die Beweisidee besteht darin, die X_{ni} durch unter sich und von den X_{ni} unabhängige normalverteilte Zufallsgrößen Y_{ni} mit den gleichen Erwartungswerten und den gleichen Varianzen σ_{ni}^2 zu ersetzen und zu zeigen, dass sich die Verteilung von S_n nur wenig von der Verteilung von $T_n := \sum_{i=1}^n Y_{ni}$ unterscheidet, welche normal $N(0, 1)$ ist.

Beachte: die Zufallsgröße Z ist $N(\mu, \sigma^2)$ verteilt, falls Z dieselbe Verteilung hat wie $\sigma X + \mu$, wobei X $N(0, 1)$ verteilt ist.

Sei f eine dreimal stetig differenzierbare Funktion mit $|f'''(x)| \leq M, \forall x$. Dann ist mit $U := X_{n1} + \dots + X_{n(n-1)}$

$$f(X_{n1} + \dots + X_{n(n-1)} + X_{nn}) = f(U) + f'(U)X_{nn} + f''(U) \cdot \frac{X_{nn}^2}{2} + r(U, X_{nn}).$$

Das Restglied $r(U, X_{nn}) = f'''(U + \eta X_{nn}) \frac{X_{nn}^3}{6}$ ist beschränkt durch $|r(U, X_{nn})| \leq \frac{M}{6} |X_{nn}|^3$.

Also gilt

$$E(f(U + X_{nn})) = E(f(U)) + E(f'(U)X_{nn}) + E\left(f''(U) \frac{X_{nn}^2}{2}\right) + E(r(U, X_{nn}))$$

und

$$E(f(U + Y_{nn})) = E(f(U)) + E(f'(U)Y_{nn}) + E\left(f''(U) \frac{Y_{nn}^2}{2}\right) + E(r(U, Y_{nn}))$$

und somit

$$|E(f(U + X_{nn})) - E(f(U + Y_{nn}))| \leq \frac{M}{6} (\mathbb{E}(|X_{nn}|^3) + E|Y_{nn}|^3),$$

denn

$$E(f'(U)X_{nn}) = E(f'(U))\mathbb{E}(X_{nn}) = 0 = E(f'(U)Y_{nn})$$

und

$$E\left(\frac{f''(U)X_{nn}^2}{2}\right) = E(f''(U))\frac{\sigma_{nn}^2}{2} = E\left(f''(U)\frac{Y_{nn}^2}{2}\right),$$

weil U, X_{nn}, Y_{nn} unabhängig sind.

Wir fahren fort und erhalten

$$\begin{aligned} & |E(f(X_{n1} + \dots + X_{n(n-1)} + X_{nn})) - E(f(X_{n1} + \dots + X_{n(n-1)} + Y_{nn}))| \\ & \leq \frac{M}{6} (\mathbb{E}(|X_{nn}|^3) + \mathbb{E}(|Y_{nn}|^3)), \\ & |E(f(X_{n1} + \dots + X_{n(n-1)} + Y_{nn})) - E(f(X_{n1} + \dots + X_{n(n-2)} + Y_{n(n-1)} + Y_{nn}))| \\ & \leq \frac{M}{6} (\mathbb{E}(|X_{n(n-1)}|^3) + \mathbb{E}(|Y_{n(n-1)}|^3)), \\ & \quad \vdots \\ & |E(f(X_{n1} + Y_{n2} + \dots + Y_{nn})) - E(f(Y_{n1} + \dots + Y_{nn}))| \\ & \leq \frac{M}{6} (\mathbb{E}(|X_{n1}|^3) + \mathbb{E}(|Y_{n1}|^3)). \end{aligned}$$

Addieren ergibt (mit Hilfe der Dreiecksungleichung)

$$|E(f(S_n)) - E(f(T_n))| \leq \frac{M}{6} \left(\sum_{i=1}^n \mathbb{E}(|X_{ni}|^3) + \sum_{i=1}^n \mathbb{E}(|Y_{ni}|^3) \right).$$

Es gilt

$$\mathbb{E}(|Y_{ni}|^3) = \sqrt{\frac{8}{\pi}} \sigma_{ni}^3 \leq \sqrt{\frac{8}{\pi}} \mathbb{E}(|X_{ni}|^3).$$

(Beweis siehe unten) und somit

$$|E(f(S_n)) - E(f(T_n))| \leq \frac{M}{6} \left(1 + \sqrt{\frac{8}{\pi}} \right) \sum_{i=1}^n \mathbb{E}(|X_{ni}|^3) =: \varepsilon_n.$$

Die rechte Seite ε_n strebt mit wachsendem n nach Voraussetzungen gegen 0. x_0 und $\delta > 0$ seien vorgegeben, aber beliebig.

a) Wähle für f eine Funktion mit

$$\begin{aligned} f(x) &= 1 & \text{für } x &\leq x_0 - \delta, \\ f(x) &= 0 & \text{für } x &\geq x_0, \\ 0 &\leq f(x) \leq 1 \quad \forall x, & |f'''(x)| &\leq M \quad \forall x. \end{aligned}$$

Dann gilt

$$\begin{aligned} \Phi(x_0 - \delta) &= P(T_n < x_0 - \delta) \leq E(f(T_n)) = E(f(S_n)) + \varepsilon_n \\ &\leq P(S_n < x_0) + \varepsilon_n, \forall n. \end{aligned}$$

Daraus folgt

$$(*) \quad \Phi(x_0 - \delta) \leq \liminf_{n \rightarrow \infty} P(S_n < x_0).$$

b) Wähle für f eine Funktion mit

$$\begin{aligned} f(x) &= 1 && \text{für } x \leq x_0, \\ f(x) &= 0 && \text{für } x \geq x_0 + \delta, \\ 0 \leq f(x) &\leq 1 \quad \forall x, && |f'''(x)| \leq M \quad \forall x. \end{aligned}$$

Dann gilt

$$\begin{aligned} P(S_n < x_0) &\leq E(f(S_n)) = E(f(T_n)) + \varepsilon_n \\ &\leq P(T_n < x_0 + \delta) + \varepsilon_n = \Phi(x_0 + \delta) + \varepsilon_n, \forall n. \end{aligned}$$

Daraus folgt

$$(**) \quad \limsup_{n \rightarrow \infty} P(S_n < x_0) \leq \Phi(x_0 + \delta).$$

c) (*) und (**) zusammen ergeben

$$\Phi(x_0 - \delta) \leq \liminf_{n \rightarrow \infty} P(S_n < x_0) \leq \limsup_{n \rightarrow \infty} P(S_n < x_0) \leq \Phi(x_0 + \delta), \forall \delta > 0.$$

Da Φ stetig ist, erhält man dann

$$\lim_{n \rightarrow \infty} P(S_n < x_0) = \Phi(x_0),$$

Um den Beweis zu vervollständigen, müssen wir noch zwei Details erledigen.

- a) Wenn $Y \sim N(0, 1)$ verteilt ist, gilt $\mathbb{E}(|Y|^3) = \sqrt{\frac{8}{\pi}}$,
- b) wenn Ψ eine konvexe Funktion ist und wenn $\mathbb{E}(|X|) < \infty$, gilt $\Psi(\mathbb{E}(X)) \leq E(\Psi(X))$ (Ungleichung von Jensen).

Beweis von a): einfache Rechnung.

Beweis von b): eine Funktion Ψ ist konvex, wenn sie in jedem Punkt x eine Stützgerade besitzt, d.h. es gibt eine *lineare* Funktion $\ell \leq \Psi$ mit $\ell(x) = \Psi(x)$. Somit, für $x := \mathbb{E}(X)$,

$$\Psi(\mathbb{E}(X)) = \ell(\mathbb{E}(X)) = E(\ell(X)) \leq E(\Psi(X)).$$

Spezialfälle:

- (i) $\Psi(x) = x^2 \implies (\mathbb{E}(X))^2 \leq \mathbb{E}(X^2)$
- (ii) $\Psi(x) = |x|^{3/2} \implies (\mathbb{E}(|X|^2))^{3/2} \leq \mathbb{E}(|X|^3)$

Im Beweis von Satz 3 haben wir den Fall (ii) benützt.

Eine Anwendung des Satzes von de Moivre-Laplace.

Gesucht ist die Wahrscheinlichkeit, bei 600 Würfeln mit einem symmetrischen Würfel mindestens 90 und höchstens 100 Sechsen zu erhalten.

S bezeichne die Anzahl von Sechsen. Gesucht ist also $P(90 \leq S \leq 100)$.

1. Genaue Lösung: Wir wissen, dass S , $B(600, \frac{1}{6})$ verteilt ist. Also gilt

$$\begin{aligned} P(90 \leq S \leq 100) &= P(S = 90) + P(S = 91) + \cdots + P(S = 100) \\ &= \sum_{k=90}^{100} \binom{600}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{100-k}. \end{aligned}$$

2. Eine gute Schätzung für die gesuchte Wahrscheinlichkeit: Nach Satz 1, mit $n = 600$ und $p = \frac{1}{6}$, wissen wir, dass

$$\begin{aligned} P\left(a \leq \frac{S - \mathbb{E}(S)}{\sqrt{\text{Var}(S)}} \leq b\right) &= P\left(a \leq \frac{S - 600 \cdot \frac{1}{6}}{\sqrt{600 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \leq b\right) \\ &\approx P\left(a \leq \frac{S - 100}{9,13} \leq b\right) \approx \Phi(b) - \Phi(a), \quad \text{wobei} \quad \Phi(u) := \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Somit gilt

$$\begin{aligned} P(90 \leq S \leq 100) &= P\left(\frac{90 - 100}{9,13} \leq \frac{S - 100}{9,13} \leq \frac{100 - 100}{9,13}\right) \\ &\approx \Phi(0) - \Phi(-1,095) = 0,5 - (1 - \Phi(1,095)) \approx 0,36, \end{aligned}$$

wobei der Wert $\Phi(1,095) \cong 0,86$ der in der Vorlesung verteilten Tafel entnommen wurde. (Wir haben die Tatsache benützt, dass $\Phi(-x) = 1 - \Phi(x)$.)