Stochastik I

Wintersemster 2018/19

Thorsten Schmidt

Inhaltsverzeichnis

1. Einführung

Die Stochastik wird in vielen Bereichen angewendet. Auf den ersten Blick erscheinen mir die folgenden Punkte am wichtigsten (insbesondere für die Hörer dieser Vorlesung):

1. Beurteilung von medizinischen, wissenschaftlichen Hypothesen (Zulassung von Medikamenten, Qualität von Bauteilen, Mobilfunk: Nachweis der Unschädlichkeit ?)
2. Lesen und Interpretieren von Statistiken
3. Finanz- und Versicherungsmathematik
4. Steuer- und Regelung (Google Cars, Apollo-Raumfähre)
5. Grundlage von maschinellem Lernen (Google Voice, Siri, ...)

Wir beginnen mit einem Beispiel.

B 1.1 ***Fahrerflucht*:** Ein Zeuge beobachtet ein Taxi, welches ein Auto beschädigt und davonfährt. Er gibt an, ein blaues Taxi gesehen zu haben. Es gibt blaue und grüne Taxis in der Stadt. Bei einem Test identifiziert der Zeuge mit 80 %-iger Wahrscheinlichkeit die richtige Farbe. Was sagen Sie dazu ?

Überraschend ist: Wir haben zunächst gar nicht alle Informationen! Wir benötigen noch die Anzahl Taxis. In diesem Fall gibt es 25 grüne und 5 blaue Taxis in der Stadt. Folgende Tabelle stellt die möglichen Fälle dar.

	Zeuge: „Blau“	Zeuge: „Grün“
Taxi blau	4	1
Taxi grün	5	20

$\downarrow$

9 mal blau, davon $\frac{4}{9}$ blau und $\frac{5}{9}$ **grün**

Tabelle 1.1: Taxis

Etwas überraschend - wäre uns nicht zufällig die Einsicht über den Weg geflattert, dass wir noch mehr Informationen brauchen, wäre gar nicht aufgefallen, dass wir einen Fehler gemacht haben. Dies führt zu folgender Einsicht: Wir benötigen einen präzisen Formalismus, um Fehler zu vermeiden!

Es wird sich erstaunlicherweise herausstellen, dass der Zufall mit mathematischen Formalismus exakt beschreibbar ist.

Als ergänzende Literatur eignet sich das hervorragende Skriptum

- *Wahrscheinlichkeitsrechnung und Statistik* von Hans Föllmer, Hansruedi Künsch mit Ergänzungen von Josef Teichmann. Es ist zu finden unter `ftp://ftp.stat.math.ethz.ch/Teaching/kuensch/skript-einf.pdf`

Es diente als Vorlage zu diesem Vorlesungsskript, geht aber deutlich über den hier präsentierten Stoff hinaus (es handelt sich ja hier nur um eine zweistündige Vorlesung).

Bücher zur Wahrscheinlichkeitstheorie gibt es viele, es lohnt sich in alle einmal einen Blick zu werfen und dann ein oder zwei Favoriten zum genauen Studium zu verwenden. Ich selbst habe im Studium nach dem Bauer gelernt, eine knappe und präzise Fassung. Mittlerweile wird auch der Georgii sehr gerne verwendet. Es gibt auch noch ein paar neuere Werke und die englischsprachige Literatur habe ich hier größtenteils ausgeklammert. Über diese kurze und unvollständige Liste hinaus gibt es noch viele hervorragende Werke - welche aber mit den einschlägigen Suchmaschinen leicht zu finden sind.

- Bauer (1990). Wahrscheinlichkeitstheorie. Ein Klassiker.
- Georgii (2015). Stochastik (5. Aufl.) Gut zu lesen und enthält auch einen Teil zur Statistik.
- Tappe (2013). Einführung in die Wahrscheinlichkeitstheorie. Hier erhält man gut dargestellt das Hintergrundwissen zur Maßtheorie und den fehlenden Beweisen zu den Grenzwertsätzen.
- Henze (2011). Stochastik für Einsteiger. Mit vielen motivierenden Beispielen.
- Chung (2001). A Course in Probability Theory.
- Jacod, Protter (2004). Probability Essentials. Sehr kondensierte Version von zwei Meistern des Fachs geschrieben.

Diese Vorlesung wurde, so wie hier, in Freiburg im WS 2018/19 gehalten. Das Skript ist sicher noch mit vielen Fehler behaftet und ich würde mich bei gefundenen Fehlern sehr über eine Nachricht freuen. Mein Dank gilt den Studierenden, die sich von dem Thema inspirieren ließen wodurch die Vorlesung sehr viel Spaß gemacht hat, und Monika Hattenbach für eine erste Version dieses Skriptes. (März 2019, Thorsten Schmidt)

Literaturverzeichnis

Bauer, H. (1990). *Wahrscheinlichkeitstheorie.* Walter de Gruyter, Berlin.

Bickel, P. J. und K. A. Doksum (2001). *Mathematical Statistics: Basic Ideas and Selected Topics Vol. I* (2nd ed.). Prentice Hall.

Casella, G. und R. L. Berger (2002). *Statistical Inference* (2nd ed.). Duxbury. Pacific Grove.

Chung, K. L. (2001). *A Course in Probability Theory.* Academic Press.

Czado, C. und T. Schmidt (2011). *Mathematische Statistik.* Springer Verlag Berlin Heidelberg

Georgii, H.-O. (2015). *Stochastik* (5th ed.). Walter de Gruyter. Berlin.

Henze, N. (2011). *Stochastik für Einsteiger: eine Einführung in die faszinierende Welt der Zufalls* Vieweg + Teubner

Irle, A. (2005). *Wahrscheinlichkeitstheorie und Statistik.* B. G. Teubner Verlag.

Jacod, J. und P. Protter (2004). *Probability Essentials.* Springer Verlag. Berlin Heidelberg New York.

Klenke, A. (2008). *Wahrscheinlichkeitstheorie* (2nd ed.). Springer Verlag. Berlin Heidelberg New York.

Lehmann, E. L. und G. Casella (1998). *Theory of Point Estimation* (2nd ed.). Springer Verlag. Berlin Heidelberg New York.

Lehmann, E. L. und J. P. Romano (2006). *Testing Statistical Hypotheses* (corr. 2nd printing ed.). Springer, New York.

Robert, C. P. und G. Casella (2008). A history of Markov chain monte carlo – subjective recollections form incomplete data. *Technical Report, University of Florida*.

Shao, J. (2008). *Mathematical Statistics.* Springer Verlag. Berlin Heidelberg New York.

Tappe, S. (2013). *Einführung in die Wahrscheinlichkeitstheorie.* Springer Verlag Berlin Heidelberg

2. Der Wahrscheinlichkeitsbegriff

Die zentrale Idee ist es, Zufallsvariablen als Abbildungen aufzufassen: Etwa $X : \Omega \to \mathbb{R}^n$, auch allgemeinere Wertebereich anstelle von $\mathbb{R}^n$ ist möglich. Der Raum Ω heißt **Grundraum**. Wieder starten wir mit ein paar Beispielen.

B 2.1 ***Grundräume*:** Als Beispiele hierfür ist möglich:

(i) Beim Würfelwurf wählen wir $\Omega = \{1, 2, 3, \dots, 6\}$.

(ii) Unendlich viele Würfe einer Münze können durch $\Omega = \{0, 1\}^{\mathbb{N}}$ abgebildet werden. (Menge aller binären Folgen)

(iii) Bewegung eines mikroskopischen Teilchens auf einer Flüssigkeit $\Omega = \mathscr{C}(\mathbb{R}_+, \mathbb{R}^2)$ (das sind stetige Abbildungen von $\mathbb{R}_+$ nach $\mathbb{R}^2$).

Neben dem Grundraum gibt es eine Menge beobachtbarer **Ereignisse** $\mathscr{A} \subset \mathscr{P}(\Omega)$ wobei $\mathscr{P}(\Omega)$ die Menge aller Teilmengen von Ω ist. Für ein Ereignis $A \in \mathscr{A}$ sagen wir: A tritt ein, falls das realisierte Ergebnis $\omega \in \Omega$ ein Element von A ist. Ein paar Beispiele folgen.

B 2.2 ***Ereignisse*:** Folgende Beispiele sind mögliche Ereignisse:

(i) Resultat ist eine gerade Zahl: $A = \{2, 4, 6\}$.

(ii) Der dritte Wurf ist Kopf: $A = \{\omega \in \Omega : \omega_3 = 1\}$.

(iii) Das Teilchen bleibt in einem Kreis mit Radius r: Hierfür wählen wir $\Omega = \mathscr{C}(\mathbb{R}_+, \mathbb{R}^2)$ und

$$A = \Big\{x \in \Omega : \sup_{t \in \mathbb{R}_+} \|x(t)\| \le r\Big\}.$$

Mit Operationen der Mengenlehre ($\cap, \cup$, Komplement) kann man nun neue Ereignisse aus bestehenden Ereignissen bilden. Darüber hinaus ist das entscheidende Konzept ein **Wahrscheinlichkeitsmaß**: Dies ist eine Abbildung $P : \mathscr{A} \to [0, 1]$, welche jedem Ereignis $A \in \mathscr{A}$ eine Wahrscheinlichkeit $P(A)$ zuordnet.

Es gibt verschiedene Motivationen und Interpretationen für das Wahrscheinlichkeitsmaß:

(i) *Subjektiv*: Man könnte Ereignissen subjektiv gewisse Wahrscheinlichkeiten zuordnen, etwa den Ausgang einer Wahl oder für das Eintreten von Regen am Folgetag usw. Die folgenden Regeln lassen sich dann hierauf anwenden und man kann bestimmte Schlussfolgerungen ableiten.

(ii) *Frequentistisch*: Aufgrund von wiederholten Experimenten kann man relative Häufigkeiten bestimmen und diese als Wahrscheinlichkeiten betrachten. In der Tat ist dies eine Hauptmotivation, denn wir werden später zeigen, dass relative Häufigkeiten von unabhängigen und identisch verteilten Zufallsvariablen sogar gegen die wahren (und unbekannten) Wahrscheinlichkeiten konvergieren.

(iii) *Gleichverteilung*: Hat man keine Vorinformation, so kann man alle ω als gleich wahrscheinlich ansehen. Dies führt zur Gleichverteilung und zu den Laplace-Experimenten.

2.1 Diskrete Wahrscheinlichkeitsräume

Der Grundraum Ω sei abzählbar oder endlich. Wir setzen

$$P(\omega) := P(\{\omega\}) \in [0,1] \tag{2.3}$$

für jedes $\omega \in \Omega$. Als Normierung nutzen wir folgende Forderung:

$$1 = P(\Omega) = \sum_{\omega\in\Omega} P(\omega) \tag{2.4}$$

und setzen $\mathscr{A} := \mathscr{P}(\Omega)$, sowie

$$P(A) = \sum_{\omega\in A} P(\omega), \quad A \in \mathscr{A}.$$

Für paarweise disjunkte $A_1, A_2, \dots (A_i \cap A_j = \varnothing$ für $i \neq j)$ ist dann

$$P\Big(\bigcup_i A_i\Big) = \sum_{\omega\in\bigcup_i A_i} P(\omega) = \sum_i \sum_{\omega\in A_i} P(\omega) = \sum_i P(A_i), \tag{2.5}$$

die Wahrscheinlichkeiten addieren sich!

In der allgemeinen Theorie verwenden wir später (2.4) und (2.5) als Axiome. Auf einer abzählbaren Menge hat jede Abbildung P, die (2.4) und (2.5) erfüllt, die Gestalt (2.3).

Man zeigt leicht, dass folgende **Rechenregeln** gelten:

$$P(A^c) = 1 - P(A) \tag{2.6}$$

$$P(A\cup B) = P(A) + P(B) - P(A\cap B) \tag{2.7}$$

$$\mathbb{P}\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{k=1}^n (-1)^{k+1} \sum_{1\le i_1<\cdots<i_k\le n} P(A_{i_1}\cap\cdots\cap A_{i_k}) \tag{2.8}$$

$$A\subseteq B \Rightarrow P(A) \le P(B) \tag{2.9}$$

B 2.10 ***Poisson-Verteilung*:** Bezeichne $\Omega = \{0, 1, 2, \dots\}$ die Anzahl Anrufe bei einem Callcenter in einer festen Zeitspanne und gelte

$$P(\omega) = e^{-\lambda}\frac{\lambda^\omega}{\omega!}, \quad \omega \in \Omega,$$

mit Parameter $\lambda > 0$. Dann gilt $\sum_{\omega\in\Omega} P(\omega) = 1$ und die Wahrscheinlichkeit, dass mindestens ein Anruf eintrifft ist

$$P(\{1, 2, \dots\}) = 1 - P(\{0\}) = 1 - e^{-\lambda}.$$

2.1.1 Erwartungswert

Wir betrachten die Zufallsvariable

$$X : \Omega \to \mathbb{R}.$$

Hierbei lassen wir ein allgemeineres Ω zu, nehmen aber an, dass $X(\Omega)$ abzählbar ist. Dann heißt X ebenfalls **diskret**.

Jeder Zufallsvariablen X ordnen wir den Erwartungswert

$$E[X] = \sum_{\omega\in\Omega} X(\omega)P(\omega) \tag{2.11}$$

zu. Für eine Funktion von X erhalten wir, dass $f(X)$ wieder eine diskrete Zufallsvariable ist und demnach

$$E[f(X)] = \sum_{\omega\in\Omega} f(X(\omega))P(\omega).$$

Hierbei müssen wir sicherstellen, dass (2.11) sinnvoll ist. Das ist z.B. der Fall, wenn $X \geq 0$ ist (!). Nimmt X positive und negative Werte an, so zerlegen wir $X = X^+ - X^-$ mit $X^+ = \max(X, 0)$ und $X^- = (-X)^+$ und setzen

$$E[X] = E[X^+] - E[X^-],$$

sofern nicht beide Summen unendlich sind, d. h. $\min(E[X^+], E[X^-]) < \infty$. In diesem Fall folgt die wichtige Rechenregel

$$E[X] = \sum_{x\in X(\Omega)} \sum_{\omega:X(\omega)=x} X(\omega)P(\omega) = \sum_{x\in X(\Omega)} x \cdot P(X = x). \tag{2.12}$$

Durch die Gewichtung

$$X(\Omega) \ni x \to P(X = x) := P(\{\omega \in \Omega \mid X(\omega) = x\})$$

ist wieder ein Wahrscheinlichkeitsmaß gegeben (erfüllt (2.4) und (2.5)) und heißt **Verteilung** von X. Wir setzen $\mu(x) = P(X = x)$.

B 2.13 ***Erwartungswert einer Poisson-Verteilung*:** In Beispiel 2.10 wählen wir $X(\omega) = \omega$ und erhalten

$$\begin{aligned} E[X] = \sum_{x=0}^{\infty} x \cdot P(X = x) &= \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{(x-1)} \cdot \lambda}{(x-1)!} = \lambda, \end{aligned}$$

d. h. der Parameter λ ist gerade die erwartete Anzahl der Anrufe.

B 2.14 ***Versicherung*:** In einem einfachen Versicherungskontrakt ist die Leistung der Versicherung gerade

$$X = \begin{cases} c, & \text{falls Ereignis A eintritt,} \\ 0, & \text{sonst,} \end{cases}$$

also zufällig. Die Prämie ist nicht zufällig und ein erster Ansatz für die Prämie ist

$$E[X] = c \cdot P(A) + 0 \cdot P(A^c) = c \cdot P(A).$$

Man wird also $c \cdot P(A)$ verlangen, zuzüglich eines Aufschlags. $\diamond$

Aus der Definition des Erwartungswertes ergibt sich

$$E[aX + bY] = aE[X] + bE[Y], \quad a, b \in \mathbb{R}, \tag{2.15}$$

sofern beide Seiten sinnvoll sind. ($\to$ Lineare Algebra!)

Lemma 2.16. *Ist $X(\Omega) \subseteq \{0, 1, 2, \dots\}$, so gilt*

$$E[X] = \sum_{n=0}^{\infty} P(X > n).$$

Beweis: Es gilt

$$\begin{aligned} \sum_{n=0}^{\infty} P(X > n) \stackrel{(2.5)}{=} \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} P(X = k) &= \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} P(X = k) \\ &= \sum_{k=1}^{\infty} k \cdot P(X = k) \stackrel{(2.12)}{=} E[X]. \end{aligned}$$

$\square$

2.1.2 Laplace-Modelle

In diesem Kapitel betrachten wir ein endliches Ω. Das ist zunächst in vielen Fällen der Fall. Die meisten Messungen etwa sind diskret, etwa in Zentimeter, so dass man mit einem diskreten Wahrscheinlichkeitsmodell bereits viele Fälle modellieren kann.

Für jedes $\omega \in \Omega$ muss man nun eine Wahrscheinlichkeit angeben. Wie in der Einleitung betont, ist ein oft verwendetes Modell dasjenige der Gleichverteilung: alle ω haben die gleiche Wahrscheinlichkeit. Das ist etwa beim Würfelwurf der Fall, oder wenn man keine Vorinformation hat, man zum Beispiel zum ersten Mal ein Butterbrot herunterfallen lässt um zu testen, auf welcher Seite es landet. Dies nennt man ein *Laplacesches Modell*, also

$$P(\omega) = \text{const.} \tag{2.17}$$

Wegen der Normierung können wir P leicht ausrechnen und es ist

$$P(\omega) = \frac{1}{|\Omega|}. \tag{2.18}$$

Für ein Ereignis $A \in \mathscr{A}$ erhalten wir die bekannte Regel

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl möglicher Fälle}}. \tag{2.19}$$

Hierbei bezeichnet $|A|$ die Anzahl der Elemente von A. Man nennt dieses P die **Gleichverteilung** auf Ω oder das Modell ein *Laplacesches Modell*.

B 2.20 ***Garderobenproblem (Montmort 1708):*** n Mäntel werden zufällig an n Personen verteilt. Wie hoch ist die Wahrscheinlichkeit, dass keine Person ihren Mantel bekommt?

Als Modell wählen wir die Menge aller Permutationen von $\{1, \dots, n\}$, $\Omega = S_n$ und P die Gleichverteilung auf Ω. Sei

$$A_i = \{\omega \in \Omega : \omega(i) = i\} \qquad i\text{-te Person bekommt ihren Mantel}$$

und $A = \left(\bigcup_{i=1}^n A_i\right)^c$ das Ereignis von Interesse.

$$\begin{aligned} P(A^c) = P\Big(\bigcup_{i=1}^n A_i\Big) &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k < n} \underbrace{P(A_{i_1} \cap \dots \cap A_{i_k})}_{=\frac{(n-k)!}{n!}} \\ &= \sum_{k=1}^n (-1)^{k+1} \underbrace{\binom{n}{k} \frac{(n-k)!}{n!}}_{=(k!)^{-1}} = -\sum_{k=1}^n \frac{(-1)^k}{k!}. \end{aligned}$$

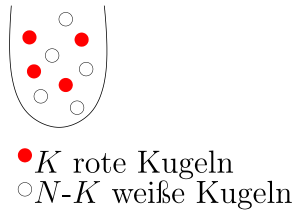
Es folgt

$$P(A) = 1 + \sum_{k=1}^n \frac{(-1)^k}{k!} \xrightarrow[n\to\infty]{} e^{-1} \approx 0{,}37.$$

B 2.21 ***Zufällige Stichprobe (Meinungsumfragen etc.):*** In einer Urne befinden sich K rote und $N-K$ weiße Kugeln. Es wird eine Stichprobe von n Kugeln gezogen. Nach dem Ziehen einer Kugel kann man diese entweder wieder zurücklegen, oder nicht. Bezeichnet ω_i die i-te Kugel, so ist der entsprechende Grundraum im Fall

mit Zurücklegen: $\Omega_1 = \{(\omega_1, \dots, \omega_n) : 1 \leq \omega_i \leq N\}$

ohne Zurücklegen: $\Omega_2 = \{(\omega_1 \dots, \omega_n) : 1 \leq \omega_i \leq N, \omega_i \neq \omega_j, i \neq j\}$



Wir bezeichnen mit P_i die Gleichverteilung auf Ω_i und berechnen die Verteilung der Zufallsvariable

$$X = \text{ Anzahl roter Kugeln in der Stichprobe.}$$

Sei $A_i^{(k)} = \{X = k\} = \{\omega \in \Omega_i : 1 \leq w_j \leq K \text{ für genau } k \text{ Indizes } j\}$ das Ereignis von Interesse. Dann ist

$$P_i(X = k) = P(A_i^{(k)}) = \frac{|A_i^{(k)}|}{|\Omega_i|}, \quad i = 1, 2.$$

Wir unterscheiden die beiden Fälle:

- *Mit Zurücklegen:* Es ist $|\Omega_1| = N^n$, und $|A_k| = K^k(N-K)^{n-k}\binom{n}{k}$, also

$$P_1(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2.22}$$

  mit $p = {}^K\!/_N$. Diese Verteilung heißt **Binomialverteilung**.

- *Ohne Zurücklegen:* Nun ist $|\Omega_2| = N(N-1)\dots(N-n+1)$, und man erhält

$$|A_k| = \binom{n}{k} K \cdot (K-1) \cdots (K-k+1) \cdot (N-K) \cdots (N-K-(n-k)+1).$$

  Es folgt

$$P_k(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}. \tag{2.23}$$

  Diese Verteilung heißt **hypergeometrische Verteilung**.

Für $N \to \infty$, $K \to \infty$, $\frac{K}{N} \to p$ konvergiert die hypergeometrische Verteilung gegen die Binomialverteilung. (Übungsaufgabe)

Möchte man die Reihenfolge der Ziehungen vernachlässigen, so wählt man folgende Grundräume:

$$\Omega_3 = \{(\omega_1, \dots, \omega_n) : 1 \leq \omega_1 \leq \omega_2 \leq \dots \leq N\},$$
$$\Omega_4 = \{(\omega_1, \dots, \omega_n) : 1 \leq \omega_1 < \omega_2 < \dots \leq N\}.$$

Ob man die Gleichverteilung auf Ω_2 oder Ω_4 ändert, spielt keine Rolle, wohl aber bei Ω_1 oder Ω_3: $(1,\dots,1) \in \Omega_3$ entspricht nur $(1,\dots,1) \in \Omega_1$, $(1,2,\dots,n) \in \Omega_3$ entspricht aber $n!$ verschiedenen Elementen in Ω_1.

Man kann diese Verteilungen sehr leicht in R[1] simulieren.

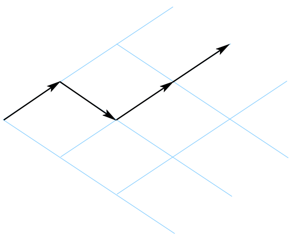
R-Code 1.

```
# Example 0 - Binomial distribution

k=0:10
plot(k,dbinom(k,10,0.5))
points(k,dbinom(k,10,0.3),col="red")

#Do it yourself: dhyper !
```

2.1.3 Die Irrfahrt



Eine Bewegung auf $\mathbb{Z}$, welche in jedem Schritt sich nur $+1$ oder -1 bewegt.

Wir modellieren dies mit $\Omega = \{\omega = (\omega_1,\dots,\omega_N) : \omega_i \in \{-1,+1\}\} =: \{-1,+1\}^N$ und setzen

$$X_k(\omega) = \omega_k \quad \text{und} \quad S_n = \sum_{k=1}^{n} X_k(\omega).$$

Die Zufallsvariable X_k ist die Bewegung im k-ten Schritt und S_n ist die Summe der ersten Bewegungen, gibt als die Stelle an, an der sich die Irrfahrt im Zeitpunkt n befindet. Wir starten in 0, d. h. $S_0(\omega) = 0$.

Sei P die Gleichverteilung auf Ω, also

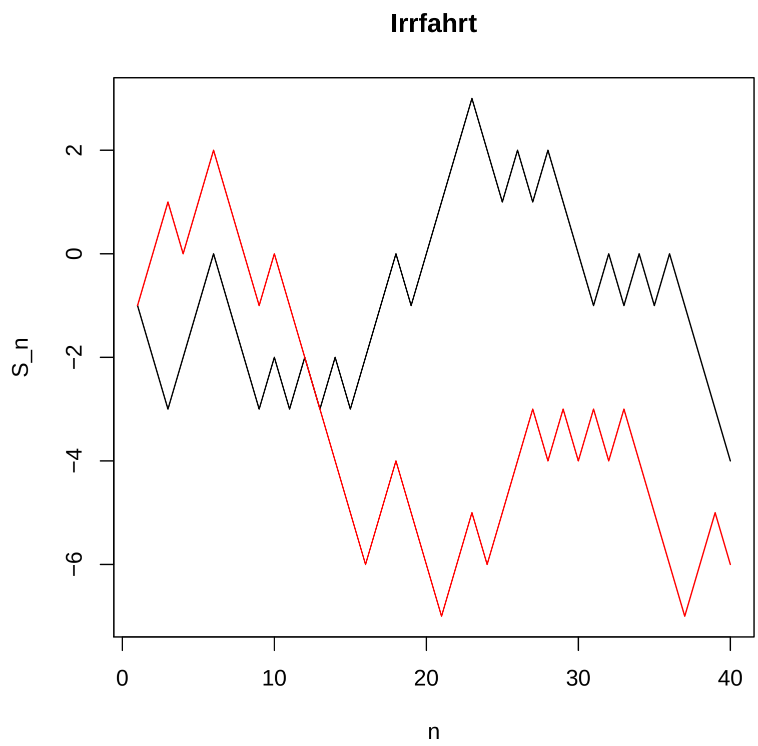
$$P(\omega) = \frac{1}{2^N} \quad \text{und} \quad \mathscr{A} = \mathscr{P}(\Omega). \tag{2.24}$$

Definition 2.25 (Irrfahrt). Sei $\Omega = \{-1,+1\}^N$ und Die Folge $(S_n)_{n=0,1,\dots,N}$ auf dem Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, P)$ nach (2.24) heißt **Irrfahrt** mit Start in 0.

[1] R ist eine Open-Source Software, zu beziehen über www.r-project.org .

R-Code 2 (Irrfahrt)**.**

```
N=40
omega1 = sample(c(-1,1),size=N,replace=TRUE)
omega2 = sample(c(-1,1),size=N,replace=TRUE)
# Der Pfad ist die kumulative Summe
S1 = cumsum(omega1)
S2 = cumsum(omega2)
# Fuer den Plot ist es nuetzlich min und max zu kennen
ymin = min(c(S1,S2))
ymax = max(c(S1,S2))
# Plot
plot(S1,type="l",main="Irrfahrt",ylim=c(ymin,ymax),xlab="n",ylab="S_n")
lines(S2,col="red")
# ---- Man kann auch unterschiedliche N probieren: 100,1.000,10.000
```



Aus (2.24) folgt, dass

$$P(X_k = 1) = \frac{2^{N-1}}{2^N} = \frac{1}{2}, \quad k = 1, \dots, N.$$

Die ersten n Schritte bilden eine Irrfahrt mit n Schritten. Es gilt

$$E[X_k] = P(X = 1) \cdot 1 + (-1) \cdot \mathbb{P}(X = -1) = 0 \quad \text{und}$$

$$E[S_n] \underset{(2.15)}{=} \sum_{k=1}^{n} E[X_k] = 0, \quad n = 0, \dots, N.$$

Satz 2.26 (Irrfahrt). *Für alle $k = 0, \dots, n$ gilt,*

$$P(S_n = 2k - n) = \binom{n}{k} 2^{-n},$$

sonst gilt $P(S_n = k) = 0$. Die Verteilung von S_n ist also eine „linear transformierte" Binomialverteilung mit $p = \frac{1}{2}$.

Beweis: Sei U_n der Anzahl der „ups"

$$U_n = \sum_{k=1}^{n} \mathbb{1}_{\{X_k = 1\}},$$

so ist $S_n = U_n - (n - U_n) = 2 \cdot U_n - n$.
Wir erhalten

$$|\{U_n = k\}| = \binom{n}{k} 2^{N-n} \Rightarrow P(U_n = k) = \binom{n}{k} \frac{2^{N-n}}{2^N} = \binom{n}{k} 2^{-n}. \qquad \square$$

Wir definieren

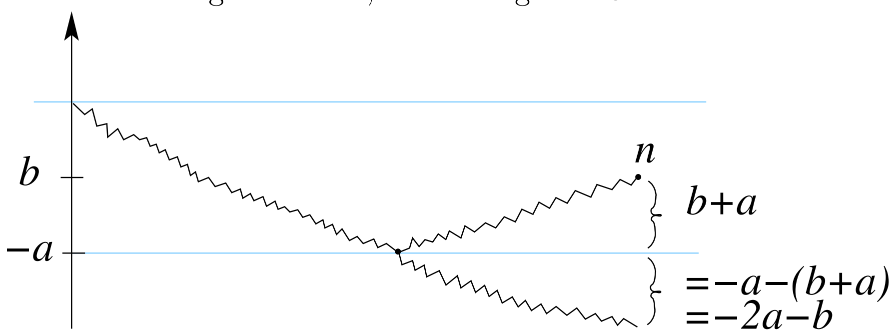
$$T_a(\omega) = \min\{n > 0 : S_n(\omega) = a\} \quad \text{mit } \min \varnothing = \infty.$$

Erstes Erreichen des Niveaus a für $a \neq 0$ bzw. erste Rückkehr nach Null sonst.

Lemma 2.27. *Für $a > 0$ und $b \geq -a$ ist*

$$P(T_{-a} \leq n, S_n = b) = P(S_n = -2a - b).$$

Der Beweis wird mittels Zählung der Pfade durchgeführt. Durch geeignetes Spiegeln erhält man eine leichte Darstellung der Pfade, was in folgende Grafik illustriert wird.



Für den allgemeinen Fall führen wir den Beweis genau aus.

Satz 2.28. *Für $a \neq 0$ gilt*

$$\begin{aligned} P(T_{-a} \leq n) &= 2P(S_n < -a) + P(S_n = -a) \\ &= P(S_n \notin (-a, a]). \end{aligned}$$

Beweis: Mit dem Additionssatz (2.5) folgt:

$$\begin{aligned} P(T_{-a} \leq n) &= \sum_{b=-\infty}^{\infty} P(T_{-a} \leq n, S_n = b) \\ &\overset{\text{Lem. 2.27}}{=} \sum_{b=-\infty}^{-a} P(S_n = b) + \sum_{b=-a+1}^{\infty} P(S_n = -2a - b) \\ &= P(S_n \leq -a) + P(S_n \leq -a - 1) \\ &= 2P(S_n < -a) + P(S_n = -a) \\ &\underset{\text{Symmetrie}}{=} (S_n < -a) + P(S_n > a) + \underbrace{P(S_n = -a)}_{P(S_n = a)} \end{aligned}$$

□

R-Code 3. *Letzter Besuch der Null bei der Irrfahrt. Wir schließen an das Listing mit der Irrfahrt an. Zunächst wird eine Illustration gezeigt, danach die Implementation der Monte-Carlo Simulation mit der Ausgabe des Histogramms. Das Histogramm zeigt die Häufigkeiten an und ist dadurch ein Schätzer für die Verteilung.*

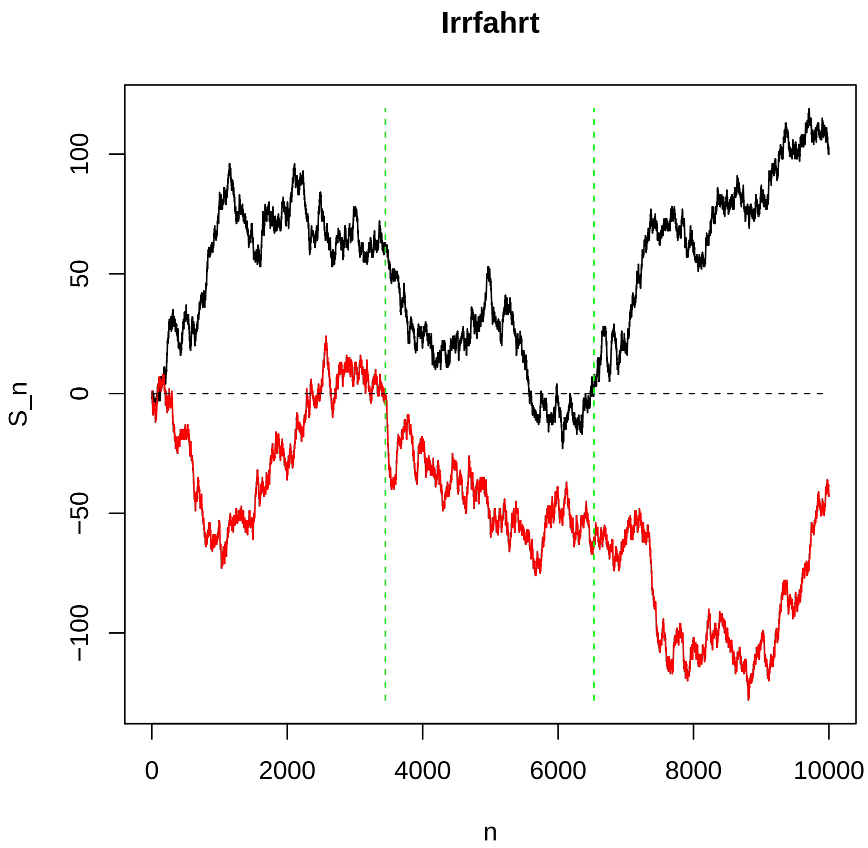
```
# Wir wiederholen obiges Beispiel und bestimmen den letzten
# Zeitpunkt an welchem S die Null erreicht hat

N=10000
omega1 = sample(c(-1,1),size=N,replace=TRUE)
omega2 = sample(c(-1,1),size=N,replace=TRUE)

# Der Pfad ist die kumulative Summe
S1 = cumsum(omega1)
S2 = cumsum(omega2)

# Fuer den Plot ist es nuetzlich min und max zu kennen
ymin = min(c(S1,S2))
ymax = max(c(S1,S2))
T1 = max(which ( S1 == 0))
T2 = max(which ( S2 == 0))

plot(S1,type="l",main="Irrfahrt",ylim=c(ymin,ymax),xlab="n",ylab="S_n")
lines(S2,col="red")
lines(c(0,N),c(0,0),lty=2)
lines (c(T1,T1),c(ymin,ymax),lty=2,col="green")
lines (c(T2,T2),c(ymin,ymax),lty=2,col="green")
```

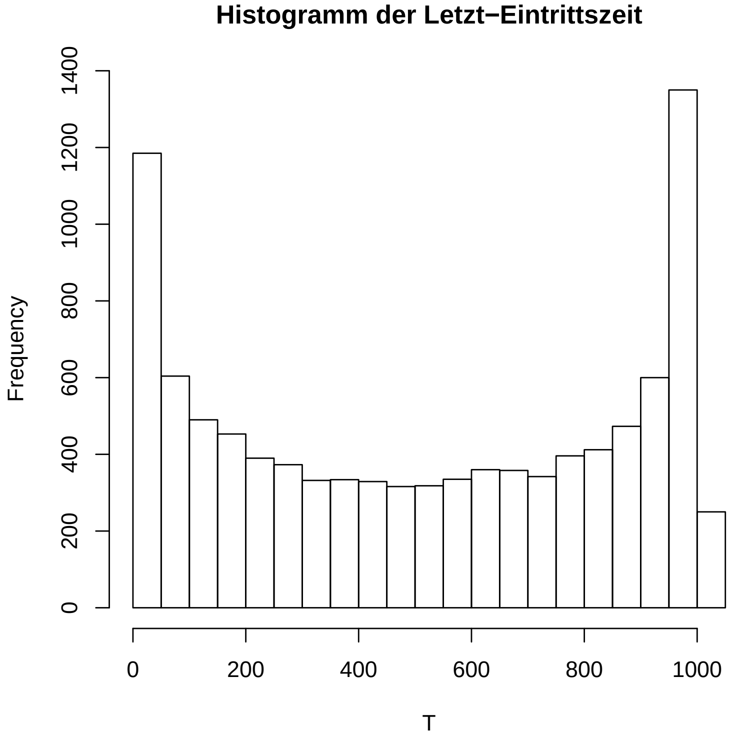



```
# Nun versuchen wir die Verteilung dieser Zeiten empirisch zu schaetzen

N=1000
N2=10000
T = rep(0,N2)

for (i in (1:N2)) {
        S = cumsum ( sample(c(-1,1),size=N,replace=TRUE) )
        # T ist die letzte Zeit wo Null erreicht wurde (oder -1)
        T[i] = max( c(which (S== 0),-1))
        # Falls die Null nicht erreicht wurde, setzen wir T=N+1
        if (T[i]==-1) {T[i]=N+1}
 }

hist (T,main="Histogramm␣der␣Letzt-Eintrittszeit")
```



Bei dem Histogramm beachte man, dass an der Stelle 1001 lediglich die Häufigkeit der Fälle, die in den ersten 1000 Schritten nicht die Null erreicht haben, aufgelistet wird. Dies gehört nicht zu der Verteilung der Letzt-Eintrittszeit.

2.1.4 Bedingte Wahrscheinlichkeiten

In diesem Kapitel führen wir ein neues Konzept auf $(\Omega, \mathscr{A}, P)$ ein: Die Wahrscheinlichkeit unter Verwendung von Zusatzinformation (etwa: die Augensumme der Würfel ist gerade oder ähnliches).

Definition 2.29 (Bedingte Wahrscheinlichkeit)**.** Seien $A, B \in \mathscr{A}$ und $P(B) > 0$. Dann heißt
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
bedingte Wahrscheinlichkeit von A gegeben B.

$P(\cdot \mid B)$ ist ein neues Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{A})$ mit
$$P(A \mid B) = 0, \;\; \forall\, A \in \mathscr{A} : A \cap B = \varnothing.$$
Die ganze Wahrscheinlichkeit konzentriert sich somit auf das Ereignis B.

B 2.30 ***Zweifacher Würfelwurf:*** $\Omega = \{1, \ldots, 6\}^2$ und

$$A_i = \text{„1. Würfel zeigt } i\text{“}$$
$$B_k = \text{„Augensumme } k\text{“}$$

Dann ist $P(A_i \mid B_7) = \frac{1/36}{1/6} = \frac{1}{6}$ für $i = 1, \ldots, 6$, die Information B_7 ist demnach nutzlos. Aber

$$P(A_i \mid B_{11}) = \begin{cases} \frac{1}{2} & \text{für } i = 5, 6 \\ 0 & \text{für } i = 1, \ldots, 4. \end{cases}$$

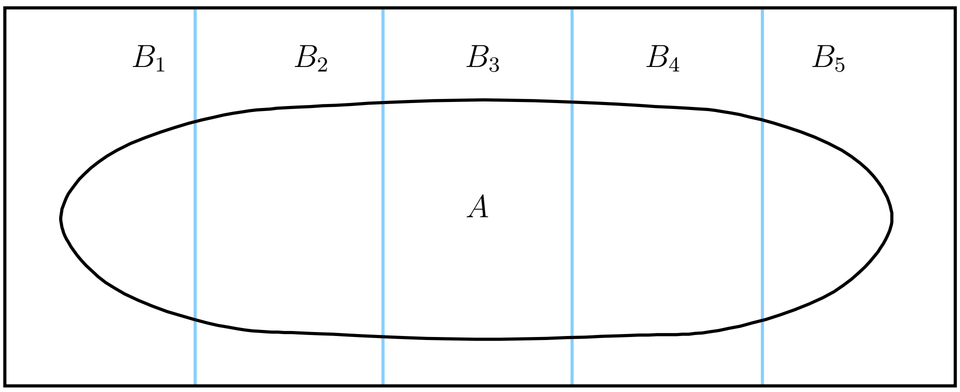
◇

Wir nennen eine abzählbare Familie von Mengen $(B_i)_{i \in I}$ eine paarweise disjunkte Zerlegung von Ω, falls $\Omega = \bigcup_i B_i$ und $B_i \cap B_j = \varnothing$ für $i \neq j$.

Satz 2.31 (Satz von der totalen Wahrscheinlichkeit)**.** *Sei $(B_i)_{i \in I}$ eine paarweise disjunkte Zerlegung von Ω mit $P(B_i) > 0$, $i \in I$. Dann gilt für alle $A \in \mathscr{A}$*

$$P(A) = \sum_{i \in I} P(A \mid B_i) \cdot P(B_i).$$

Dieser Satz ist in folgender Grafik illustriert:



Beweis. Zunächst ist $A = \bigcup_i (A \cap B_i)$, da die B_i eine paarweise disjunkte Zerlegung (eine so genannte Partition) von Ω sind. Daraus folgt, dass

$$P(A) = \sum_i P(A \cap B_i)$$

und die Behauptung folgt durch Einsetzen der Definition der bedingten Wahrscheinlichkeit. □

Satz 2.32. *Seien $A_1, \dots, A_n \in \mathscr{A}$ mit $P(A_1 \cap \dots \cap A_n) > 0$. Dann ist*

$$P\Big(\bigcap_{i=1}^{n} A_i\Big) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \dots \cap A_{n-1}).$$

Der Beweis ist eine Übungsaufgabe.

B 2.33 ***Geburtstagsproblem*:** Wie hoch ist die Wahrscheinlichkeit, dass n Personen alle an verschiedenen Tagen Geburtstag haben?

A_i = die ersten i Personen haben an verschiedenen Tagen Geburtstag

$$P(A_1 \cap \dots \cap A_n) = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - n + 1}{365}$$

Für $n = 50 \Rightarrow 0{,}03$.

Seien $A, B \in \mathscr{A}$ mit $P(A) > 0$, $P(B) > 0$. Dann folgt aus der Definition der bedingten Wahrscheinlichkeit, dass

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{P(B) \cdot P(A \mid B)}{P(A)} \tag{2.34}$$

$\Rightarrow$ die bedingte Wahrscheinlichkeit ist nicht symmetrisch!

Satz 2.35. *Ist $(B_i)_{i \in J}$ eine paarweise disjunkte Zerlegung von Ω mit $B_i \in \mathscr{A}$, $i \in J$ und ist $A \in \mathscr{A}$ mit $P(A) \neq 0$, so gilt*

$$P(B_i \mid A) = \frac{P(B_i) \cdot P(A \mid B_i)}{\sum_{j \in J} P(B_j) \cdot P(A \mid B_j)}.$$

Beweis. Es gilt $P(A) = \sum_{j \in I} P(A \cap B_j) = \sum_{j \in I} P(B_j) \cdot P(A \mid B_j)$ und die Behauptung folgt aus (2.34). Im Fall, wo $P(B_i) = 0$ ist, ist der Wert von $P(A|B_i)$ beliebig, und wir setzen ihn gleich 0. $\square$

Interpretation: Wir haben verschiedene Hypothesen (B_i) mit a-priori-Wahrscheinlichkeit $P(B_i)$. Unter der Hypothese B_i tritt A mit Wahrscheinlichkeit $P(A \mid B_i)$ auf. Ist nun A eingetreten, kann man aus diesen Größen die **a-posteriori**-Wahrscheinlichkeit $P(B_i \mid A)$ ermitteln. ($\to$ Taxi-Beispiel)

B 2.36 ***Signalübertragung*:** Wir betrachten einen Nachrichtenkanal auf dem Nullen und Einsen übertragen werden, aber mit Rauschen empfangen werden. Wir setzen:

$$A_i = \{i \text{ empfangen}\} \qquad \text{und} \qquad B_i = \{i \text{ gesendet}\}.$$

Wir kennen $P(B_i)$ und $P(A_j \mid B_i)$. Gesucht ist die **Dekodierung** $\varphi : \{0,1\} \to \{0,1\}$, so dass $[A_0 \cap A_1 = \varnothing]$,

$$P(\text{richtig dekodiert}) = P(A_0 \cap B_{\varphi(0)}) + P(A_1 \cap B_{\varphi(1)})$$

maximal ist. Wir erhalten folgende Tabelle:

Dekodierung	„richtig“	P(„richtig“)
$\varphi_1 \equiv 1$	B_1	$P(B_1) =: \alpha$
$\varphi_2 \equiv 0$	B_0	$1-\alpha$
$\varphi_3(1) = 1, \varphi_3(0) = 0$	$A_1 \cap B_1 + A_0 \cap B_0$	$\alpha \cdot \underbrace{P(A_1 \mid B_1)}_{=:p_1} + (1-\alpha)\underbrace{P(A_0 \mid B_0)}_{=:p_0}$
$\varphi_4(1) = 0, \varphi_4(0) = 1$	$A_1 \cap B_0 + A_0 \cap B_1$	$\alpha(1-p_1) + (1-\alpha)(1-p_0)$

Tabelle 2.1: Signalübertragung

Nehmen Sie an, dass $p_1, p_0 \geq 0,5$ sind und bestimmen Sie die optimale Dekodierung.

2.2 Der bedingte Erwartungswert

Sei $B \in \mathscr{A}$ mit $P(B) > 0$. Dann definieren wir den Erwartungswert bedingt auf B durch

$$E[X \mid B] = \sum_{x \in X(\Omega)} x \cdot P(x = x \mid B) = \frac{\sum_{x \in X(\Omega)} x \cdot P(X = x \cap B)}{P(B)}$$
$$= \frac{E[X 1_B]}{P(B)}.$$

Dieser Begriff ist oft nicht ausreichend. Für eine abzählbare, paarweise disjunkte Zerlegung $(B_i)_{i \in I} = \mathscr{B}$ definieren wir die Zufallsvariable $E[X \mid \mathscr{B}]$ durch

$$E[X \mid \mathscr{B}](\omega) := \sum_{i \in I, P(B_i) > 0} E[X \mid B_i] \mathbb{1}_{B_i}(\omega).$$

Diese Zufallsvariablen nennen wir bedingten Erwartungswert von X gegeben $\mathscr{B}$. Sie hat den kleinsten Abstand zu X (im quadratischen Sinne) von den Zufallsvariablen, die auf $\mathscr{B}$ konstant sind.

Satz 2.37. *Sei X eine Zufallsvariable mit $E[X^2] < \infty$ und $\mathscr{B} = (B_i)_{i \in I}$ eine paarweise disjunkte Partition von Ω. Setze $I^* = \{i \in I : P(B_i) > 0\}$. Dann ist*

$$E\Big[\Big(X - \sum_{i \in I^*} c_i \mathbb{1}_{B_i}\Big)^2\Big]$$

minimal für $c_i = \frac{E[X\mathbb{1}_{B_i}]}{P(B_i)}$, $i \in I^*$.

Beweis. Unter den Voraussetzungen des Satzes gilt

$$\begin{aligned} E\Big[X \sum_{i \in I^*} c_i \mathbb{1}_{B_i}\Big] &= \sum_{i \in I^*} x_i E\big[X \mathbb{1}_{B_i}\big] \\ &= \sum_{i \in I^*} c_i \frac{E[X\mathbb{1}_{B_i}]}{P(B_i)} \cdot P(B_i) \\ &= E\Big[\sum_{i \in I^*} c_i \mathbb{1}_{B_i} \sum_{j \in I^*} \mathbb{1}_{B_j} \frac{E[X\mathbb{1}_{B_j}]}{P(B_j)}\Big] = E\Big[E[X \mid \mathscr{B}] \sum_{i \in I^*} c_i \mathbb{1}_{B_i}\Big] \end{aligned}$$

und wir erhalten

$$E\Big[(X - E(X \mid \mathscr{B}]) \sum_{i \in I^*} c_i 1_{B_i}\Big] = 0. \tag{2.38}$$

Da $E[X \mid \mathscr{B}]$ ebenfalls von der Gestalt $\sum c_i \mathbb{1}_{B_i}$ ist, gilt auch:

$$0 = E\Big[(X - E[X \mid \mathscr{B}]) \cdot (E[X \mid \mathscr{B}] - \sum_{i \in I^*} c_i \mathbb{1}_{B_i})\Big],$$

also folgt

$$\begin{aligned} E\Big[\Big(X - \sum_{i \in I^*} c_i 1_{B_i}\Big)^2\Big] &= E\Big[(X - E[X \mid \mathscr{B}])^2 + 2(X - E(X \mid \mathscr{B}))\Big(E[X \mid \mathscr{B}] - \sum_{i \in I^*} c_i 1_{B_i}\Big) \\ &\quad + E\Big([X \mid \mathscr{B}] - \sum_{i \in I^*} c_i 1_{B_i}\Big)^2\Big]. \end{aligned}$$

Der Mittelteil verschwindet wie oben bemerkt und die Behauptung folgt. □

Formel (2.38) zeigt, dass $E[X \mid \mathscr{B}]$ die orthogonale Projektion von X auf dem Unterraum der Funktion $\sum_{i \in I^*} c_i \mathbb{1}_{B_i}$ bezüglich des Skalarproduktes

$$\langle x, y \rangle = \sum X(\omega) Y(\omega) P(\omega)$$

auf dem Raum der diskreten Zufallsvariablen ist. Man erkennt ebenfalls, dass man die Konstanten c_i außerhalb von I^* beliebig modifizieren kann: Zufallsvariablen kann man oft auf Mengen mit Wahrscheinlichkeit Null beliebig verändern, ohne eine bestimmte Eigenschaft (wie etwa in Satz 2.37) zu verlieren.

Man kann allgemeiner einen effektiveren Begriff von „Information“ einführen (σ-Algebra) als „Partition“, benötigt hierfür aber Maßtheorie.

2.3 Unabhängigkeit

Definition 2.39. Eine Familie von Ereignissen $(A_i, i \in I)$ heißt *unabhängig*, falls für alle endlichen $J \subseteq I$ gilt, dass

$$P\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} P(A_j).$$

Bemerkung 2.40. (i) Für zwei Ereignisse mit $P(A), P(B) > 0$ erhalten wir

$$\begin{aligned} A, B \text{ unabhängig } &\Leftrightarrow P(A\cap B) = P(A)\cdot P(B) \\ &\Leftrightarrow P(A \mid B) = P(A) \\ &\Leftrightarrow P(B \mid A) = P(B). \end{aligned}$$

(ii) Paarweise Unabhängigkeit ist strikt schwächer: Beim zweifachen Münzwurf etwa definieren wir

$$\begin{aligned} A &= \text{ 1. Wurf Kopf,} \\ B &= \text{ 2. Wurf Kopf,} \\ C &= \text{ 1. und 2. Wurf verschieden,} \end{aligned}$$

so sind die Ereignisse paarweise unabhängig, aber $P(A\cap B\cap C) = 0$!

(iii) Es genügt für $A_1, \dots, A_n$ nicht einfach

$$P(A_1 \cap \dots \cap A_n) = \prod_{i=1}^n P(A_i) \quad \text{zu fordern.}$$

Dies ist eine Übungsaufgabe. Hinweis: Wählen Sie etwa $A_1 = \varnothing$ und Lemma 2.43 gilt nicht mehr.

Lemma 2.41. *Die Ereignisse $(A_i)_{i\in I}$ seien unabhängig. Wir setzen B_i entweder A_i oder A_i^c. Dann sind auch $(B_i)_{i\in I}$ unabhängig.*

Die Idee ist einfach: für A, B unabhängig ist

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) \cdot (1 - P(B)).$$

Beweis. Es ist zu zeigen, dass für disjunkte endliche Mengen J, K gilt:

$$P\Big(\bigcap_{i\in J} A_i \cap \bigcap_{i\in K} A_i^c\Big) = \prod_{i\in J} P(A_i) \prod_{j\in K}(1 - P(A_i)).$$

Wir nutzen Induktion nach $k = |K|$. Für $k > 0$ folgt die Aussage aus der Unabhängigkeit. Weiterhin ist

$$\begin{aligned} P\Big(\bigcap_{i\in J} A_i \cap \bigcap i \in K A_i^c \cap A_j^c\Big) &= P\Big(\bigcap_J A_i \bigcap_K A_i^c\Big) - P\Big(\bigcap_J A_i \cap A_j \bigcap_K A_i^c\Big) \\ &= \prod_{i\in J} P(A_i) \prod_{i\in K} P(A_i^c)(1 - P(A_j)) \end{aligned}$$

und die Behauptung folgt. □

Definition 2.42. Eine Familie von diskreten Zufallsvariablen $(X_i : i \in I)$ heißt unabhängig, falls die Ereignisse

$$(\{X_i = x_i : i \in I\})$$

unabhängig sind für jede Wahl der x_i aus dem Wertebereich von X_i.

Lemma 2.43. *Sind $X_i, \dots, X_n$ diskrete und unabhängige Zufallsvariablen, so gilt*

$$E\Big[\prod_{i=1}^n g(X_i)\Big] = \prod_{i=1}^n E[g(X_i)].$$

Beweis. Wir rechnen nach. Bezeichen wir den Wertebereich von $\boldsymbol{X} = (X_1, \dots, X_n)$ mit $\boldsymbol{X}(\Omega)$, so ist

$$\begin{aligned} E\Big[\prod_{i=1}^n g(X_i)\Big] &= \sum_{\boldsymbol{x}\in\boldsymbol{X}(\Omega)} \prod_{i=1}^n g(x_i) P(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{\boldsymbol{x}\in\boldsymbol{X}(\Omega)} \prod_{i=1}^n g(x_i) P(X_i = x_i) \\ &= \prod_{i=1}^n E[g(X_i)] \end{aligned}$$

und der Satz ist bewiesen. □

2.3.1 Ein erster Grenzwertsatz

Grenzwertsätze werden einen ganz zentralen Platz in der Stochastik einnehmen. Wir beginnen mit einem einfachen Beispiel, welches eine interessante Verbindung zwischen der Binomialverteilung und der Poissonverteilung aufzeigt. Intuitiv sagt es aus, dass bei einer Binomialverteilung mit $n \to \infty$ eine Konvergenz gegen eine Poissonverteilung (die ja auf $\mathbb{N}$ lebt) entsteht, falls nur die Wahrscheinlichkeiten p_n sich in einem richtigen Sinn stabilisieren.

Als Konvergenz betrachten wir hier die *Konvergenz von Verteilungen*, d.h. wir haben eine Folge von Wahrscheinlichkeiten P^n, welche auf einem festen Grundraum Ω (in diesem Fall $\Omega = \mathbb{N}$) gegeneinander konvergieren, genauer:

$$P^n(\omega) \longrightarrow P(\omega), \qquad \text{für alle } \omega \in \Omega.$$

Man könnte sich auch andere Konvergenzen vorstellen (welche ?).

Satz 2.44. *Die Binomialverteilgung konvergiert für $n \to \infty$ und $p \to 0$ mit $np \to \lambda > 0$ gegen eine Poisson(λ)-Verteilung, d.h. für alle $k \in \mathbb{N}_0$ ist*

$$\binom{n}{k} p^k (1-p)^{n-k} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

Beweis. Wir rechnen nach:

$$\begin{aligned} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{1}{k!}(np)^k \Big(1 - \frac{np}{n}\Big)^n \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{1}{(1-p)^k} \\ &\to \frac{1}{k!}\lambda^k e^{-\lambda} \cdot 1 \cdot 1 \end{aligned}$$

und die Behauptung folgt. □

Die Poissonverteilung hat auch folgende, interessante Abschlusseigenschaft: Die Summe zweier *unabhängiger* Poisson-Verteilungen ist wieder Poisson-verteilt. Solche Eigenschaften machen die Klasse zum Einen interessant, sind zum anderen aber auch zum Rechnen ungemein hilfreich.

Lemma 2.45. *Seien X_1 und X_2 unabhängig und Poisson-verteilt zum Parameter λ_1 und λ_2. Dann ist $X_1 + X_2$ Poisson-verteilt zum Parameter $\lambda_1 + \lambda_2$.*

Beweis. Wieder rechnen wir nach:

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

und, in der Tat, die Summe ist Poisson-verteilt zum gewünschten Parameter $\lambda_1 + \lambda_2$. □

3. Stetige Modelle

In diesem Kapitel gehen wir den Schritt zu komplizierteren Modellen und verlassen das bis dato bekannte Terrain von endlichen oder abzählbaren Grundräumen.

3.1 Dichten und Wahrscheinlichkeiten im Licht von Maßen

Wir nutzen nun die axiomatische Einführung der Wahrscheinlichkeit nach A. Kolmogorov.

Sei $\Omega \neq \varnothing$ eine Menge und $\mathscr{A} \subset \mathscr{P}(\Omega)$ eine Familie von Teilmengen von Ω und $P : \mathscr{A} \to [0, 1]$ eine Abbildung.

Definition 3.1. $(\Omega, \mathscr{A}, P)$ heißt *Wahrscheinlichkeitsraum*, falls gilt:

(i) $\mathscr{A}$ ist eine *σ-Algebra*, d. h.

$\Omega \in \mathscr{A}$

$A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$

$A_1, A_2, \ldots \in \mathscr{A} \Rightarrow \bigcup_i A_i \in \mathscr{A}$

(ii) P ist ein *Wahrscheinlichkeitsmaß*, d. h.

$P(\Omega) = 1$

$A_1, A_2, \ldots \in \mathscr{A}$ und p. d. $\Rightarrow P\big(\sum_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} P(A_i)$.

Die beiden Eigenschaften eines Wahrscheinlichkeitsmaßes tragen folgende Namen: Die erste, $P(\Omega) = 1$, ist die *Normiertheit* des Maßes P. Sie impliziert, dass P ein endliches Maß ist. Man beachte außerdem, dass stets $P(\cdot) \geq 0$; es gibt allerdings auch Maße, die negative Werte annehmen können (diese heißen *signiert*).

Die zweite Eigenschaft heißt *σ-Additivität*. Ein Wahrscheinlichkeitsmaß ist also eine auf 1 normierte, nicht negative und σ-additive Funktion auf der σ-Algebra $\mathscr{A}$.

B 3.2 ***Spezialfall: diskrete Wahrscheinlichkeitsräume*:** Jeder diskrete Wahrscheinlichkeitsraum wie im vorigen Kapitel erfüllt die Bedingungen von Definition 3.1.

B 3.3 ***Borel σ-Algebra*:** Sei $\Omega = [0,1]$ und $\mathscr{A}$ die kleinste σ-Algebra, die jedes Intervall $[a,b] = [0,1]$ enthält. Dann gibt es genau ein P, so dass

$$P([a,b]) = b - a.$$

Damit folgt $P([a]) = 0$ für alle $a \in [0,1]$! Nun gilt nicht mehr

$$P(A) = \sum_{\omega \in A} p(\{\omega\}),$$

dies gilt nach Definition 3.1 (ii) nur für abzählbare Mengen A. ◇

Einfache Folgerungen:

- $A_1, A_2, \ldots \in \mathscr{A} \Rightarrow \bigcap A_i = \left(\bigcup_i A_i^c\right)^c \in \mathscr{A}$
- $A_\infty =$ „unendlich viele Ereignisse A_i treten ein“

$$= \bigcap_n \bigcup_{k \geq n} A_k.$$

- Starten wir von einem Mengensystem $\mathscr{A}_0$, so definieren wir

$$\sigma(\mathscr{A}_0) = \bigcap_{\substack{B \supseteq \mathscr{A}_0 \\ B \text{ ist } \sigma\text{-Algebra}}} \mathscr{B}$$

  die kleinste σ-Algebra, die $\mathscr{A}_0$ enthält. Das ist in der Tat (!) eine σ-Algebra.

B 3.4 ***Würfelwurf*:** Sei $\Omega = \{1, \ldots, 6\}$. Als mögliche Ereignis soll $A = \{\{1,3,5\}$ verwendet werden. Die kleinste σ-Algebra, die A enthält ist $\sigma(A) := \{\varnothing, \Omega, \{1,3,5\}, \{2,4,6\}\}$. Ganz natürlich gehören also die leere Menge und Ω, sowie das Komplement von A dazu.

Beachten Sie: Die bekannten Rechenregel für endliche Wahrscheinlichkeiten gelten hier natürlich auch. Überprüfen Sie das einmal als gute Übungsaufgabe!

Eine Funktion $P : \mathscr{A} \to [0,1]$ heißt *additiv*, falls

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \text{ für alle } n \in \mathbb{N} \text{ und } A_1, \ldots, A_n \text{ p. d.}$$

Satz 3.5. *Ist $P : \mathscr{A} \to [0,1]$ additiv, so sind äquivalent:*

(i) P *ist* σ*-additiv*

(ii) $A_1 \subset A_2 \subset \ldots, \quad A_n \in \mathscr{A} \Rightarrow P\left(\bigcup_n A_n\right) = \lim_n P(A_n)$

(iii) $A_1 \supset A_2 \supset \ldots, \quad A_n \in \mathscr{A} \Rightarrow P\left(\bigcap_n A_n\right) = \lim_n P(A_n)$

Beweis. **(i) ⇒ (ii)** Wir setzen $B_1 = A_1$, $B_n = A_n - A_{n-1}$, $n \geq 2$. Dann sind (B_n) paarweise disjunkt und $\bigcup_n B_n = \bigcup_n A_n$. Aus der σ-Additivität folgt

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} B_i\Big) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{\infty} P(B_i)$$
$$= P\Big(\bigcup_{i=1}^{\infty} B_i\Big) = P\Big(\bigcup_{i=1}^{\infty} A_i\Big).$$

(ii) ⇒ (i) Wie oben folgt $P\big(\bigcap_{i=1}^{\infty} B_i\big) = \sum_{i=1}^{\infty}(B_i)$ für $A_1 = B_1$, $A_n = A_{n-1} \bigcup B_n$.

(ii) ⇔ (iii) Durch Komplementbildung. □

Übungsaufgabe: Zeigen Sie, dass für $A_1, A_2, \dots$ unabhängig gilt, dass

$$P(\cap A_i) = \prod P(A_i).$$

Korollar 3.6. *Für beliebige $A_1, A_2, \dots \in \mathscr{A}$ gilt*

$$P\Big(\bigcup_k A_k\Big) \leq \sum_k P(A_k).$$

Beweis:
$$P\Big(\bigcup_k A_k\Big) = \lim_{n\to\infty} P\Big(\bigcup_{k=1}^{n} A_k\Big) \leq \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{\infty} P(A_k)$$
□

Lemma 3.7 (Borel-Cantelli)**.** *Wir definieren*

$$A_\infty = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k = \text{„unendlich viele der } A_i \text{ treten ein.“}$$

(i) $\sum_{k\geq 1} P(A_k) < \infty \Rightarrow P(A_\infty) = 0$

(ii) *Sind die (A_k) unabhängig, so folgt aus $\sum_{k\geq 1} P(A_k) = \infty$, dass $P(A_\infty) = 1$.*

Beweis:

(i) Die Folge $(\bigcup_{k\geq n} A_k)_n$ ist absteigend und somit ist die Voraussetzung für Bedinung (iii) von Satz 3.5 erfüllt. Wir erhalten

$$P(A_\infty) = \lim_{n\to\infty} P\Big(\bigcup_{k\geq n} A_k\Big) \overset{\text{Kor. 3.6}}{\leq} \lim_{n\to\infty} \sum_{k\geq n} P(A_n) = 0.$$

(ii) Da die Folge $\left(\bigcap_{k\leq n} A_k^c\right)_n$ wachsend ist, ist die Voraussetzung für Teil (ii) von Satz 3.5 erfüllt und es folgt

$$P(A_\infty^c) = P\Big(\bigcup_n \bigcap_{k\geq n} A_k^c\Big) = \lim_n P\Big(\bigcap_{k\geq n} A_k^c\Big).$$

Wir haben

$$P\Big(\bigcap_{k=n}^{\infty} A_k^c\Big) = \prod_{k=n}^{\infty} P(A_k^c) = \prod_{k=n}^{\infty} (1 - P(A_k)).$$

Da $1 - x \leq e^{-x}$ ist, gilt (für alle n - denn natürlich ist auch $\sum_{k\geq n} P(A_n) = \infty$)

$$P\Big(\bigcap_{k\geq n} A_k^c\Big) \leq e^{-\sum_{k\geq n} P(A_n)} = 0. \qquad \square$$

B 3.8 ***Unendliche Wiederholungen*:** Wiederholen wir den Würfelwurf unendlich oft, so können wir uns folgende Fragen stellen:

(i) Ich werfe nie eine 6:

$$A_i = \text{keine 6 in den ersten } i \text{ Würfen}$$

Wir erhalten $A_1 \supseteq A_2 \supseteq \dots$ und $P(A_i) = (\frac{5}{6})^i$. Es folgt, dass

$$P\Big(\bigcap A_i\Big) = \Big(\frac{5}{6}\Big)^\infty = 0.$$

Dies gilt übrigens für alle $q < 1$! $\quad\rightarrow$ Wohnungssuche in Freiburg :-)

(ii) Werfe ich unendlich viele 6-en?
Hier wählen wir $B_i =$ eine 6 im i-ten Wurf mit $P(B_i) = \frac{1}{6}$. Dann ist

$$\sum_{i=1}^{\infty} P(B_i) = \infty \Rightarrow \text{ JA!}$$

Definition 3.9. Eine Zufallsvariable ist eine Abbildung

$$X : \Omega \rightarrow \mathbb{R},$$

so dass $X^{-1}([a, b]) \in \mathscr{A}$ für alle $a \leq b \in \mathbb{R}$.

Man kann dies auch noch formeller formulieren[1]. Setze $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. Die Menge der reellen Zahlen $\mathbb{R}$ betrachten wir mit der Borel-σ-Algebra $\mathscr{B} = \sigma([a, b] : a, b \in \mathbb{R})$. In diesem Sinne ist eine Zufallsvariable eine Abbildung die *messbar* ist, d. h.

$$X^{-1}(B) \in \mathscr{A} \quad \forall\, B \in \mathscr{B}.$$

[1] Die Borel σ-Algebra wird genauer besprochen in Beispiel 3.3 und Gleichung (3.18).

Definition 3.10. Die durch

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}$$

definierte Funktion heißt **Verteilungsfunktion** von X. Die auf $\mathscr{B}$ definierte Funktion

$$\mu(B) := P(X \in B)$$

heißt **Verteilung** von X.

Satz 3.11. *Sei F eine Verteilungsfunktion. Dann gilt*

(i) $a \leq b \Rightarrow F(a) \leq F(b)$ *(Monotonie)*

(ii) $F(a) = \lim_{\varepsilon \to 0} F(a + \varepsilon)$ *(Rechtsstetig)*

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

Es folgt $\mu((a, b]) = F(b) - F(a)$ für $a < b$ und $\mu(\{a\}) = F(a) - F(a^-)$.

B 3.12 ***Diskrete Zufallsvariable*:** Nimmt X die Werte $\{0, 1, 2, \dots\}$ an, so ist

$$F(x) = \sum_{k=0}^{x} P(X = k), \quad x = 0, 1, 2, \dots .$$

Somit ist die Kenntniss der Verteilungsfunktion gleichwertig mit der Kenntniss aller $P(X = k)$, $k = 0, 1, 2, \dots$.

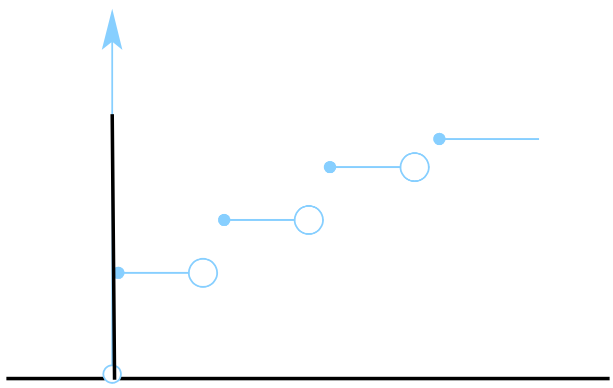


Bild einer Verteilungsfunktion.

Die Verteilungsfunktion F ist monoton, also können wir die *verallgemeinerte Inverse*

$$F^{-1}(t) = \inf\{x : F(x) \geq t\}$$

definieren.

Lemma 3.13. *Es gilt: F^{-1} ist monoton wachsend und*

(i) $F^{-1}(F(x)) \leq x$

(ii) $F(F^{-1}(t)) \geq t$

Beweis: $\{x : F(x) \geq t\} = [F^{-1}(x), \infty) \Rightarrow F^{-1}$ monoton wachsend

(i) $F^{-1}(F(x)) = \inf\{y : F(y) \geq F(x)\}$, da für $x = y$ gilt $F(y) \geq (F(x)$, folgt $\leq x$

(ii) $F(F^{-1}(t)) = P(x \leq F^{-1}(t))$
$F^{-1}(t) = \inf\{P(x \leq x) \geq t\} := y \overset{F \text{ rechtsstet.}}{\Rightarrow} P(X \leq y) \geq t$
$\geq t.$ □

Wir erhalten

$$P(X < F^{-1}(t)) \leq t \leq P(X \leq F^{-1}(t))$$

Man nennt $F^{-1}(t)$ auch das t-Quantil von X. (Median $\to$ 50 %-Quantil)

Definition 3.14. Eine Zufallsvariable X heißt **absolut stetig**, falls eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ existiert, so dass

(i) $f(x) \geq 0$ für alle $x \in \mathbb{R}$

(ii) $\displaystyle\int_{\mathbb{R}} f(x)\, dx = 1$

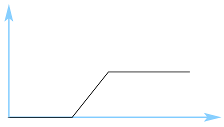
(iii) $F(b) = P(X \leq b) = \displaystyle\int_{-\infty}^{b} f(x)\, dx \quad$ für alle $b \in \mathbb{R}$

Die Funktion f heißt **Dichte** von X.

Ist f stetig an x, so gilt $f(x) = F'(x)$.

B 3.15 ***Gleichverteilung*:** Für die Gleichverteilung auf $[a,b]$ schreiben wir $U(a,b)$. Wir setzen $f(x) = \mathbb{1}_{[a,b]}\frac{1}{b-a}$ und erhalten

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & x \in (a,b) \\ 1, & x \geq b. \end{cases}$$



B 3.16 ***Exponentialverteilung*:** Für die Exponentialverteilung schreiben wir $\text{Exp}(\lambda)$. Sei $\lambda > 0$ und $f(x) = \mathbb{1}_{\{x \geq 0\}}\lambda e^{-\lambda x}$. Die Verteilungsfunktion der Exponentialverteilung ist

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

B 3.17 ***Normalverteilung*:** Für die Normalverteilung schreiben wir $\mathcal{N}(\mu, \sigma^2)$. Sei $\mu \in \mathbb{R}$, $\sigma > 0$. Die Dichte der Normalverteilung ist

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Hieraus erhalten wir die Verteilungsfunktion:

$$\begin{aligned} F_{\mu,\sigma^2} &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-y)^2}{2\sigma^2}}\, dx \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \Phi\Big(\frac{x-\mu}{\sigma}\Big). \end{aligned}$$

◇

Aus Lemma 3.13 erhalten wir für $0 < t < 1$

$$\left.\begin{aligned} t \leq F(x) &\quad \Rightarrow F^{-1}(t) \leq F^{-1}(F(x)) \leq x \\ F^{-1}(t) \leq x &\quad \Rightarrow \underbrace{F(F^{-1}(t))}_{\geq t} \leq F(x) \text{ also } t \leq F(x) \end{aligned}\right\} \Rightarrow F^{-1}(t) \leq x \Leftrightarrow t \leq F(x).$$

Ist $U \sim U[0,1]$ und

$$X = F^{-1}(U), \quad \text{so gilt}$$

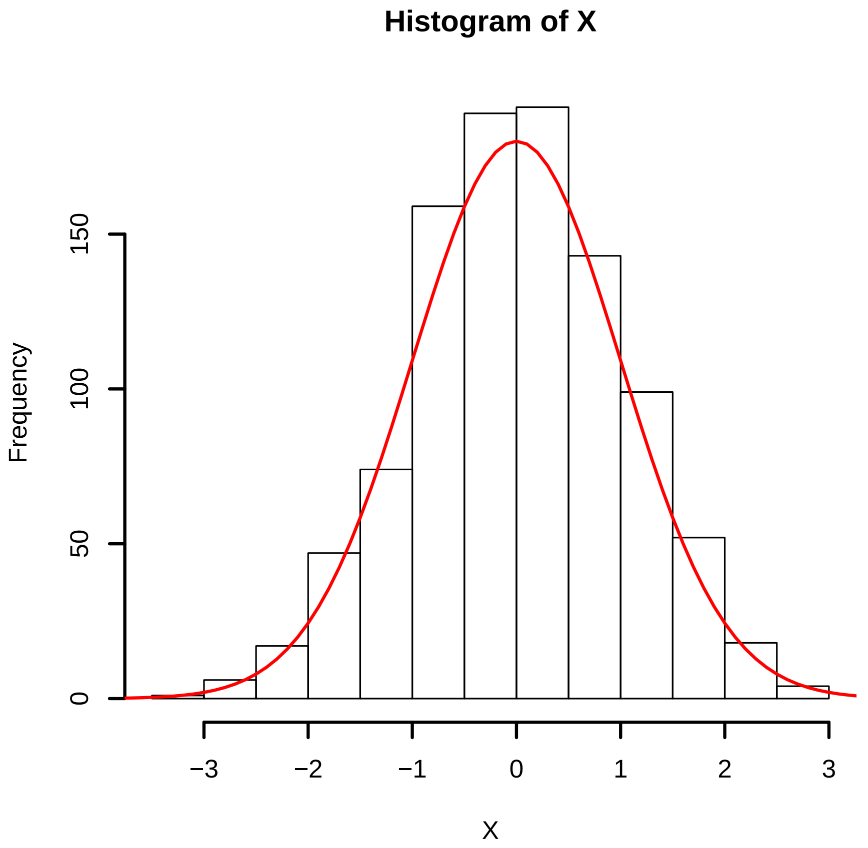
$$P(X \leq b) = P(\{U : F^{-1}(U) \leq b\}) = P(\{U : U \leq F(b)\}) = F(b), \quad \text{also } X \sim F!$$

R-Code 1.

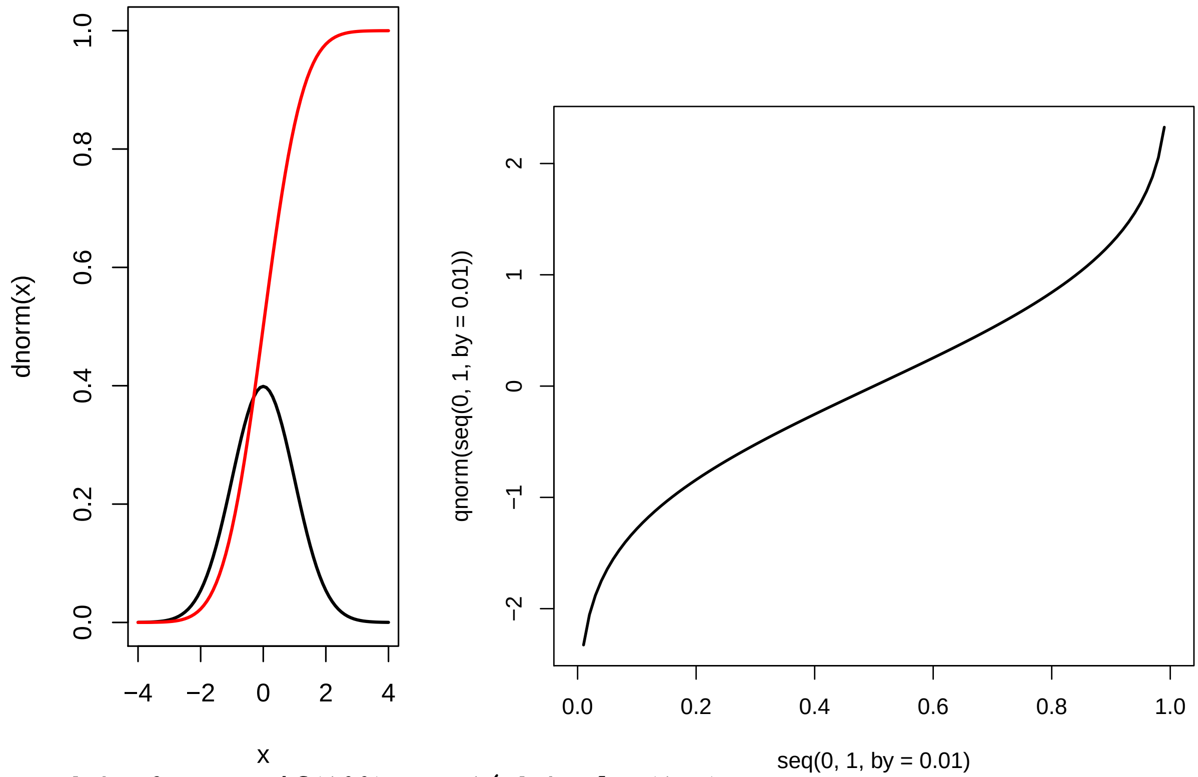
```
# Example 1 - Normal Verteilung

X=rnorm(1000)
hist(X,lwd=2)
lines(seq(-4,4,by=0.1),sqrt(2*pi)*180*dnorm(seq(-4,4,by=0.1)),col=2,lwd
    =2)
# Dichte und Stammfunktion
```

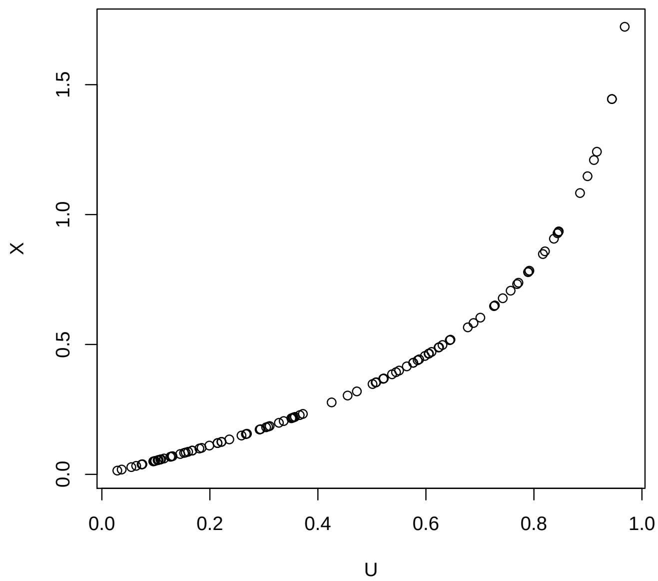
```
x=seq(-4,4,by=0.1)
plot(x,dnorm(x),ylim=c(0,1),type="l",lwd=2)
# Dann die Stammfunktion
lines(x,pnorm(x),col=2,lwd=2)
```



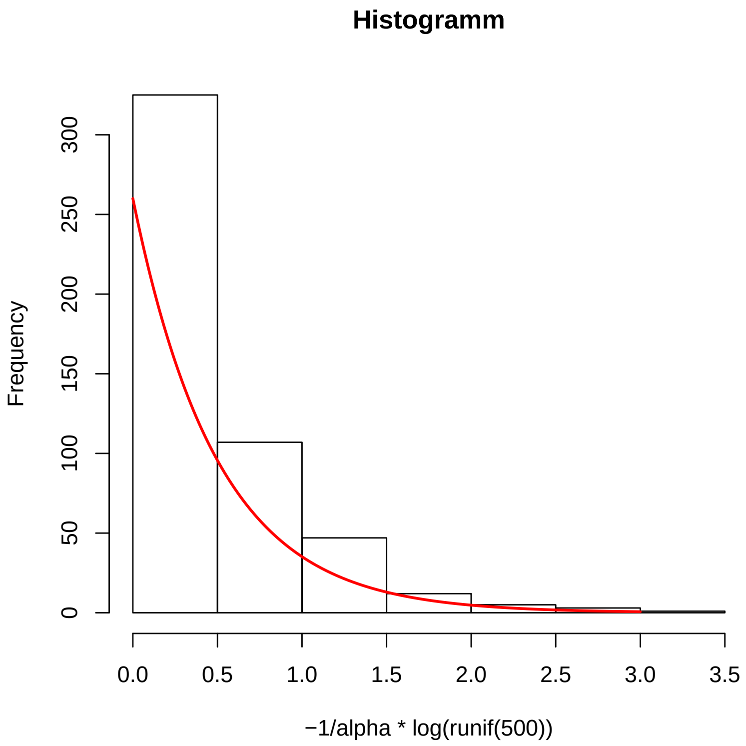

```
# Nun die Inverse der Verteilungsfunktion
par(mfrow=c(1,2)); x=seq(-4,4,by=0.1)
plot(x,dnorm(x),ylim=c(0,1),type="l",lwd=2)
lines(x,pnorm(x),col=2,lwd=2)
plot(seq(0,1,by=0.01),qnorm(seq(0,1,by=0.01)),lwd=2,type="l")
```




```
alpha=2; U=runif(100); X=-1/alpha*log(1-U)
plot(U,X)
```




```
# Histogramm - es funktioniert !
hist(-1/alpha*log(runif(500)),main="Histogramm")
lines(seq(0,3,by=0.01),dexp(seq(0,3,by=0.01),rate=alpha)*130,col=2,lwd
    =2)
```



3.1.1 Transformation von Zufallsvariablen

Wir definieren die Borel-σ-Algebra durch

$$\mathscr{B}(\mathbb{R}) = \mathscr{B} := \sigma([a,b], a < b \in \mathbb{R}) = \sigma((a,b), a < b \in \mathbb{R}) = \sigma((-\infty, a], a \in \mathbb{R}). \qquad (3.18)$$

Die Borel-σ-Algebra ist die kleinste σ-Algebra, die alle abgeschlossenen Intervalle enthält. Man kann leicht zeigen, dass sie auch alle offenen Intervalle und auch alle Halbstrahlen enthält. Eine Funktion $g : \mathbb{R} \to \mathbb{R}$ heißt **messbar**, falls sie $\mathscr{B}(\mathbb{R}) - \mathscr{B}(\mathbb{R})$-messbar ist, also gilt, dass

$$g^{-1}(B) \in \mathscr{B}(\mathbb{R}) \quad \forall B \in \mathscr{B}(\mathbb{R}).$$

Ist X eine Zufallsvariable und g messbar, so ist $y = g(x)$, definiert durch

$$Y(\omega) = g(X(\omega))$$

wieder eine Zufallsvariable (Übungsaufgabe). Sie hat die Verteilungsfunktion

$$F_y(b) = P(g(X) \leq b) = P(X \in g^{-1}((-\infty, b])).$$

B 3.19 ***Transformation:*** **X^2:** Sei $g(x) = x^2$ und $Y = g(X)$. Dann ist

$$P(X^2 \leq b) = P(-\sqrt{b} \leq X \leq \sqrt{b}) = F_X(\sqrt{b}) - F_X(-\sqrt{b}),$$

falls F_X stetig ist. (Was, wenn nicht?) Ist F_X sogar absolut stetig mit Dichte f_X, so folgt

$$f_Y(b) = \frac{1}{2\sqrt{b}}[f_X(\sqrt{b}) + f_X(-\sqrt{b})].$$

B 3.20 ***Transformation:*** $aX + b$**:** Sei $g(x) = ax + b$ und wie oben $Y = g(X)$. Dann ist

$$F_Y(x) = F_X(\frac{x-b}{a})$$

und, falls F_X absolut stetig,

$$f_Y(x) = \frac{1}{a} f_X\Big(\frac{x-b}{a}\Big).$$

3.1.2 Erwartungswert

Definition 3.21. Für $X \geq 0$ ist der Erwartungswert von X definiert durch

$$E[X] = \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}} x\mu(dx) \in [0, \infty).$$

Die Integrale sind im Sinne der Maßtheorie zu verstehen. Man beginnt mit diskreten Zufallsvariablen

$$X = \sum_{i=1}^n c_i \mathbb{1}_{A_i} \quad \text{mit } A_i.$$

(Dann ist $E[X] = \sum_{i=1}^n c_i P(A_i)$ wie bekannt.)

Ein nicht negatives X kann man immer als Limes einer aufsteigenden Folge (X_n) solcher Zufallsvariablen schreiben und man setzt $E[X] = \lim_n E[X_n]$.

Für eine beliebige Zufallsvariable X definiert man

$$X^+ = \max(X_0) \quad \text{und} \quad X^- = \max(-X, 0)$$

und setzt $E[X] = E[X^+] - E[X^-]$, falls nicht beide Seiten $= +\infty$ sind. (Dann kann man auch $E[X] = \infty$ setzen.)

In allen praktischen Fällen können wir $E[X]$ direkter ausrechnen: Hat X die Dichte f, so ist

$$E[X] = \int x f(x) dx$$

(da $\mu(dx) = f(x)\,dx$ bzw. $dF(x) = f(x)\,dx$).

Bemerkung 3.22. Der Erwartungswert erbt folgende Eigenschaften des Integrals:

(1)	Linearität	$E[aX + bY] = aE[X] + bE[Y]$
(2)	Monotonie	$X \leq Y \;\Rightarrow\; E[X] \leq E[Y]$
(3)	monotone Konvergenz	$X_1 \leq X_2 \leq \cdots \;\Rightarrow\; E[\lim_n X_n] = \lim_n E[X_n]$
(4)	Konvergenz nach Lebesgue	konvergiert $X_n(\omega) \to X(\omega)$ für $\omega \in A$ mit $P(A) = 1$ und ist $\lvert X_n(\omega)\rvert \leq X(\omega)$ für alle n mit $E[\lvert X\rvert] < \infty$, so gilt $$E[\lim_n X_n] = \lim_n E[X_n]$$
(5)	Transformation	$E[g(X)] = \int g(x)f(x)\,dx$ falls X absolut stetig.

Definition 3.23. Wir definieren die *Varianz* von X durch

$$\mathrm{Var}(X) = E[(X - E(X))^2],$$

und die *Standardabweichung* durch $\sigma(x) = \sqrt{\mathrm{Var}(X)}$ mit der Konvention $\sqrt{\infty} = \infty$.

A 3.1 ***Varianz*:** Zeigen Sie, dass gilt: $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(x)$, $\mathrm{Var}(X) = E[X^2] - (E[X])^2$.

$\boxed{\text{B 3.24}}$ ***Erwartungswerte und Varianzen für verschiedene Verteilungen*:** Wir geben diese Werte für die Binomialverteilung, die hypergeometrische Verteilung, die Poisson-Verteilung, die geometrische Verteilung, die Gleichverteilung, die Exponentialverteilung und die Normalverteilung an. Die Berechnung kann als Übungsaufgabe durchgeführt werden.

Verteilung	$E[X]$	$\mathrm{Var}(X)$
$\mathrm{Bin}(n,p)$	np	$np(1-p)$
$\mathrm{Hyp}(n,N,K)$	$n\frac{K}{N}$	$n\frac{K}{N}\left(1-\frac{K}{N}\right)\left(\frac{N-n}{N-1}\right)$
$\mathrm{Poisson}(\lambda)$	λ	λ
$\mathrm{Geometr.}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathrm{Uniform}[a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathrm{Exp}(\alpha)$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$
$\mathcal{N}(\mu,\sigma^2)$	μ	σ^2

B 3.25 ***Poisson-Verteilung*:** Als Beispiel berechnen wir Erwartungswert einer Poissonverteilung. Sei dazu $X \sim \text{Poisson}(\lambda)$. Dann gilt

$$\begin{aligned} E[X] &= \sum_{k\geq 0} kP(x=k) \\ &= \sum_{k\geq 0} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda \sum_{k\geq 1} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = \lambda. \end{aligned}$$

B 3.26 ***Exponentialverteilung*:** Und für eine Exponentialverteilung: Sei dazu $X \sim \text{Exp}(\lambda)$. Dann gilt

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)\,dx \\ &= \int_0^{\infty} x\lambda e^{-\lambda x} dx = \underbrace{x(-e^{-\lambda x})\,\Big|_0^{\infty}}_{=0} + \int_0^{\infty} e^{-\lambda x} dx \\ &= -\frac{1}{\lambda}e^{-\lambda x}\,\Big|_0^{\infty} = \frac{1}{\lambda} \end{aligned}$$

B 3.27 ***Gleichverteilung*:** Schließlich berechnen wir noch die Varianz einer Standard-Gleichverteilung. Sei dazu $X \sim U[0,1]$. Dann gilt

$$\begin{aligned} E[X^2] &= \int_0^1 x^2\,dx = \frac{1}{3}x^3\,\Big|_0^1 = \frac{1}{3}, \\ E[X] &= \int_0^1 x\,dx = \frac{1}{2} \end{aligned}$$

und so folgt

$$\text{Var}(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{4}{12} - \frac{3}{12} = \frac{1}{12}.$$

3.1.3 Ungleichungen

Für eine nichtlineare Funktion ist im Allgemeinen

$$E[g(X)] \neq g(E[x]).$$

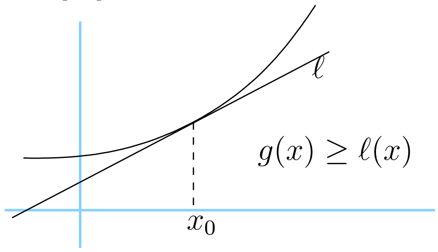
Für konvexes (konkaves) g hat man wenigstens eine Ungleichung.

Satz 3.28 (Jensensche Ungleichung)**.** *Sei X eine Zufallsvariable mit $E[|X|] < \infty$ und g konvex. Dann ist*

$$E[g(X)] \geq g(E[X]).$$

Beweis. Sei $x \to \ell(x)$ die Stützgerade von g für $x_0 = E[X]$. Dann ist

$$\begin{aligned} g(E[X]) &= \ell(E[X] \\ &= E[\ell(x)] \quad \text{(Linearität)} \\ &\leq E[g(x)] \quad \text{(Monotonie)} \end{aligned}$$



□

Satz 3.29 (Verallgemeinerte Tschebyscheff-Ungleichung)**.** *Sei die Funktion $g \geq 0$ und monoton wachsend. Dann gilt für jedes c mit $g(c) > 0$, dass*

$$P(X \geq c) \leq \frac{E[g(x)]}{g(c)}$$

Beweis. Zunächst gilt

$$\mathbb{1}_{\{X \geq c\}} \leq \frac{g(x)}{g(c)} \quad \forall x \in \mathbb{R}.$$

Wegen Monotonie des Erwartungswertes erhalten wir

$$P(X \geq c) = E[\mathbb{1}_{\{X \geq c\}}] \leq E\left[\frac{g(x)}{g(c)}\right]$$

und der Beweis ist bereits fertig. □

Als Spezialfall erhalten wir zwei wichtige Ungleichungen.

B 3.30 ***Markov-Ungleichung*:** Wenden wir $g(x) = x^+ = \max(x, 0)$ auf $Y = |X|$ an, so erhalten wir für alle $c > 0$ die Markov-Ungleichung:

$$P(|X| > c) \leq \frac{E[|X|]}{c}.$$

Insbesondere folgt aus $E[|X|] = 0$, dass

$$P(X = 0) = P\Big(\bigcap_{n=1}^{\infty} \Big\{|X| \leq \frac{1}{n}\Big\}\Big) \underset{\substack{\uparrow \\ \sigma\text{-Stetigkeit}}}{=} \lim_n P\Big(|X| \leq \frac{1}{n}\Big) = 1.$$

B 3.31 ***Tschebyscheff-Ungleichung*:** Betrachten wir $Y = |X - E(X)|$ mit $g(x) = (X^+)^2$, so folgt die berühmte Tschebyscheff-Ungleichung,

$$P(|X - E[X]| > c) \leq \frac{\text{Var}(X)}{c^2}. \tag{3.32}$$

Insbesondere folgt aus $\text{Var}(X) = 0$, dass X f.s. konstant ist.

4. Konvergenz von Zufallsvariablen

Ein wichtiges Hilfsmittel in der Statistik ist bei einer Stichprobe die Anzahl der gezogenen Proben gegen Unendlich gehen zu lassen. Es wird sich zeigen, dass man mit diesem Trick den Zufall ganz exakt in den Griff bekommt und präzise Aussagen treffen kann. Das Hilfsmittel hierzu sind die folgenden Grenzwertsätze.

4.1 Die Gesetze der großen Zahl

Wir betrachten einen fixierten Wahrscheinlichkeitraum $(\Omega, \mathscr{A}, P)$. Zunächst einmal sind Zufallsvariablen Funktionen. Es liegt nahe den bekannten Begriff der *punktweisen Konvergenz* von Funktionen anzuwenden, d.h. die Folge (X_n) konvergiert gegen X falls

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

für alle $\omega \in \Omega$. Das folgende Beispiel zeigt, dass dieser Begriff zu stark ist.

B 4.1 ***Konvergenz beim Münzwurf*:** Haben wir unabhängige und identische verteilte (i.i.d.) Münzwürfe (X_n), so erwarten wir, dass $\frac{1}{n}\sum X_n \to \frac{1}{2}$. Werfen wir allerdings nur Kopf $(X_n(\omega) = 1, \quad \omega \in \Omega)$, so ist $\frac{1}{n}\sum X_n(\omega) = 1$. Genau so verhält es sich wenn wir höchstens endlich oft Zahl werfen. Allerdings haben diese Ereignisse eine Wahrscheinlichkeit von Null, was den folgenden Konvergenzbegriff "fast sicher" motiviert.

Wie in der Analysis gibt es eine Reihe von wichtigen, ganz unterschiedlichen Konvergenzbegriffen.

Definition 4.2. Für Zufallsvariablen $Z, Z_1, Z_2, \dots$ definieren wir

(i) stochastische Konvergenz $Z_n \xrightarrow{P} Z$, falls $P(|Z_n - Z| > \varepsilon) \to 0 \quad \forall \varepsilon > 0$

(ii) fast-sichere Konvergenz $Z_n \to Z$ f.s., falls

$$P(\{\omega \in \Omega : \lim Z_n(\omega) = z(\omega)\}) = 1.$$

Satz 4.3. *Es gelten folgende Aussagen:*

(i) Fast-sichere impliziert stochastische Konvergenz

(ii) $\sum_n P(|Z_n - Z| > \varepsilon) < \infty \;\; \forall \varepsilon > 0 \quad \Rightarrow \quad Z_n \to Z$ *fast sicher.*

Wir erhalten

Korollar 4.4. *Gilt* $Z_n \xrightarrow{P} Z$, *so existiert eine Teilfolge* $Z_{n(j)}$, *welche fast sicher gegen* Z *konvergiert.*

Beweis. Wähle $n(j)$ so dass $P\big(|Z_{n(j)} - Z| \geq \frac{1}{j}\big) \leq \frac{1}{j^2}$. Behauptung folgt aus (ii) des vorigen Satzes. □

Beweis von Satz 4.3. Teil (i): Zunächst einmal ist

$$1 = P(\lim_n Z_n = Z) = P\Big(\bigcap_{\varepsilon>0} \bigcup_{n\geq 1} \bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big). \tag{4.5}$$

Die Mengen $\bigcap_{k\geq n}\{|Z_k - Z| \leq \varepsilon\}$ sind monoton wachsend in n und ε, was wir nun geschickt ausnutzen. Mit der σ-Stetigkeit von P (Satz 3.5) folgt, dass (4.5) äquivalent dazu ist, dass

$$1 = P\Big(\bigcup_{n\geq 1} \bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big) \tag{4.6}$$

für alle[1] $\varepsilon > 0$. Eine zweite Anwendung der σ-Stetigkeit liefert, dass (4.5) äquivalent dazu ist, dass

$$\lim_n P\Big(\bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big) = 1 \;\; \forall \varepsilon > 0.$$

Hieraus folgt

$$0 = \lim_n P\Big(\bigcup_{k\geq n} \{|Z_k - Z| > \varepsilon\}\Big) \geq \lim_n P(|Z_n - Z| > \varepsilon)$$

und somit die gesuchte stochastische Konvergenz.

Teil (ii): Für ein festes $\varepsilon > 0$ folgt mit Borel-Cantelli (Lemma 3.7), dass

$$\sum_{n=1}^{\infty} P(|Z_n - Z| > \varepsilon) < 0 \quad \Rightarrow \quad P\Big(\bigcap_n \bigcup_{k\geq n} \{|Z_k - Z| > \varepsilon\}\Big) = 0$$

Das Komplement ist

$$P\Big(\bigcup_n \bigcap_{k\geq n} \{|Z_k - Z| \leq \varepsilon\}\Big) = 1. \tag{4.7}$$

Da ε beliebig gewählt war, liefert die Äquivalenz (4.6) nun die fast sichere Konvergenz. □

[1] In der Tat: Ist $1 = P(\cap_{\varepsilon>0} A_\varepsilon)$ und gilt $A_\varepsilon \supset A_{\varepsilon'}$ für $\varepsilon > \varepsilon'$, so folgt $1 = P(\cap_{n\geq 0} A_{n^{-1}}) = \lim_n P(A_{n^{-1}})$. Da $\lim_n P(A_{n^{-1}}) \leq P(A_{n'^{-1}})$ für alle $n' \geq 1$, folgt sogar, dass $1 = P(A_{n^{-1}})$ für alle $n \geq 1$ und schließlich $1 = P(A_\varepsilon)$ für alle $\varepsilon > 0$.

B 4.8 ***Stochastische impliziert nicht fast sichere Konvergenz*:** Betrachten wir Borel-Mengen (A_n) mit der Gleichverteilung auf $[0,1]$ und die Folge von Zufallsvariablen $(\mathbb{1}_{A_n})$, so konvergiert diese Folge genau dann stochastisch gegen 0, falls $P(A_n) \to 0$. Dies folgt unmittelbar, da für $0 < \varepsilon < 1$, $P(\mathbb{1}_{A_n} > \varepsilon) = P(A_n)$ ist.

Für das Gegenbeispiel betrachten wir die Intervalle

$$X_{n,j} = \Big[\frac{(j-1)}{n}, \frac{j}{n}\Big], \qquad 1 \leq j \leq n$$

und konstruieren eine Folge (Y_n) von Zufallsvariablen, indem wir, beginnend mit $n = 1$ zunächst n festhalten und alle $j = 1, \dots, n$ durchlaufen, dann n um eins erhöhen, und so fort. Diese Folge konvergiert nach obiger Bemerkung stochastisch gegen 0. Allerdings ist $\limsup Y_n = 1$ und $\liminf Y_n = 0$, so dass keine fast sichere Konvergenz gelten kann. Wählt man aber die Teilfolge $(X_{n,1})$, so erhält man sogar fast sichere Konvergenz.

Lemma 4.9. *Seien $X_1, \dots, X_n$ unabhängig mit $E[X_i^2] < \infty$. Dann ist*

$$\operatorname{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \operatorname{Var}(X_i). \tag{4.10}$$

Beweis. Mit Induktion reicht es zu zeigen, dass

$$\begin{aligned}\operatorname{Var}(X+Y) &= E[(X - E[X] + Y - E[Y])^2)] \\ &= E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2E[(X-E[X])(Y-E[Y])] \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y),\end{aligned}$$

da $E[(X-E[X])(Y-E[Y])] = 0$ aufgrund der Unabhängigkeit. □

Satz 4.11 (Schwaches Gesetz der großen Zahlen)**.** *Seien $X_1, X_2, \dots$ unabhängig und identisch verteilt mit $\operatorname{Var}(X_i) < \infty$. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} E[X_1].$$

Beweis. Mit der Tschebyscheff-Ungleichung gilt

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n X_i - E[X_1]\Big| > \varepsilon\Big) \leq \frac{\operatorname{Var}(\frac{1}{n}\sum X_i)}{\varepsilon^2} = \frac{\sum_{i=1}^n \operatorname{Var}(X_i)}{n^2 \cdot \varepsilon^2} = \frac{\operatorname{Var}(X_1)}{n \cdot \varepsilon^2} \to 0 \tag{4.12}$$

und die Behauptung folgt. □

Bemerkung 4.13. Für die zentrale Formel (4.10) braucht man nicht unbedingt, dass die (X_i) unabhängig sind, es reicht dass sie unkorreliert sind, d.h. es gilt $E[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = 0$ für alle $i \neq j$. Somit gilt das schwache Gesetz der großen Zahlen bereits für unkorrelierte und identisch verteilte (X_i) mit $\sum_{i=1}^n \mathrm{Var}(X_i) = o(n^2)$.

B 4.14 ***Der Weierstraßsche Approximationssatz*:** Als Anwendung des Gesetzes der großen Zahl beweisen wir diesen wichtigen Satz - Polynome liegen dicht in der Menge der stetigen Funktionen auf Kompakta: Wir betrachten das Intervall $[0,1]$ und die stetigen Funktionen $C[0,1]$, versehen mit der Supremumsnorm $\| f \|_\infty = \sup_{x\in[0,1]} |f(x)|$. Die *Bernstein-Polynome* sind definiert durch

$$B_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}, \qquad x \in [0,1],\ k = 0,\ \dots, n. \tag{4.15}$$

Die Approximation einer stetigen Funktion f durch die Bernstein-Polynome ist gegeben durch

$$B_n^f(x) = \sum_{k=0}^n f\Big(\frac{k}{n}\Big) B_{n,k}(x).$$

Man kann $B_n^f(x)$ als $E[f(S_n/n)]$ auffassen, wobei S_n die Summe von identisch verteilten Bernoulli-Variabeln mit Erfolgsparameter $x \in (0,1)$ ist (siehe (2.22)). Das Gesetz der großen Zahl liefert nun $S_n/n \to x$. Genauer gilt mit der Jensenschen Ungleichung, dass

$$\begin{aligned} \| B_n^f(x) - f(x) \|_\infty &= \big\| \mathbb{E}[f(S_n/n)] - f(x) \big\|_\infty \\ &\leq \mathbb{E}\Big[\| f(S_n/n) - f(x) \|_\infty \Big]. \end{aligned} \tag{4.16}$$

Im nächsten Schnitt nutzen wir die (sogar gleichmäßige !) Stetigkeit und teilen das Supremum auf in $|u-v| \leq \delta$ und $|u-v| > \delta$: Es gilt

$$\begin{aligned} (4.16) &= \mathbb{E}\Big[\mathbb{1}_{\{|S_n/n - x| \leq \delta\}} \| f(S_n/n) - f(x) \|_\infty \Big] + \mathbb{E}\Big[\mathbb{1}_{\{|S_n/n - x| > \delta\}} \| f(S_n/n) - f(x) \|_\infty \Big] \\ &\leq \sup_{|u-v| \leq \delta} |f(u) - f(v)| \cdot P(|S_n/n - x| \leq \delta) + 2 \| f \|_\infty \cdot P(|S_n/n - x| > \delta). \end{aligned}$$

Wegen der gleichmäßigen Stetigkeit von f finden wir für jedes $\varepsilon > 0$ ein $\delta > 0$, so dass $\sup_{|u-v| \leq \delta} |f(u) - f(v)| \leq \varepsilon$ ist (und damit auch der ganze erste Term). Das schwache Gesetz der großen Zahl liefert für jedes $\varepsilon' > 0$ ein n_0, so dass für alle $n \geq n_0$, $P(|S_n/n - x| > \delta) \leq \varepsilon'$. Da ε und ε' beliebig waren, folgt $\| B_n^f(x) - f(x) \|_\infty \to 0$ für $n \to \infty$.

Erhalten wir vielleicht auch fast-sichere Konvergenz? Unter der Annahme der zweiten Momente folgt direkt folgende Aussage.

Satz 4.17 (Starkes Gesetz der großen Zahlen)**.** *Seien $X_1, X_2, \dots$ unabhängig und identisch verteilt mit* $\mathrm{Var}(X_i) < \infty$. *Dann gilt*

$$\frac{1}{n} \sum_{i=1}^n X_i \longrightarrow E[X_1] \quad f.s.$$

Beweis. Wir können ohne Einschränkung der Allgemeinheit annehmen, dass $X \geq 0$ (zerlege $X = X^+ - X^-$ und beweise die Aussagen getrennt). Hierdurch gewinnen wir eine wichtige Monotonie in $S_n = \sum_{i=1}^n X_i$. Zunächst folgt mit (4.12) und Satz 4.3 (ii), dass die Teilfolge $S_{n^2}/(n^2)$ fast sicher gegen $E[X_1]$ konvergiert.

Für die verbleibenden k mit $n^2 < k < (n+1)^2$ erhalten wir folgende obere und untere Schranken: Da $S_k \geq S_{n^2}$ ist, folgt

$$\frac{S_k}{k} \geq \frac{S_{n^2}}{(n+1)^2} = \frac{n^2}{(n+1)^2}\frac{S_{n^2}}{n^2}$$

und

$$\frac{S_k}{k} \leq \frac{S_{(n+1)^2}}{n^2} = \frac{S_{(n+1)^2}}{(n+1)^2}\frac{(n+1)^2}{n^2}.$$

Sowohl obere als auch untere Schranke konvergieren fast sicher (nach unserer obigen Beobachtung) gegen $E[X_1]$ und die Behauptung folgt. □

Notwending und hinreichend für das starke Gesetz der große Zahl in dieser Form ist lediglich die Annahme $E[|X_1|] < \infty$, was von Kolmogorov bewiesen wurde (siehe *Strong law of large numbers* (2011), Kolmogorov (1930)). Dieser letzte Schritt findet sich sehr schön bewiesen in Etemadi (1981):

Satz 4.18 (Starkes Gesetz der großen Zahlen). *Seien $X_1, X_2, \dots$ unabhängig und identisch verteilt mit $E[|X_1|] < \infty$. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^n X_i \longrightarrow E[X_1] \quad f.s.$$

4.2 Der zentrale Grenzwertsatz

Bisher haben wir uns vornehmlich für die Konvergenz von Zufallsvariablen interessiert, etwa Konvergenz des arithmetischen Mittels gegen den Erwartungwert. In diesem Kapitel interessiert uns nun die Variation um den Limes. Hierzu skaliert man mit einer anderen Folge und erhält eine neue Art von Konvergenz, Konvergenz in Verteilung (schwache Konvergenz). Ein erstes Beispiel haben wir bereits in Abschnitt 2.3.1 kennengelernt.

Zentral für das starke Gesetz der großen Zahl ist, dass die Varianz verschwindet für $n \to \infty$. Nun betrachten wir aber Folgen von Zufallsvariablen, für die Erwartungswert und Varianz stets gleich bleiben (0, bzw. 1 - so genannte standardisierte Zufallsvariablen). Die ist zum Beispiel der Fall für

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} = \frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n\,\mathrm{Var}(X_1)}}, \tag{4.19}$$

wobei die letzte Gleichheit identisch Verteiltheit und Unabhängigkeit voraussetzt und, wie oben, $S_n = \sum_{i=1}^n X_i$.

Wenn wir nun Verteilungen von Zufallsvariablen selbst betrachten, so sind dies Wahrscheinlichkeitsmaße auf $\mathbb{R}$ versehen mit der Borel σ-Algebra $\mathscr{B}$, siehe Definition 3.10.

Definition 4.20. Sein μ und $\mu_1, \mu_2, \ldots$ Wahrscheinlichkeitsverteilungen auf $(\mathbb{R}, \mathscr{B})$. Wir sagen μ_n *konvergiert schwach* gegen μ, falls

$$\int f d\mu_n \to \int f d\mu,$$

für alle Funktionen $f : \mathbb{R} \to \mathbb{R}$, welche stetig und beschränkt sind.

Für eine Zufallsvariable X schreiben wir F_X für die Verteilung von X. Dies führt zu folgender Definition:

Definition 4.21. Sind $X, X_1, X_2, \ldots$ Zufallsvariablen, so sagen wir (X_n) konvergiert in *Verteilung* gegen X, falls F_{X_n} schwach gegen F_X konvergiert.

Wir schreiben $X_n \xrightarrow{\mathscr{L}} X$ für die Konvergenz in Verteilung. Eine Kombination der beiden obigen Aussagen zeigt, dass dies äquivalent dazu ist, dass

$$E[f(X_n)] \to E[f(X)]$$

für alle $f \in C_b(\mathbb{R})$ (alle stetigen und beschränkten Funktionen auf $\mathbb{R}$).

B 4.22 ***Normalverteilung konvergiert gegen ein Dirac-Maß:*** Die Verteilung $\mu_n = \mathcal{N}(c, \frac{1}{n})$, die Normalverteilungen mit Mittelwert c und Varianzen $1/n$ konvergiert gegen die Verteilung μ, die in c konzentriert ist (das so genannte Dirac-Maß, charakterisiert durch $\mu(\{c\}) = 1$). Ist $\mu_{c,\sigma^2} = \mathcal{N}(c, \sigma^2)$, so erhält man mit der Substitution $y = \sigma^{-1}(x - c)$

$$\int f(x) d\mu_{c,\sigma^2} = \int f(x) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-c}{\sigma})^2} dx = \int f(c + \sigma y) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Außerdem ist

$$\int f(x) d\mu = f(c) = f(c) \int \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Es folgt

$$\int f(x) d\mu_n - \int f(x) d\mu = \frac{1}{\sqrt{2\pi}} \int \Big(f(c + n^{-1/2} y) - f(c) \Big) e^{-\frac{y^2}{2}} dy.$$

Mit dem Konvergenzsatz von Lebesgue konvergiert dieser Ausdruck gegen 0 für $n \to \infty$.

Betrachten wir nun die Verteilungsfunktionen: Die zu μ_n gehörige Verteilungsfunktion ist $F_n(x) = \Phi(\sqrt{n}(x-s))$, diejenige von μ ist $F(x) = \mathbb{1}_{\{x \geq c\}}$. Nun konvergiert für $x < c$ $F_n(x) \to 0 = F(x)$ und für $x > c$ $F_n(x) \to 1 = F(x)$, allerdings nicht für $x = c$, denn hier ist $F_n(c) = 1/2$, aber $F(c) = 1$. ◇

Mit $C_b^\infty(\mathbb{R})$ bezeichnen wir die glatten und beschränkten Funktionen auf $\mathbb{R}$ mit beschränkten Ableitungen. Zu einer Verteilung μ gehört die Verteilungsfunktion F gegeben durch $F(x) = \mu((-\infty, x])$.

Lemma 4.23. *Seien $\mu, \mu_1, \mu_2, \ldots$ Wahrscheinlichkeitsverteilungen auf $(\mathbb{R}, \mathscr{B})$ mit Verteilungsfunktionen $F, F_1, F_2, \ldots$. Dann sind äquivalent:*

(i) (μ_n) *konvergiert schwach gegen* μ,

(ii) $F_n(x)$ *konvergiert gegen* $F(x)$ *für jede Stetigkeitsstelle* x *von* F,

(iii) $\int f d\mu_n \to \int f d\mu$ *für alle* $f \in C_b^\infty(\mathbb{R})$.

Beweis. Zunächst folgt die Implikation (i) $\Rightarrow$ (iii) sofort aus Definition 4.20.

Wir zeigen (iii) $\Rightarrow$ (ii): Dazu wählen wir uns zwei Punkte x und $x + \delta$ und nutzen, dass $F(x) = \int \mathbb{1}_{(-\infty,x]} d\mu$. Nun müssen wir nur noch die Indikatorfunktionen durch ein $f \in C_b^\infty(\mathbb{R})$ approximieren und Grenzwerte bilden.

Sei dazu $f \in C_b^\infty(\mathbb{R})$ so, dass[2]

$$\mathbb{1}_{(-\infty,x]} \leq f \leq \mathbb{1}_{(-\infty,x+\delta]}.$$

Dann folgt

$$\begin{aligned} F_n(x) &= \int \mathbb{1}_{(-\infty,x]} d\mu_n \leq \int f d\mu_n, \\ F(x+\delta) &= \int \mathbb{1}_{(-\infty,x+\delta]} d\mu \geq \int f d\mu. \end{aligned}$$

Nach Voraussetzung folgt aus $f \in C_b^\infty(\mathbb{R})$, dass $\int f d\mu_n \to \int f d\mu$. Wir bilden den Grenzwert und erhalten,

$$\limsup F_n(x) \leq \lim \int f d\mu_n = \int f d\mu \leq F(x+\delta).$$

[2] In der Tat liegen die glatten Funktionen in L^p dicht, so dass man eine solche Funktion immer finden kann, siehe etwa Kapitel 18.2 in Schröder (2007). Die (stetigen) Ableitungen dieser Funktionen verschwinden außerhalb von $[x, x+\delta]$ und sind somit ebenfalls beschränkt.

Wiederholen wir diese Schritte analog für $x - \delta$ und x, mit Funktion $\widetilde{f} \in C_b^\infty(\mathbb{R})$, so erhalten wir

$$\liminf F_n(x) \geq \lim \int \widetilde{f} d\mu_n = \int \widetilde{f} d\mu \geq F(x - \delta).$$

Ist F stetig an der Stelle x, so folgt mit $\delta \to 0$, dass

$$F(x) \leq \liminf F_n(x) \leq \limsup F_n(x) \leq F(x), \tag{4.24}$$

und somit kann man $\leq$ überall durch $=$ ersetzen und (ii) folgt.

Als letztes zeigen wir, dass (ii)$\Rightarrow$(i). Die Idee ist hierbei, eine Funktion $f \in C_b(\mathbb{R})$ an genügend Stetigkeitsstellen von F zu approximieren. Zunächst beobachten wir, dass eine Verteilungsfunktion als wachsende, durch 0 und 1 beschränkte Funktion nur abzählbar viele Sprünge haben kann (in der Tat, für jedes k hat sie sogar nur endlich viele Sprünge mit einer Höhe in $(2^{-k}, 2^{-(k+1)}]$).

Wir fixieren $f \in C_b(\mathbb{R})$ und ein $\varepsilon > 0$. Die Approximation soll auf einem kompakten Intervall stattfinden, und der erste Schritt ist, ein geeignetes zu finden. Zunächst finden wir zu μ ein Intervall $[a, b]$, so dass $\mu([a, b]) > 1 - \varepsilon$ (dies folgt, da die Verteilungsfunktionen gegen 0 bzw. 1 konvergieren für $x \to -\infty$ bzw. $x \to \infty$). Da nur abzählbar viele Unstetigkeitsstellen von F existieren, können wir a und b als Stetigkeitsstellen wählen.

Weiterhin gibt es ein n_0, so dass für alle $n \geq n_0$ F_n beliebige nahe an F ist, wir finden also ebenfalls ein Intervall (wieder können wir die Grenzen als Stetigkeitsstellen wählen), so dass sogar

$$\inf_{n \geq n_0} \mu_n([a, b]) > 1 - \varepsilon$$

gilt. Für die verbleibenden endlich vielen Intervalle finden wir jeweils wieder ein solches Intervall und die Vereinigung gibt, dass wir Stetigkeitsstellen a und b von F finden können, so dass

$$\inf_n \mu_n([a, b]) > 1 - \varepsilon, \quad \mu([a, b]) > 1 - \varepsilon.$$

Auf diesem Intervall werden wir nun f geeignet approximieren. Zunächst ist f gleichmäßig stetig auf $[a, b]$, d.h. wir finden ein $\delta > 0$, so dass $|f(x) - f(y)| < \varepsilon$ für alle $a \leq x, y \leq b$ mit $|x - y| < \delta$.

Nun können wir ein genügend großes m finden und zugehörige Stetigkeitsstellen (x_i) (von F) mit $a = x_0 < x_1 < \cdots < x_n = b$ so dass jeweils $|x_i - x_{i-1}| < \delta$. Die gesuchte Approximation ist

$$f_m(x) = \sum_{i=1}^{m} f(x_{i-1}) \mathbb{1}_{(x_{i-1}, x_i]}.$$

Mit ihr gilt, dass für $x \in [a, b]$

$$|f(x) - f_m(x)| \leq \varepsilon.$$

Für $x \notin [a,b]$ ist $|f(x) - f_m(x)| = |f(x)| \leq \sup_x |f(x)|$. Damit erhalten wir, dass

$$\begin{aligned}\Big|\int (f-f_m)d\mu_n\Big| &\leq \Big|\int_a^b (f-f_m)d\mu_n\Big| + \mu_n(\overline{[a,b]})\\ &\leq \varepsilon + \varepsilon \sup_x |f(x)| = \varepsilon(1+\sup_x |f(x)|).\end{aligned}$$

Dies folgt analog, wenn wir μ_n durch μ ersetzen. Da die (x_i) Stetigkeitsstellen sind, folgt mit der Voraussetzung (ii),

$$\int f_m d\mu_n = \sum_{i=1}^m f(x_{i-1})\big(F_n(x_i) - F_n(x_{i-1})\big) \xrightarrow[n\to\infty]{} \int f_m d\mu$$

und wir finden ein n groß genug, so dass $|\int f_m d\mu_n - \int f_m d\mu_n| \leq \varepsilon$. Zusammenfassend erhalten wir

$$\begin{aligned}\Big|\int f d\mu_n - \int f d\mu\Big| &\leq \Big|\int (f-f_m)d\mu_n\Big| + \Big|\int f_m d\mu_n - \int f_m d\mu\Big| + \Big|\int (f-f_m)d\mu\Big|\\ &\leq \varepsilon(3 + 2\sup_x |f(x)|).\end{aligned}$$

Da ε beliebig war, folgt die Behauptung. □

Wir sind nun in der Lage den zentralen Grenzwertsatz zu formulieren und zu beweisen.

Satz 4.25. *Seien $(X_i)_{i\geq 1}$ i.i.d. mit $Var(X_i) < \infty$. Dann konvergiert S_n^* in Verteilung gegen eine Standardnormalverteilung.*

Wir erinnern, dass S_n^* in (4.19) eingeführt wurde. Im Folgenden setzen wir $m = E[X_1]$ und $\sigma^2 = \mathrm{Var}(X_1)$. Dann ist die Aussage des Satzes, dass für alle $x \in \mathbb{R}$

$$\lim_{n\to\infty} P\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - m}{\sigma} \leq x\Big) = \Phi(x).$$

Diesen Satz beweisen wir mit Hife eines wesentlich allgemeineren Resultates über so genannte Dreiecksschamata. Die Bedingung (iii) ist die berühmte Lindeberg-Bedingung, sie geht auf den Artikel Lindeberg (1922) zurück. Für den folgenden Satz nutzen wir (4.19) mit $S_n = \sum_{i=1}^n X_{i,n}$ (der Einfachheit halber mit der gleichen Notation).

Satz 4.26 (Lindeberg)**.** *Seien $(X_{n,i})_{1\leq i\leq n, n\in\mathbb{N}}$ Zufallsvariablen, so dass*

(i) für alle n sind $X_{n,1},\dots,X_{n,n}$ unabhängig,

(ii) *für alle n, i gilt $E[X_{n,i}] = 0$, $\mathrm{Var}(X_{n,i}) < \infty$, und für alle n ist $\sum_{i=1}^n \mathrm{Var}(X_{n,i}) = 1$,*

(iii) *für alle $\varepsilon > 0$ ist $\lim_{n\to\infty} \sum_{i=1}^n E\Big(X_{n,i}^2 \mathbb{1}_{\{|X_{n,i}|>\varepsilon\}}\Big) = 0$.*

Dann konvergiert S_n^* in Verteilung gegen eine Standardnormalverteilung.

Der Beweis nutzt, zurückgehend auf Lindeberg, einen geschickten sukzessiven Austausch der $X_{n,i}$ gegen unabhänige, normalverteilte Variable $Y_{n,i}$ mit den gleichen Momenten: Setzen wir

$$Z_{n,i} = X_{n,1} + \cdots + X_{n,i-1} + Y_{n,i+1} + \cdots + Y_{n,n}, \tag{4.27}$$

so ist

$$f(\sum_{i=1}^n X_{n,i}) - f(\sum_{i=1}^n Y_{n,i}) = \sum_{i=1}^n \Big[f(Z_{n,i+X_{n,i}}) - f(Z_{n,i} + Y_{n,i})\Big] \tag{4.28}$$

und wir können uns auf die einfacheren Terme

$$f(Z_{n,i} + X_{n,i}) - f(Z_{n,i} + Y_{n,i})$$

konzentrieren. Hierfür nutzen wir dann eine geschickte Taylor-entwicklung von f.

Lemma 4.29. *Für $f \in C_b^3(\mathbb{R})$ gilt für alle $\varepsilon > 0$, dass*

$$|f(z+x) - f(z) - f'(z)x - \frac{1}{2}f''(z)x^2| \le const \cdot x^3, \qquad und \tag{4.30}$$

$$|f(z+x) - f(z) - f'(z)x - \frac{1}{2}f''(z)x^2| \le const \cdot (\varepsilon x^2 + x^2 \mathbb{1}_{\{|x|>\varepsilon\}}). \tag{4.31}$$

Beweis. Wir wenden die Taylor-formel mehrfach auf verschiedene Weise an. Zunächst ist mit einer Taylor-Entwicklung dritter Ordnung

$$|f(z+x) - f(z) - f'(z)x - \frac{1}{2}f''(z)x^2| \le |\frac{1}{6}f'''(x+\theta z)^3 \cdot x^3|$$

mit einem $\theta \in [0,1]$. Da f''' beschränkt ist, folgt bereits (4.30). Allerdings können wir auch eine Entwicklung zweiter Ordnung nutzen, so dass ebenfalls

$$|f(z+x) - f(z) - f'(z)x - \frac{1}{2}f''(z)x^2| \le \frac{1}{2}|f''(x+\theta z)^2 \cdot x^2 - f''(z) \cdot x^2|,$$

wieder mit einem $\theta \in [0,1]$. Beschränktheit der zweiten Ableitung liefert als obere Schranke const$\cdot x^2$. Für den letzen Schritt nutzen wir beide Abschätzungen: Das ergibt $\mathbb{1}_{\{|x|\le\varepsilon\}}|x|^3 + \mathbb{1}_{\{|x|>\varepsilon\}}x^2$ als obere Schranke. Nun aber ist

$$\mathbb{1}_{\{|x|\le\varepsilon\}}|x|^3 \le \varepsilon x^2$$

und (4.31) folgt. □

Setze

$$\sigma_{n,i}^2 = \mathrm{Var}(X_{n,i}).$$

Bemerkung 4.32 (Die Lindeberg-Bedingung)**.** Die Bedingung (iii) aus dem Satz 4.26 erzwingt, dass die $X_{n,i}$ beliebig klein werden müssen, denn für alle $\varepsilon > 0$ folgt, dass

$$\max_i \sigma_{n,i}^2 \leq \max_i \Big(\varepsilon^2 + E[X_{n,i}^2 \mathbb{1}_{\{|X_{n,i}|>\varepsilon\}}]\Big) \leq \varepsilon^2 + \sum_{i=1}^n E[X_{n,i}^2 \mathbb{1}_{\{|X_{n,i}|>\varepsilon\}}]. \tag{4.33}$$

Dies bedeutet aber, dass aus (iii) folgt, dass $\max_{1\leq i\leq n} \sigma_{n,i}^2 \to 0$ für $n \to \infty$.

Wir nutzen noch eine Eigenschaft der Normalverteilung.

A 4.1 ***Drittes Moment der Normalverteilung*****:** Zeigen Sie, dass für $X \sim \mathcal{N}(0,1)$ gilt

$$E[|X|^3] = \sqrt{\frac{8}{\pi}}.$$

Beweis. (von Satz 4.26) Als Notation verwenden wir $\sigma_{n,i}^2 = \mathrm{Var}(X_{n,i}) = E[X_{n,i}^2]$. Wir nutzen Lemma 4.23(iii). Sei also $f \in C_b^\infty(\mathbb{R})$. Wie bereits angedeutet, wählen wir $Y_{n,i}$ derart, dass für jedes n

$$Y_{n,1}, \ldots, Y_{n,n}, X_{n,1}, \ldots, X_{n,n}$$

unabhängig sind und $Y_{n,i} \sim \mathcal{N}(0, \sigma_{n,i}^2)$. Weiterhin definieren wir $(Z_{n,i})$ wie in (4.27). Mit einer Taylorentwicklung von f um $Z_{n,i}$ folgt

$$\begin{aligned} f(Z_{n,i} + X_{n,i}) - f(Z_{n,i} + Y_{n,i}) &= f'(Z_{n,i})(X_{n,i} - Y_{n,i}) + \frac{1}{2} f''(Z_{n,i})(X_{n,i}^2 - Y_{n,i}^2) \\ &\quad + R(Z_{n,i}, X_{n,i}) - R(Z_{n,i}, Y_{n,i}). \end{aligned} \tag{4.34}$$

Wir erhalten

$$\begin{aligned} &\big|E[f(Z_{n,i} + X_{n,i})] - E[f(Z_{n,i} + Y_{n,i})]\big| \\ &\qquad \leq \big|E[f'(Z_{n,i})(X_{n,i} - Y_{n,i})]\big| + \frac{1}{2}\big|E[f''(Z_{n,i})(X_{n,i}^2 - Y_{n,i}^2)]\big| \\ &\qquad + E[|R(Z_{n,i}, X_{n,i})|] + E[|R(Z_{n,i}, Y_{n,i})|] =: (1) + (2) + (3) + (4). \end{aligned}$$

Wegen Unabhängigkeit und Lemma 4.29 ist

$$\begin{aligned} (1) &\leq \mathrm{const} \cdot \big|E[X_{n,i}] - E[Y_{n,i}]\big| = 0, \\ (2) &\leq \mathrm{const} \cdot \big|E[X_{n,i}^2] - E[Y_{n,i}^2]\big| = 0, \\ (3) &\leq \mathrm{const} \cdot (\varepsilon \sigma_{n,i}^2 + E[X_{n,i}^2 \mathbb{1}_{\{|X_{n,i}|>\varepsilon\}}]), \\ (4) &\leq \mathrm{const} \cdot E[|Y_{n,i}|^3] = \mathrm{const} \cdot \sqrt{\frac{8}{\pi}} \sigma_{n,i}^3. \end{aligned}$$

Schließlich summieren wir noch und erhalten durch (4.28), dass

$$\Big|E\Big[f(\sum_{i=1}^n X_{n,i}) - f(\sum_{i=1}^n Y_{n,i})\Big]\Big|$$
$$\le \text{const} \cdot \bigg(\varepsilon \sum_{i=1}^n \sigma_{n,i}^2 + \sum_{i=1}^n E[X_{n,i}^2 \mathbb{1}_{\{|X_{n,i}|>\varepsilon\}}] + \max_i \sigma_{n,i} \cdot \sum_{i=1}^n \sigma_{n,i}^2\bigg).$$

Mit $\sum \sigma_{n,i}^2 = 1$, (iii) und $\max_i \sigma_{n,i} \to 0$ nach (4.33) konvergiert dieser Ausdruck gegen $\text{const} \cdot \varepsilon$ für $n \to \infty$ und die Behauptung folgt. □

B 4.35 ***Konvergenz des Medians*:** Der Median ist ein alternativer Schätzer für den Mittelwert, der insbesondere bei Ausreißern ein robustes Verhalten zeigt (im Gegensatz etwa zu dem arithmetischen Mittel). Außerdem spielt er eine wichtige Rolle in der nichtparametrischen Statistik. Mit dem Satz von Lindeberg können wir auch für den Median eine Asymptotik ableiten.

Der *Median* einer Verteilung F ist $F^{-1}(1/2)$. Ist F konstant in einer Umgebung dieser Stelle, so gibt es eine Reihe von Möglichkeiten den Median etwa durch Mitteln der beiden Nachbarwerte festzulegen. Wir betrachten $X_1, X_2, \dots$ i.i.d. $\sim F$ mit Median $M = 0$ und nehmen an, dass $F'(0) > 0$.

Den *Stichprobenmedian* als Schätzer des Medians führen wir wie folgt ein: Wir ordnen $X_1, \dots, X_n$ der Größe nach und bezeichnen die *Ordnungsgrößen* mit

$$X_{1:n} \le \dots \le X_{n:n}.$$

Der Stichprobenmedian ist definiert durch

$$Z_n = X_{k:n}$$

mit[3] $k = \lfloor n/2 + 1 \rfloor$. Wir zeigen nun, dass $\sqrt{n} Z_N$ in Verteilung gegen $\mathcal{N}(0, \frac{1}{4(F'(0))^2})$ konvergiert: Fixiere ein $x \in \mathbb{R}$ und setze $Y_{n,i} = \mathbb{1}_{\{X_i > x/\sqrt{n}\}}$. Dann ist

$$\sqrt{n} Z_n \le x \Leftrightarrow X_{k:n} \le \frac{x}{\sqrt{n}} \Leftrightarrow \sum_{i=1}^n Y_{n,i} \le n - k. \tag{4.36}$$

Setzen wir $p_n = E[Y_{n,i}] = 1 - F(x/\sqrt{n})$, so ist $\text{Var}(Y_{n,i}) = p_n(1 - p_n)$. Wir standardisieren und betrachten

$$X_{n,i} = \frac{Y_{n,i} - p_n}{\sqrt{n\, p_n(1 - p_n)}}. \tag{4.37}$$

Wie oben sei $S_n = \sum_{i=1}^n X_{n,i}$. Aus (4.36) und (4.37) folgt, dass

$$\sqrt{n} Z_n \le x \Leftrightarrow S_n \le \frac{n - k - np_n}{\sqrt{np_n(1 - p_n)}} =: a_n.$$

[3] Mit $\lfloor x \rfloor$ bezeichnen wir den ganzzahligen Teil von x.

Da $F(0) = F(m) = F(F^{-1}(1/2)) = 1/2$ (wegen $F'(0) > 0$ ist F in einer Umgebung von 0 invertierbar) und $F(x/\sqrt{n}) = F(0) + F'(0)x/\sqrt{n} + O(n^{-1})$ erhalten wir

$$a_n = \frac{1}{\sqrt{p_n(1-p_n)}} \cdot \Big(\frac{n-k}{\sqrt{n}} - \sqrt{n}p_n\Big) \approx 2 \cdot \Big(\frac{\sqrt{n}}{2} - \sqrt{n}\frac{1}{2} + F'(0)x\Big) \to 2F'(0)x.$$

Für ein festes $\delta > 0$ und genügend großem n gilt demnach

$$P(S_n \leq 2F'(0)x - \delta) \leq P(S_n \leq a_n) = P(\sqrt{n}Z_n \leq x) \leq P(S_n \leq 2F'(0)x + \delta).$$

Der Zentrale Grenzwertsatz, Satz 4.26, liefert nun $P(S_n \leq 2F'(0)x \pm \delta) \to \Phi(2F'(0)x \pm \delta)$ und die Behauptung folgt, da δ beliebig gewählt war.

Listen der Beispiele und Aufgaben

Liste der Beispiele

Liste der Aufgaben

Literaturverzeichnis

Bauer, H. (1990). *Wahrscheinlichkeitstheorie.* Walter de Gruyter, Berlin.

Bickel, P. J. und K. A. Doksum (2001). *Mathematical Statistics: Basic Ideas and Selected Topics Vol. I* (2nd ed.). Prentice Hall.

Casella, G. und R. L. Berger (2002). *Statistical Inference* (2nd ed.). Duxbury. Pacific Grove.

Chung, K. L. (2001). *A Course in Probability Theory.* Academic Press.

Etemadi, N. (1981), 'An elementary proof of the strong law of large numbers', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **5**5(1), 119–122.

Georgii, H.-O. (2015). *Stochastik* (5th ed.). Walter de Gruyter. Berlin.

Henze, N. (2011). *Stochastik für Einsteiger: eine Einführung in die faszinierende Welt der Zufalls* Vieweg + Teubner

Irle, A. (2005). *Wahrscheinlichkeitstheorie und Statistik.* B. G. Teubner Verlag.

Jacod, J. und P. Protter (2004). *Probability Essentials.* Springer Verlag. Berlin Heidelberg New York.

Klenke, A. (2008). *Wahrscheinlichkeitstheorie* (2nd ed.). Springer Verlag. Berlin Heidelberg New York.

Kolmogorov, A. N. (1930), 'Sur la loi forte des grands nombres', *Comptes rendus de l'Académie des Sciences* **1**91, 910–912.

Lehmann, E. L. und G. Casella (1998). *Theory of Point Estimation* (2nd ed.). Springer Verlag. Berlin Heidelberg New York.

Lehmann, E. L. und J. P. Romano (2006). *Testing Statistical Hypotheses* (corr. 2nd printing ed.). Springer, New York.

Lindeberg, J. W. (1922), 'Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung', *Mathematische Zeitschrift* **1**5(1), 211–225.

Robert, C. P. und G. Casella (2008). A history of Markov chain monte carlo – subjective recollections form incomplete data. *Technical Report, University of Florida*.

Schröder, B. (2007), *Mathematical Analysis: A Concise Introduction*, Wiley.

Shao, J. (2008). *Mathematical Statistics.* Springer Verlag. Berlin Heidelberg New York.

Tappe, S. (2013). *Einführung in die Wahrscheinlichkeitstheorie*. Springer Verlag Berlin Heidelberg

Strong law of large numbers (2011), Encyclopedia of Mathematics.
URL: *http://www.encyclopediaofmath.org/index.php?title=Strong_law_of_large_numbers&oldid=26960*

Index