# Stochastik I

Mitschrift der Vorlesung

**Stochastik I**

von

**Dr. habil. H. Finner**

an der

Universität Bayreuth

im WS 95/96

Diese Mitschrift erhebt keinen Anspruch auf Richtigkeit und/oder Vollständigkeit. Sie ist kein Ersatz für das Studium einschlägiger Literatur. Das Script wurde parallel zur Vorlesung erstellt. Dabei wurde große Sorgfalt auf eine „1 zu 1"-Übernahme von der Tafel auf Papier geachtet. Trotzdem sind Übertragungsfehler sowie Fehler an der Tafel nicht auszuschließen.

(w) Mai 1996

# Inhaltsverzeichnis

**0 Einleitung** **2**

**1 Wahrscheinlichkeitsräume** **6**

1.1 Maßräume . . . 6

1.2 Rechenregeln für Maße . . . 9

1.3 Laplace-Experimente . . . 14

1.4 Elementare bedingte Wahrscheinlichkeiten . . . 18

1.5 Stochastische Unabhängigkeit von Ereignissen . . . 21

1.6 Diskrete Wahrscheinlichkeitsverteilungen . . . 24

**2 Zufallsvariable und Integrationstheorie** **26**

2.1 Zufallsvariable und stochastische Unabhängigkeit . . . 26

2.2 Integrationstheorie in Kürze . . . 37

2.3 Maße und Dichten . . . 43

2.4 Produktmaße und der Satz von Fubini . . . 45

2.5 $\lambda^n$-stetige Verteilungen: Zusammenhang Verteilungsfunktion / Dichte . . . 49

2.6 Erwartungswert, Varianz, Kovarianz, Momente und Ungleichungen . . . 51

2.7 Faltungsformel und spezielle Verteilungen . . . 60

**3 Grenzwertsätze** **68**

3.1 Fragestellungen . . . 68

3.2 Null-Eins-Gesetz . . . 69

3.3 Gesetze der großen Zahlen . . . 74

3.4 Schwache Konvergenz und zentraler Grenzwertsatz (ZGWS) . . . 83

**4 Einführung in die Statistik** **95**

4.1 Problemstellungen . . . 95

4.2 Das Fundamentallemma von Neyman und Pearson . . . 100

4.3 Monotone Dichtequotienten . . . 106

4.4 Konfidenzbereiche . . . 118

4.5 Maximum-Likelihood-Schätzer (ML-Schätzer) . . . 122

**A Diskrete Verteilungen** **124**

**B Absolut stetige Verteilungen** **125**

**C Mengen und Indikatoren** **127**

# 0 Einleitung

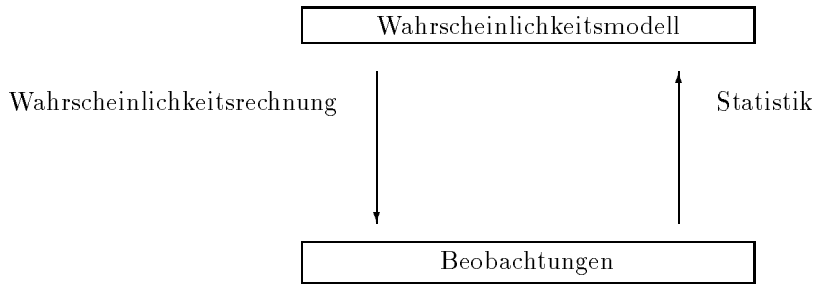
Stochastik ist

- Wahrscheinlichkeitsrechnung
- Statistik

Geschichtliche Grundlagen:

1. Glücksspiele (Würfel, Karten, Roulette) für die Wahrscheinlichkeitsrechnung (17. und 18. Jahrhundert).
2. Erhebungen für die Zwecke des Staates („Staatskunde"), Rekrutierungslisten, Steuerlisten für die Statistik.

Beziehung zwischen Wahrscheinlichkeitsrechnung und Statistik:




Lange Zeit unklar: Vernünftige Axiome zum Aufbau einer Wahrscheinlichkeitsrechnung. Die Wichtigkeit der Wahrscheinlichkeitsrechnung wurde zum Beispiel von David Hilberth (1862 - 1943) in seiner berühmten Rede 1900 auf der Pariser Weltausstellung betont. Unter 23 Problemen, die er für vordringlich hielt, nannte er als sechstes Problem:

„6. Mathematische Behandlung der Axiome der Physik: Durch die Untersuchungen über die Grundlagen der Geometrie wird uns die Aufgabe nahegelegt, nach diesem Vorbild diejenigen physikalischen Disziplinen axiomatisch zu behandeln, in denen schon heute die Mathematik eine hervorragende Rolle spielt; dies sind in erster Linie die Wahrscheinlichkeitsrechnung und die Mechanik. Was die Axiome der Wahrscheinlichkeitsrechnung angeht, so scheint es mir wünschenswert, daß mit der logischen Untersuchung derselben zugleich eine strenge und befriedigende Entwicklung der Methode der mittleren Werte in der mathematischen Physik, speziell in der kinetischen Gastheorie Hand in Hand gehe."

Typische Beispiele zufälliger Ergebnisse sind Würfeln mit den Augenzahlen $1, \ldots, 6$, Münzwurf mit den Ergebnissen 0 und 1, Ziehung der Lottozahlen 6 aus 49, aber auch Temperatur morgen um 12 Uhr oder der Dollarverlauf der kommenden Woche.

Charakteristisch: Ergebnisse sind nicht deterministisch vorhersagbar. Es lassen sich jedoch Gesetze über Häufigkeiten bzw. Wahrscheinlichkeiten von zufälligen Ergebnissen ableiten.

Schwierigkeit: Es gibt verschiedene, nicht immer verträgliche Wahrscheinlichkeitsbegriffe, davon hier nur vier:

1. Die klassische Definition

$$\text{Wahrscheinlichkeit} = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$$

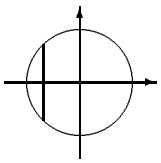
stammt von Piere de Laplace (1749 - 1827). Dieser Ansatz setzt die „Gleichwahrscheinlichkeit“ aller möglichen Fälle voraus.

**Beispiel 0.1:**

(a) Fairer Münzwurf: Wk(0)=Wk(1)=$\frac{1}{2}$
Fairer Würfel: Wk(1)=...=Wk(6)=$\frac{1}{6}$

(b) Bertrand'sches Paradoxon: Gegeben ist ein Kreis mit Radius 1. Es werde zufällig eine Sehne gezogen. Die Frage ist, wie groß die Wahrscheinlichkeit ist, daß die Sehne länger wird als eine Seite des einbeschriebenen gleichseitigen Dreiecks. Diese ist $\sqrt{3}$ lang. Mit verschiedenen Lösungsansätzen erhält man jedoch auch unterschiedliche Wahrscheinlichkeiten.

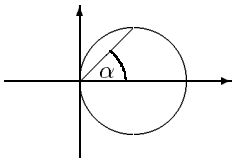
i. Wähle im oberen Halbkreis einen beliebigen Punkt und ziehe eine Sehne zum der X-Achse gegenüberliegenden Punkt.



Dann ist $\Omega = [-1, 1]$ und $A = \left[-\frac{1}{2}, \frac{1}{2}\right]$ und es gilt:

$$\text{Wahrscheinlichkeit}(A) = \frac{\text{Länge von } A}{\text{Länge von } \Omega} = \frac{1}{2}$$

ii. Ziehe die Sehne durch Wahl eines beliebigen Winkels:




Dann ist $\Omega = [0, \pi]$ und $A = \left[\frac{1}{3}\pi, \frac{2}{3}\pi\right]$ und es gilt:

$$\text{Wahrscheinlichkeit}(A) = \frac{\text{Länge von } A}{\text{Länge von } \Omega} = \frac{1}{3}$$

iii. Wähle zufällig einen Punkt aus der Kreisfläche aus. Dieser Punkt repräsentiere den Mittelpunkt der Sehne. Die Sehne ist damit genau dann größer als $\sqrt{3}$, wenn der Mittelpunkt nicht aus dem Innenkreis mit Radius 1 gewählt wurde. Es gilt somit:

$$\text{Wahrscheinlichkeit}(A) = \frac{\text{Fläche des Innenkreises mit Radius } \frac{1}{2}}{\text{Fläche des Außenkreises mit Radius } 1} = \frac{1}{4}$$

(c) Paradoxon von Chevalier de Meré (1607 - 1685), der ein Freund von Pascal und Fermat gewesen war. Zur Diskussion stehen zwei Varianten eines Würfelspiels:

i. Ein Würfel wird viermal geworfen. Ein Spieler gewinnt, wenn dabei keine „6“ auftritt. Die Gewinnwahrscheinlichkeit ist

$$\left(\frac{5}{6}\right)^4 = \frac{625}{1296} \approx 48,2\%.$$

ii. Zwei Würfel werden 24 mal geworfen. Ein Spieler gewinnt, wenn dabei keine Doppelsechs auftritt. Die Gewinnwahrscheinlichkeit ist

$$\left(\frac{35}{36}\right)^{24} \approx 50,8\%.$$

Die falsche Argumentation von de Meré ist

$$\begin{aligned} 1 - \text{GWk} &= 24 \cdot \text{Wk(Doppelsechs bei Doppelwurf)} = \frac{24}{36} = \\ &= 4 \cdot \frac{1}{6} = 4 \cdot \text{Wk(„6“ bei einem Wurf)} \end{aligned}$$

De Meré verwechselte dabei Erwartungswert und Wahrscheinlichkeit.

2.

$$\text{Wahrscheinlichkeit} = \text{Grenzwert relativer Häufigkeiten}$$

vgl. R. van Mines (1936): Wahrscheinlichkeit, Statistik und Wahrheit, Springer, Wien oder auch P. Martin-Löf (1966): Definition of random sequences, Information and Control 6, Seite 602 - 619.

Es bezeichne $h_n(A)$ die Anzahl der Versuche mit Ergebnis A (Häufigkeit) bei $n$ Versuchswiederholungen. Dann ist $\frac{1}{n}h_n(A)$ die relative Häufigkeit.

$$\frac{1}{n}h_n(A) \overset{n\to\infty}{\longrightarrow} \text{Wk}(A)$$

Problem: Existiert so ein Limes und welcher Konvergenzbegriff muß angewendet werden.

3. Subjektive Wahrscheinlichkeiten sind ein Maß des persönlichen Glaubens.

**Beispiel 0.2:**
Wir betrachten das Tennismatch Becker - Muster. Ein Fan zahlt 10,- DM, falls Muster gewinnt und gewinnt 5,- DM, falls Becker gewinnt. Was heißt das für seine subjektive Wahrscheinlichkeit $p$, daß Becker gewinnt?

$$5p - (1-p) \cdot 10 \geq 0 \quad \Longleftrightarrow \quad p \geq \frac{2}{3}$$

Problem: Die Wahrscheinlichkeit wird psychologisch angelegt. Mathematische Theorie dazu in L. J. Savage (1954), Foundations of Statistics, Wiley, New York.

4. Axiomatischer Wahrscheinlichkeitsbegriff nach A. N. Kolmogoroff (1933): Sei $\Omega$ der Ergebnisraum (Menge). $\mathcal{A} \subseteq 2^\Omega$ System von Ereignissen $A \subseteq \Omega$ mit gewissen Eigenschaften. Dabei bezeichne $2^\Omega$ die Potenzmenge von $\Omega$. Sei $P : \mathcal{A} \to [0,1]$ Wahrscheinlichkeit, Funktion mit Werten in $[0,1]$ mit gewissen Eigenschaften. $(\Omega, \mathcal{A}, P)$ ist der Wahrscheinlichkeitsraum oder das Zufallsexperiment.

**Beispiel 0.3:**
Eine Kugel wird zufällig aus einer Urne mit den Kugeln $1, \ldots, k$ gezogen. Der Ergebnisraum ist $\Omega = \{1, \ldots, k\}$, ein Elementarereignis ist $\{j\}$ mit $j \in \Omega$. Dabei steht $\{j\}$ dafür, daß die Kugel $j$ gezogen wurde. $p_j = P(\{j\})$ bezeichne die Wahrscheinlichkeit, daß die Kugel $j$ gezogen wird. Falls alle Kugeln (physikalisch) gleich sind, gilt

$$p_j = \frac{1}{k}, \quad p_j \in [0,1], \quad \sum_{j=1}^{k} p_j = 1 \qquad \forall\, j \in \Omega$$

Sei $A \subseteq \Omega$, gesucht ist die Wahrscheinlichkeit, daß die gezogene Kugel in $A$ ist.

$$\Rightarrow \quad P(A) = \sum_{j \in A} p_j = \sum_{j \in A} P(\{j\})$$

$A = 2^\Omega$. $P : \mathcal{A} \to [0,1]$ mit folgenden Eigenschaften:

(a) - Hier fehlt noch etwas -

(b) Für disjunkte $A, B \subseteq \Omega$ gilt: $P(A \cup B) = P(A) + P(B)$

**Beispiel 0.4:**
Es werden zwei Würfel geworfen.

(a) Werden die beiden Würfel unterschieden, so ist

$$\Omega = \big\{(i,j) \ : \ i,j \in \{1,\dots,6\}\big\}$$

und somit ist

$$P(\{i,j\}) = \frac{1}{36} = p_{ij}, \quad P(A) = \sum_{(i,j)\in A} p_{ij}, \quad A \in 2^{\Omega}$$

(b) Werden die beiden Würfel nicht unterschieden, so ist

$$\Omega = \big\{(i,j) \ : \ 1 \le i \le j \le 6\big\}, \qquad |\Omega| = \binom{6}{2} + 6 = 21$$

Desweiteren sieht man, daß hier keine Gleichverteilung vorliegt:

$$P(\{(i,j)\}) = \begin{cases} \frac{1}{18} & \text{falls } i < j \\ \frac{1}{36} & \text{falls } i = j \end{cases}$$

**Beispiel 0.5:**
Verabrede ein Treffen mit deinem Freund zwischen 14.00 Uhr und 14.30 Uhr. Erfahrungsgemäß hält er Verabredungen ein. Die Chance, daß er in einem Zeitintervall $[t_1,t_2]$ oder $[t_3,t_4]$ eintrifft, sei gleich groß, wenn $t_2 - t_1 = t_4 - t_3$ und $t_i \in [14, 14.5]$ ist für $i = 1,\dots,4$. Dann ist $\Omega = [14, 14.5]$ und $P(\Omega) = 1$.

$$P([t_1,t_2]) = 2 \cdot \text{Länge von } [t_1,t_2] = \int_{t_1}^{t_2} 2\,dt$$

Sei $A = [t_1,t_2]$ mit $14 \le t_1 \le t_2 \le 14.5$ und $B = [t_3,t_4]$. $A$ und $B$ seien disjunkt. Dann gilt

$$P(A \cup B) = P(A) + P(B) = \int_{t_1}^{t_2} 2\,dt + \int_{t_3}^{t_4} dt$$

Was geschieht mit $P(A)$, wenn $A \subseteq \Omega$ beliebig gewählt werden kann? Zum Beispiel ist

$$P(\{t\}) = \int_{t}^{t} 2\,dt = 0, \qquad 1 = P(\Omega) \neq \sum_{t\in\Omega} P(\{t\}) = 0$$

# 1 Wahrscheinlichkeitsräume

## 1.1 Maßräume

**Definition 1.1:**
Sei $\Omega$ eine Menge, $\mathcal{A} \subseteq 2^\Omega$.

1. $\mathcal{A}$ heißt $\sigma$-Algebra (über $\Omega$), falls

   $(\sigma 1)$ $\Omega \in \mathcal{A}$

   $(\sigma 2)$ $\forall\, A \in \mathcal{A}$ gilt $A^c \in \mathcal{A}$

   $(\sigma 3)$ $\forall\, (A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ gilt $\bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{A}$

2. Ist $\mathcal{A}$ eine $\sigma$-Algebra über $\Omega$, so heißt $(\Omega, \mathcal{A})$ Meßraum und $A \in \mathcal{A}$ Ereignis oder auch meßbare Menge (genauer $\mathcal{A}$-meßbare Menge).

3. Ersetzt man $(\sigma 3)$ durch

   $(a3)$ $\forall\, A, B \in \mathcal{A}$ gilt $A \cup B \in \mathcal{A}$

   so heißt $\mathcal{A}$ eine (Mengen-) Algebra.

**Bemerkung 1.2:**
Es ist $\mathcal{A} = 2^\Omega$ die größte $\sigma$-Algebra über $\Omega$ und $\mathcal{A} = \{\emptyset, \Omega\}$ die kleinste $\sigma$-Algebra.

**Lemma 1.3:**
Sei $\mathcal{A} \subseteq 2^\Omega$ mit

$(\sigma 1')$ $\emptyset \in \mathcal{A}$

$(\sigma 2')$ $\forall\, A \in \mathcal{A}$ gilt $A^c \in \mathcal{A}$

$(\sigma 3')$ $\forall\, (A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ gilt $\bigcap\limits_{n\in\mathbb{N}} A_n \in \mathcal{A}$

Dann ist $\mathcal{A}$ eine $\sigma$-Algebra.

**Lemma 1.4:**
Sei $\mathcal{A}' \subseteq \Omega$. Dann existiert eine kleinste $\mathcal{A}'$ umfassende $\sigma$-Algebra $\mathcal{A}$. Diese wird mit $\sigma(\mathcal{A}')$ bezeichnet.

Beweis: Sei $\mathcal{E} = \{\mathcal{B} \,:\, 2^\Omega \supseteq \mathcal{B} \supseteq \mathcal{A}' \text{ und } \mathcal{B} \text{ } \sigma\text{-Algebra}\}$ und $\mathcal{A} = \bigcap\limits_{\mathcal{B}\in\mathcal{E}}$. Dann gilt

$$\emptyset \neq \mathcal{A} \supseteq \mathcal{A}'$$

da $2^\Omega \in \mathcal{E}$ und $\mathcal{B} \supseteq \mathcal{A}'$. Es ist noch zu zeigen, daß $\mathcal{A}$ eine $\sigma$-Algebra ist.

$(\sigma 1)$ $\Omega \in \mathcal{A}$, da $\Omega \in \mathcal{B}\ \forall\, \mathcal{B} \in \mathcal{E}$

$(\sigma 2)$ $A \in \mathcal{A} \Rightarrow \forall\, \mathcal{B} \in \mathcal{E} :\ A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B}\ \forall\, \mathcal{B} \in \mathcal{E} \Rightarrow A^c \in \mathcal{A}$

$(\sigma 3)$ Sei $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A} \Rightarrow \forall\, \mathcal{B} \in \mathcal{E}$ gilt $(A_n)_{n\in\mathbb{N}} \subset \mathcal{B} \Rightarrow \forall\, \mathcal{B} \in \mathcal{E} :\ \bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{B} \Rightarrow \bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{A}$

$\mathcal{A}$ ist auch kleinste $\sigma$-Algebra mit $\mathcal{A} \supseteq \mathcal{A}'$, denn ist $\tilde{\mathcal{B}} \supseteq \mathcal{A}'$, $\tilde{\mathcal{B}}$ $\sigma$-Algebra, so ist

$$\tilde{\mathcal{B}} \supseteq \bigcap_{\mathcal{B} \in \mathcal{E}} \mathcal{B} = \mathcal{A}$$

**Definition 1.5:**
Sei $(\Omega, \mathcal{A})$ ein Meßraum. Dann heißt $\mu$ ein Maß auf $\mathcal{A}$ über $(\Omega, \mathcal{A})$, falls $\mu : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$ und

(m1) $\mu(\emptyset) = 0$ (Nulltreue)

(m2) $\forall\, A \in \mathcal{A}$ gilt $\mu(A) \geq 0$ (Nichtnegativität)

(m3) $\forall\, (A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$ und $A_n$ paarweise disjunkt gilt

$$\mu\Big( \bigcup_{n \in \mathbb{N}} A_n \Big) = \sum_{n \in \mathbb{N}} \mu(A_n) \qquad (\sigma\text{-Additivität})$$

Ist $\mu$ ein Maß auf $\mathcal{A}$, so heißt $(\Omega, \mathcal{A}, \mu)$ Maßraum. Ist $(\Omega, \mathcal{A}, P)$ ein Maßraum mit $P(\Omega) = 1$, so heißt $P$ Wahrscheinlichkeitsmaß und $(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum oder Zufallsexperiment.

**Beispiel 1.6:**
Urne mit $k$ Kugeln, vgl. Beispiel 0.3. Es ist $\Omega = \{1, \ldots, k\}$, $\mathcal{A} = 2^\Omega$, $p_j = P(\{j\})$. Für alle $A \in \mathcal{A}$ ist $P(A) = \sum\limits_{j \in A} p_j$. $P$ heißt diskrete Gleichverteilung über $(\Omega, 2^\Omega)$, falls $p_1 = \ldots = p_k = \frac{1}{k}$.

**Beispiel 1.7:**
$|\Omega| = |\mathbb{N}|$, also $\Omega$ habe abzählbar unendlich viele Elemente, ohne Einschränkung sei $\Omega = \mathbb{N}$, $\mathcal{A} = 2^\mathbb{N}$. Seien $\mu_j \in [0, \infty]$ für $j \in \mathbb{N}$. Dann ist $\mu$ definiert durch

$$\mu(A) = \sum_{j \in A} \mu_j \qquad \forall\, A \in 2^\mathbb{N}$$

ein Maß über $(\Omega, \mathcal{A})$.

Wahrscheinlichkeitsmaße $P$ über $(\Omega, \mathcal{A})$ sind also in diesem Fall charakterisiert über die Darstellung

$$P(A) = \sum_{j \in A} p_j, \qquad A \subseteq \mathbb{N}$$

wobei $p_j \in [0, 1]$ mit $\sum\limits_{j = \in \mathbb{N}} p_j = 1$.

**Beispiel 1.8 (Dirac-Maß):**
Sei $(\Omega, \mathcal{A})$ ein Meßraum, $\omega \in \Omega$ fest vorgegeben. Dann heißt $\delta_\omega$, definiert durch

$$\forall\, A \in \mathcal{A} : \quad \delta_\omega(A) := I_A(\omega) := \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Dirac-Maß oder Einpunktmaß im Punkt $\omega$. Im Beispiel 1.7 läßt sich $\mu$ mit Hilfe von Dirac-Maßen darstellen:

$$\forall\, A \in 2^\mathbb{N} : \quad \mu(A) = \sum_{j \in \mathbb{N}} \mu_j \delta_j(A)$$

**Beispiel 1.9 (Stetige Gleichverteilung, vgl. Beispiel 0.5):**
$\Omega = [a, b]$, hätten gerne $(\Omega, \mathcal{A}, P)$ mit $\mathcal{A} \supseteq I = \{[c, d] :\ a \leq c < d \leq b\}$ und

$$P([c, d]) = \frac{d - c}{b - a} \qquad \forall\, a \leq c < d \leq b.$$

Führt zu Konstruktions- und Fortsetzungsproblem: Setze $P$ vom System der Intervalle $I$ fort zu einem Wahrscheinlichkeitsmaß auf $\mathcal{A} = \sigma(I)$. $\sigma(I)$ heißt Borel-$\sigma$-Algebra auf $[a,b]$. Das Fortsetzungsproblem in der Maßtheorie behandelt zum Beispiel Bauer (1990), Maß- und Integrationstheorie, de Gruyter.

**Beispiel 1.10 ($n$-dim. Borel-$\sigma$-Algebra, Lebesgue-Maß):**
Sei $n \in \mathbb{N}$, $\Omega = \mathbb{R}^n$ und

$$I_n = \Big\{ \times_{i=1}^n [a_i, b_i] : \; a_i < b_i, \; i = 1, \ldots, n \Big\}.$$

Dann heißt $\mathbb{B}^n := \sigma(I_n)$ die $n$-dimensionale Borel-$\sigma$-Algebra. Das ist per Definition die kleinste $\sigma$-Algebra, die alle Quader (also $n$-dimensionale Intervalle) der Form $[a,b] := [a_1, b_1] \times \ldots \times [a_n, b_n]$ mit $a_i < b_i$, $1 \leq i \leq n$, enthält. Für das Volumen eines Quaders gilt

$$\mathrm{Vol}\Big( \times_{i=1}^n [a_i, b_i] \Big) = \prod_{i=1}^n (b_i - a_i)$$

**Resultat aus der Maßtheorie:**
Es gibt genau ein Maß $\lambda^n$ auf $\mathbb{B}^n$, welches jedem Intervall

$$[a,b] := \times_{i=1}^n [a_i, b_i]$$

sein Volumen zuordnet. Dieses Maß heißt $n$-dimensionales Lebesgue-Maß.

**Satz 1.10:**
Das Lebesgue-Maß $\lambda^1$ kann nicht auf $2^{\mathbb{R}}$ so fortgesetzt werden, daß gilt

$$(\star) \qquad \forall\, \nu \in \mathbb{R} : \; \forall\, E \subseteq \mathbb{R} : \qquad \lambda^1(E + \nu) = \lambda^1(E).$$

Dabei ist $E + \nu := \{x : \; \exists\, y \in E : \; x = y + \nu\}$.

Beweis: Definiere Äquivalenzrelation

$$x \sim y \iff x - y \in \mathbb{Q}.$$

Sei $E \subset [0,1)$ ein Repräsentantensystem für $\sim$.

Annahme: $\forall\, \nu \in \mathbb{R}$ gelte $\lambda(E) = \lambda(E + \nu) = \alpha \geq 0$. Es folgt

(a) $\forall\, x \in (0,1) : \; \exists\, \nu \in \mathbb{Q}$, so daß $|\nu| \leq 1$ und $x \in E + \nu$.

(b) $\forall\, r, s \in \mathbb{Q}$ mit $r \neq s = 0$ gilt

$$(E + r) \cap (E + s) = \emptyset.$$

Beweis der Zwischenbehauptung:

(a) $\exists\, y \in E : \; x \sim y \Rightarrow \exists\, \nu \in \mathbb{Q}$, so daß $|\nu| \leq 1$ und $\nu = x - y$. Es folgt: $x \in E + \nu$.

(b) Angenommen für $x, y \in E$, $r, s \in \mathbb{Q}$ mit $r \neq s$ und $x + r = y + s$ folgt $x - y = s - r \in \mathbb{Q} \Rightarrow x \sim y$, d.h. $x = y$, da $E$ ein Repräsentantensystem ist. Somit ist $s = r$, was aber ein Widerspruch ist.

Aus (b) folgt, daß $S := \bigcup\limits_{\substack{\nu \in \mathbb{Q} \\ |\nu| \leq 1}} (E + \nu)$ eine abzählbare Vereinigung von paarweise disjunkten Mengen ist.
Wegen $S \subset [-1, 2]$ ist $\lambda(S) \leq 3$ und

$$\lambda(S) = \sum_{\substack{r \in \mathbb{Q} \\ |r| \leq 1}} \underbrace{\lambda(E + r)}_{\alpha},$$

also $\alpha = 0$ und somit $\lambda(S) = 0$. Aus (a) folgt jedoch, daß $(0,1) \subseteq S$, d.h. $\lambda(S) \geq 1$, was aber ein Widerspruch ist.

**Bemerkung 1.11:**

(i) Für alle $n \in \mathbb{N}$ gilt $\mathbb{B}^n \neq 2^{\mathbb{R}}$.

(ii) Die Elemente von $\mathbb{B}^n$ lassen sich nicht konstruktiv angeben.

**Bemerkung 1.12:**
Axiome von Kolmogoroff 1933: Es ist $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum genau dann, wenn

(i) $\Omega \neq \emptyset$

(ii) $\mathcal{A}$ ist $\sigma$-Algebra (wie oben definiert)

(iii) $P$ ist ein Wahrscheinlichkeitsmaß (wie oben definiert)

## 1.2 Rechenregeln für Maße

Häufig sind Wahrscheinlichkeiten von Ausdrücken wie

$$\bigcup_{i \in \mathbb{N}} A_i, \qquad \bigcap_{i \in \mathbb{N}} A_i, \qquad A \setminus B, \qquad \text{etc.}$$

zu berechnen. Geschicktes Umrechnen erleichtert manchmal die Aufgabe. Man muß auch wissen, ob die Mengen überhaupt meßbar sind, d.h. in A liegen.

**Lemma 1.13 (de Morgan):**
Sei $I$ eine beliebige Indexmenge, $(A_i)_{i \in I} \subset 2^{\Omega}$. Dann gilt

(a) $\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c$

(a) $\left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c$

**Lemma 1.14 (Eigenschaften von $\sigma$-Algebren):**
Sei $(\Omega, \mathcal{A})$ ein Meßraum.

(a) $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A} \quad \Rightarrow \quad \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$

(b) $A_1, \ldots, A_n \in \mathcal{A} \quad \Rightarrow \quad \bigcap_{i=1}^{n} A_i \in \mathcal{A}, \quad \bigcup_{i=1}^{n} A_i \in \mathcal{A}$

(c) $A, B \in \mathcal{A} \quad \Rightarrow \quad A \setminus B \in \mathcal{A}$

(d) $A, B \in \mathcal{A} \quad \Rightarrow \quad A \Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B) \in \mathcal{A}$ (Symmetrische Differenz)

(e) Ist $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$, dann ist

$$\limsup_{n \to \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \in \mathcal{A}$$

$$\liminf_{n \to \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} a_m \in \mathcal{A}$$

Beweis: selbst als Übung.

**Bemerkung 1.15:**
Es ist

$$\begin{aligned} \limsup_{n\to\infty} A_n &= \{\omega \in \Omega : \omega \in A_n \text{ für unendlich viele } n\} \\ \liminf_{n\to\infty} A_n &= \{\omega \in \Omega : \omega \in A_n \text{ für alle bis auf endlich viele } n\} \end{aligned}$$

Wiederholt man zum Beispiel den Wurf einer fairen Münze unendlich oft, so interessiert zum Beispiel die Wahrscheinlichkeiten für die Ereignisse, daß

- die 1 unendlich oft erscheint oder
- die 1 nur endlich oft erscheint.

Dann ist $\Omega = \{0,1\}^{\mathbb{N}}$ und $A_n = \{\omega = (\omega_1, \omega_2, \ldots) \in \Omega : \omega_n = 1\}$ Es interessieren

$$P(\limsup_{n\to\infty} A_n) = ?, \qquad P(\liminf_{n\to\infty} A_n) = ?.$$

**Satz 1.16 (Rechenregeln und Ungleichungen für Maße):**
Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $A, B \in \mathcal{A}$ und $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$. Dann gelten die folgenden Aussagen:

(a) Sind $A_1, \ldots, A_n$ paarweise disjunkt, gilt die endliche Additivität:

$$\mu\Big(\sum_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mu(A_i).$$

(b) Ist $A \subseteq B$, so ist $\mu(A) \leq \mu(B)$ (Isotonie)

(c) Ist $A \subseteq B$ und $\mu(A) < \infty$, folgt $\mu(B - A) = \mu(B) - \mu(A)$ (Substraktivität)

(d) Ist $\mu(A_i) < \infty$ für alle $i = 1, \ldots, n$, so gilt

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leq i_1 < \ldots < i_k \leq n} \mu\Big(\bigcap_{j=1}^{k} A_{i_j}\Big)$$

(Siebformel von Sylvester/Poincare)

(e) Es gilt die Sub-$\sigma$-Additivität

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

Beweis:

(a) Wähle $A_j = \emptyset \; \forall \; j \geq n+1$. Mit (m3) folgt die Behauptung.

(b) Mit (a) und dem Trick $B = (B - A) + A$ folgt die Behauptung.

(c) Beachte die Konvention $\infty - \nu = \infty \; \forall \; |\nu| < \infty$, dann ist die Behauptung klar.

(d) siehe Übungsaufgabe

(e) Sei $B_n := A_n \setminus \bigcup_{j=1}^{n-1} A_j$, $n \in \mathbb{N}$. Dann sind die $(B_n)_{n\in\mathbb{N}}$ paarweise disjunkt und $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$. Dann gilt

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \;=\; \mu\Big(\bigcup_{n=1}^{\infty} B_n\Big) \overset{(\mathrm{m3})}{=} \sum_{n=1}^{\infty} \mu(B_n) \le \sum_{n=1}^{\infty} \mu(A_n)$$

**Folgerung 1.17:**
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Dann gilt

(a) Für alle $A, B \in \mathcal{A}$ gilt
$$A \subseteq B \quad \Rightarrow \quad P(B - A) = P(B) - P(A).$$

(b) Für alle $A \in \mathcal{A}$ gilt: $P(A) = 1 - P(A^c)$.

(c) Ist $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$, dann gilt die Sub-Additivität:

(i) $P\big(\bigcup_{n=1}^{\infty} A_n\big) \le \sum_{n=1}^{\infty} P(A_n)$

(ii) $P\big(\bigcup_{n=1}^{k} A_n\big) \le \sum_{n=1}^{k} P(A_n)$

Die Ungleichung (ii) wird auch Bonterroni-Ungleichung genannt, insbesondere wenn

$$P(A_n) \le \frac{\alpha}{k} \qquad \text{für ein } \alpha \in (0,1).$$

Dann ist

$$P\Big(\bigcup_{n=1}^{k} A_n\Big) \le \sum_{n=1}^{k} \frac{\alpha}{k} = \alpha.$$

(d) Ist $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$, dann gilt

$$P\Big(\bigcap_{n=1}^{k} A_n\Big) \ge 1 - \sum_{n=1}^{k} \big(1 - P(A_n)\big).$$

**Bemerkung 1.18:**
Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Dann heißt $\mu$

(a) endlich (oder endliches Maß), falls $\mu(\Omega) < \infty$,

(b) $\sigma$-endlich (oder $\sigma$-endliches Maß), falls

$$\exists\, (A_n)_{n\in\mathbb{N}} \subset \mathcal{A} : \qquad A_n \uparrow \Omega \quad \text{und} \quad \forall\, n \in \mathbb{N} : \;\; \mu(A_n) < \infty.$$

**Beispiel 1.19 (Recente-Problem - Anwendung der Siebformel):**
$n$ Briefe $B_1, \ldots, B_n$ werden zufällig in $n$ Briefumschläge $U_1, \ldots, U_n$ gesteckt. Wie groß ist die Wahrscheinlichkeit, daß kein Brief im richtigen Umschlag steckt, d.h. $B_i$ nicht in $U_i$ für $i = 1, \ldots, n$?

Ansatz:

$$\Omega = \big\{(\omega_1, \ldots, \omega_n) : \{\omega_1, \ldots, \omega_n\} = \{1, \ldots, n\}\big\} \hat{=} \text{ Menge der Permutationen von } (1, \ldots, n)$$

Dann ist $P(\{\omega\}) = \frac{1}{n!}$ und $\mathcal{A} = 2^{\Omega}$ sowie $(\Omega, \mathcal{A}, P)$ ist ein Wahrscheinlichkeitsraum.

Interpretation: Ein Brief $B_{\omega_i}$ landet im Umschlag $U_i,\ i=1,\dots,n \iff \omega=(\omega_1,\dots,\omega_n)$.

$$\begin{aligned} A &\hat{=} \{\text{kein Brief im richtigen Umschlag}\} = \{\omega\in\Omega:\ \omega_i\neq i\ \forall\, i=1,\dots,n\} \\ A_i &\hat{=} \{\text{Brief } i \text{ im richtigen Umschlag}\} = \{\omega\in\Omega:\ \omega_i = i\} \end{aligned}$$

Gesucht ist $P(A)$. Es gilt: $P(A)=1-P(A^c)=1-P\big(\bigcup_{i=1}^n A_i\big)$. Ein wenig Kombinatorik (siehe später) ergibt:

$$\begin{aligned} P(A_i) &= \frac{1}{n} \\ P(A_i\cap A_j) &= \frac{1}{n(n-1)} \quad \forall\, i\neq j \\ P(A_{i_1}\cap\ldots\cap A_{i_k}) &= \frac{(n-k)!}{n!} \quad i_1,\dots,i_k\in\{1,\dots,n\} \end{aligned}$$

Folglich gilt

$$\begin{aligned} P(A) &= 1-P\Big(\bigcup_{i=1}^n A_i\Big) = \\ &= 1-\sum_{k=1}^n (-1)^{k+1} \sum_{1\le i_1<\ldots<i_k\le n} P(A_{i_1}\cap\ldots\cap A_{i_k}) = \\ &= 1-\sum_{k=1}^n (-1)^{k+1}\frac{(n-k)!}{n!} \underbrace{\sum_{1\le i_1<\ldots<i_k\le n} 1}_{\binom{n}{k}} = \\ &= 1-\sum_{k=1}^n (-1)^{k+1}\frac{1}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!} \overset{n\to\infty}{\longrightarrow} \sum_{k=0}^\infty \frac{(-1)^k}{k!} = e^{-1} \end{aligned}$$

Wir sehen

| $n$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $P(A)$ | 0 | $\frac{1}{2}$ | $\frac{1}{3}$ | $\frac{3}{8}$ | $\frac{11}{30}$ |

**Bemerkung:**
Es gilt:

$$\sum_{k=0}^{2n-1}\frac{(-1)^k}{k!} < e^{-1} < \sum_{k=0}^{2n}\frac{(-1)^k}{k!}$$

**Vorbemerkung zur Konvergenz von Mengen:**
Monotone Mengenfolgen konvergieren:

$$(A_n)_{n\in\mathbb{N}}\uparrow,\ \text{d.h. } A_1\subseteq A_2\subseteq\ldots\ (\text{isoton}) \ \Rightarrow\ \lim_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n$$

$$(A_n)_{n\in\mathbb{N}}\downarrow,\ \text{d.h. } A_1\supseteq A_2\supseteq\ldots\ (\text{antiton}) \ \Rightarrow\ \lim_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} A_n$$

**Satz 1.20 (Stetigkeit von Maßen):**
Seien $(\Omega,\mathcal{A},\mu)$ ein Maßraum und $(A_n)_{n\in\mathbb{N}}\subset\mathcal{A}$. Dann gilt

(a) Ist $(A_n)_{n\in\mathbb{N}}\uparrow \Rightarrow \lim_{n\to\infty}\mu(A_n)=\mu\big(\bigcup_{n\in\mathbb{N}} A_n\big)$ (Stetigkeit von unten)

(b) Ist $(A_n)_{n\in\mathbb{N}} \downarrow$ und existiert ein $k \in \mathbb{N}$, so daß $\mu(A_k) < \infty$, dann gilt die Stetigkeit von oben:

$$\lim_{n\in\mathbb{N}} \mu(A_n) = \mu\Big(\bigcap_{n\in\mathbb{N}} A_n\Big).$$

(c) Existiert $\lim_{n\to\infty} A_n$ und für alle $n \in \mathbb{N}$ gilt $\mu(A_n) < \infty$, so gilt die Stetigkeit

$$\lim_{n\to\infty} \mu(A_n) = \mu\big(\lim_{n\to\infty} A_n\big).$$

Beweis:

(a) Falls $\mu(A_n) = \infty$ für ein $n \in \mathbb{N}$, so folgt wegen $A_n \uparrow$ sofort

$$\lim_{n\to\infty} \mu(A_n) = \infty = \mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big).$$

Sei nun $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$. Aus $A_n \uparrow = 0$ folgt

$$\lim_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n = A_1 + A_2A_1^c + A_3A_2^c + \ldots = \underbrace{\sum_{n=1}^{\infty} A_nA_{n-1}^c}_{\in\mathcal{A}}$$

mit $A_0 := \emptyset$. Es folgt

$$\begin{aligned}\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) &= \mu\Big(\sum_{n=1}^{\infty} A_nA_{n-1}^c\Big) = \sum \mu(A_nA_{n-1}^c) = \\ &= \sum_{n=1}^{\infty} \mu(A_n - A_{n-1}) = \lim_{n\to\infty} \underbrace{\sum_{k=1}^{n} \big[\mu(A_n) - \mu(A_{n-1})\big]}_{\text{Teleskopsumme}} = \lim_{n\to\infty} \mu(A_n)\end{aligned}$$

(b) Ohne Einschränkung ist $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$. Definiere $B_n := A_1 - A_n$ für alle $n \in \mathbb{N}$. Dann gilt: $(B_n)_{n\in\mathbb{N}} \uparrow$.

$$\begin{aligned}\mu(A_1) - \mu(\bigcap A_n) &= \mu(A_1 - \bigcap A_n) = \mu\Big(A_1\big(\bigcap A_n\big)^c\Big) = \mu(A_1 \bigcup A_n^c) = \\ &= \mu\big(\bigcup (A_1A_n^c)\big) = \mu\big(\bigcup B_n\big) \overset{(a)}{=} \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \mu(A_1 - A_n) = \\ &= \mu(A_1) - \lim_{n\to\infty} \mu(A_n)\end{aligned}$$

(c) Der $\lim A_n$ existiert genau dann, wenn $\limsup A_n = \liminf A_n = \lim A_n$ ist. Definiere $B_n := \bigcup_{m\geq n} A_m$ und $C_n := \bigcap_{m\geq n} A_m$. Dann gilt $(B_n) \downarrow$ und $(C_n) \uparrow$ und es folgt

$$\begin{aligned}\mu(\liminf A_n) &= \mu(\lim C_n) \overset{(a)}{=} \lim \mu(C_n) = \liminf \mu(C_n) = \liminf \mu(C_n) = \\ &= \liminf \mu\Big(\bigcap_{m\geq n} A_m\Big) \overset{(Isot.)}{\leq} \liminf \mu(A_n) \leq \limsup \mu(A_n) \leq \\ &\overset{(Isot.)}{\leq} \limsup \mu\Big(\bigcup_{m\geq n} A_m\Big) = \limsup \mu(B_n) = \lim \mu(B_n) \overset{(a)}{=} \mu(\lim B_n) = \\ &= \mu(\limsup A_n) = \mu(\liminf A_n)\end{aligned}$$

## 1.3 Laplace-Experimente

**Definition 1.21:**
$(\Omega, 2^\Omega, P)$ oder kurz $(\Omega, P)$ heißt Laplace-Experiment genau dann, wenn

(i) $\Omega \neq \emptyset$, $\Omega$ endlich, d.h. $0 < |\Omega| < |\mathbb{N}|$.

(ii) Für alle $A \in \Omega$ gilt: $P(A) = \frac{|A|}{|\Omega|}$.

$P$ heißt (diskrete) Gleichverteilung oder Laplace-Verteilung auf $\Omega$. Kurz $P = G(\Omega)$ und $P(A)$ ist die Wahrscheinlichkeit von $A$.

$$p : \Omega \to [0,1] \qquad \text{mit } p(\omega) := P(\{\omega\})$$

heißt Wahrscheinlichkeitsdichte oder Zähldichte von $P$.

**Rechenregeln für Laplace-Experimente:**

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \sum_{\omega \in A} \frac{1}{|\Omega|} = \sum_{\omega \in A} p(\omega)$$

Beachte: $P : 2^\Omega \to [0,1]$, dagegen $p : \Omega \to [0,1]$ und $P(A) = \sum\limits_{\omega \in A} p(\omega)$ mit $p(\omega) = P(\{\omega\})$.

Einfache Beispiele für Laplace-Experimente sind:

(i) Münzwurf: $\Omega = \{0,1\}$, $p(\omega) = \frac{1}{2}$

(ii) Roulette: $\Omega = \{0,1,\ldots,36\}$, $p(\omega) = \frac{1}{37}$

Man benötigt für Laplace-Experimente die richtige Gleichverteilungsannahme und die Kunst des Zählens (Kombinatorik), s.u.

**Beispiel 1.22 (Skat):**
Wir haben 32 verschiedene Karten, 3 Spieler bekommen je 10 Karten. Die restlichen 2 Karten landen im Skat. Wie groß ist die Wahrscheinlichkeit, daß ein Spieler beim Geben alle vier Buben erhält?

Möglicher Ansatz:

$$\begin{aligned} \Omega &\hat{=} \text{ alle möglichen } (10,10,10,2)\text{-Aufteilungen} \\ &= \Big\{(S_1,\ldots,S_4) : \sum_{i=1}^{4} S_i = S,\ |S_1| = \ldots = |S_3| = 10\Big\} \end{aligned}$$

wobei $S$ die Menge der Karten des Kartenspiels ist.

$$\begin{aligned} A &\hat{=} \text{ ein Spieler erhält alle vier Buben } B_1,\ldots,B_4 \in S \\ &= A_1 + A_2 + A_3 \end{aligned}$$

mit $A_i \hat{=} i-$ter Spieler erhält alle vier Buben, d.h.

$$A_i = \big\{(S_1,\ldots,S_4) \in \Omega : \{B_1,\ldots,B_4\} \in S_i\big\}.$$

Es folgt $|A| = 3 \cdot |A_1| = 3 \cdot \binom{28}{6}\binom{22}{10}\binom{12}{10}\binom{2}{2}$ und somit

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3 \cdot \binom{28}{6}\binom{22}{10}\binom{12}{10}\binom{2}{2}}{\binom{32}{10}\binom{22}{10}\binom{12}{10}\binom{2}{2}} = \frac{3\binom{28}{6}}{\binom{32}{10}} = \frac{3 \cdot 28! \cdot 10! \cdot 22!}{6! \cdot 22! \cdot 32!} = \frac{63}{3596} \approx \frac{1}{57}$$

**Beispiel 1.23:**
Gleichzeitiger Wurf von drei fairen Würfeln. Sind die Wahrscheinlichkeiten für Augensumme 11 und 12 gleich?

| $\sum$ | Augenpaare | | | | | |
|---|---|---|---|---|---|---|
| 11 | (641) | (632) | (551) | (542) | (533) | (443) |
| 12 | (651) | (642) | (633) | (552) | (543) | (444) |

Chevalier de Meré war sicher, daß $P(\sum = 11) > P(\sum = 12)$. Die richtige Erklärung von Pascal:

$$\Omega := \{\omega \in \Omega_0 : \omega_1 \geq \omega_2 \geq \omega_3\} \qquad \text{mit } \Omega_0 := \{1, \ldots, 6\}^3$$

Wählt man $\Omega$, so ist $(\Omega, P)$ kein Laplace-Experiment, wohl aber $(\Omega_0, P_0)$. Denkt man sich die Würfel unterscheidbar, so steckt zum Beispiel hinter

$$\begin{aligned}
&(641) \in \Omega && \text{einer von } (641), (614), (461), (416), (146), (164)\\
&(533) \in \Omega && \text{einer von } (533), (353), (335)\\
&(444) \in \Omega && \text{einer von } (444) \text{ in } \Omega_0
\end{aligned}$$

Dann ist

$$\begin{aligned}
P(\textstyle\sum = 11) &= P_0(\textstyle\sum = 11) = \dfrac{|\{\omega_0 \in \Omega_0 : \sum_{i=1}^{3} \omega_{0_i} = 11\}|}{|\Omega_0|} =\\
&= \frac{1}{6^3}\big(6 \cdot |\{(641), (632), (542)\}| + 3 \cdot |\{(551), (533), (443)\}|\big) = \frac{27}{6^3} = 0.125\\
P(\textstyle\sum = 12) &= P_0(\textstyle\sum = 12) = \frac{1}{6^3}\big(6 \cdot 3 + 3 \cdot 2 + 1 \cdot 1\big) = \frac{25}{6^3} \approx 0.116
\end{aligned}$$

**Kombinatorische Hilfsmittel:**

**Definition/Lemma 1.24:**
Seien $r, n \in \mathbb{N}$, $M = \{x_1, \ldots, x_n\}$ und

$$\begin{aligned}
M^r &:= \underbrace{M \times \ldots \times M}_{r-\text{mal}}\\
M^{(r)} &:= \big\{(a_1, \ldots, a_r) \in M^r : a_i \neq a_j \text{ für } i \neq j,\ i, j \in \{1, \ldots, r\}\big\}\\
K^r &:= \big\{(x_{j_1}, \ldots, x_{j_r}) \in M^r : i_1 \leq \ldots \leq i_r\big\}\\
K^{(r)} &:= \big\{(x_{j_1}, \ldots, x_{j_r}) \in M^r : i_1 < \ldots < i_r\big\}.
\end{aligned}$$

Dann heißt jedes $r$-Tupel $(a_1, \ldots, a_r) \in$

(a) $M^r$ $r$-Permutation mit Wiederholung,

(b) $M^{(r)}$ $r$-Permutation ohne Wiederholung (falls $r \leq n$),

(c) $K^r$ $r$-Kombination mit Wiederholung,

(d) $K^{(r)}$ $r$-Kombination ohne Wiederholung (falls $r \leq n$).

Es gilt:

(a) $|M^r| = n^r$

(b) $|M^{(r)}| = \frac{n!}{(n-1)!} = r!\binom{n}{r}$

(c) $|K^r| = \binom{n+r-1}{r} = \binom{n+r-1}{n-1}$

(d) $|K^{(r)}| = \binom{n}{r}$,

d.h. aus einer Menge $M = \{x_1, \ldots, x_n\}$ kann man auf

(a) $n^r$ verschiedene Arten geordnete Proben mit Wiederholung vom Umfang $r$ entnehmen,

(b) $\frac{n!}{(n-r)!}$ verschiedene Arten geordnete Proben ohne Wiederholung vom Umfang $r$ entnehmen,

(c) $\binom{n+r-1}{r}$ verschiedene Arten ungeordnete Proben mit Wiederholung vom Umfang $r$ entnehmen,

(d) $\binom{n}{r}$ verschiedene Arten ungeordnete Proben ohne Wiederholung vom Umfang $r$ entnehmen.

Seien weiter $i_j \in \mathbb{N}$, $j = 1, \ldots, r$, mit $\sum_{j=1}^{r} i_j \leq n$ und

$$Q^{i_1,\ldots,i_r} = \{(M_1, \ldots, M_r) : M_j \subseteq M \text{ paarw. disjunkt}, |M_j| = i_j, j = 1, \ldots, r\}$$

Dann gilt

(e) $|Q^{i_1,\ldots,i_r}| = \binom{n}{i_1}\binom{n-i_1}{i_2}\binom{n-i_1-i_2}{i_3}\cdots\binom{n-\sum_{j=1}^{r-1} i_j}{i_r} = \frac{n!}{i_1!\cdots i_r!(n-\sum_{j=1}^{r} i_j)!}$

Falls $\sum_{j=1}^{r} i_j = n$ ist, so heißt $\binom{n}{i_1 \cdots i_r} := \frac{n!}{i_1!\cdots i_r!}$ Multinominalkoeffizient.

**Bemerkung 1.25:**
Es sei eine Urne mit den Kugeln $1, \ldots, n$ gegeben. Es werden $r$ Ziehungen durchgeführt.

(a) Ziehen mit Zurücklegen mit Berücksichtigung der Reihenfolge entspricht $M^r$.

(b) Ziehen ohne Zurücklegen mit Berücksichtigung der Reihenfolge entspricht $M^{(r)}$.

(c) Ziehen mit Zurücklegen ohne Berücksichtigung der Reihenfolge entspricht $K^r$.

(d) Ziehen ohne Zurücklegen ohne Berücksichtigung der Reihenfolge entspricht $K^{(r)}$.

**Beispiel:**
Sei $n = 3$, $r = 2$ und $M = \{1, 2, 3\}$.

| Beispiel | Augenpaare | Formel |
|---|---|---|
| (a) | (11), (12), (13), (21), (22), (23), (31), (32), (33) | $n^r = 3^2 = 9$ |
| (b) | (12), (13), (21), (23), (31), (32) | $\frac{n!}{(n-r)!} = \frac{3!}{1!} = 6$ |
| (c) | (11), (12), (13), (22), (23), (33) | $\binom{n+r-1}{r} = \binom{4}{2} = 6$ |
| (d) | (12), (13), (23) | $\binom{n}{r} = \binom{3}{2} = 3$ |

**Bemerkung 1.26:**
Seien $\Omega_i \neq \emptyset$ Mengen mit $|\Omega_i| = n_i \in \mathbb{N}$ für $i = 1, \ldots, n$. Es sei $\Omega = \Omega_1 \times \ldots \times \Omega_n$. Dann ist

$$|\Omega| = \prod_{i=1}^{n} |\Omega_i| = \prod_{i=1}^{n} n_i.$$

Beweis von 1.24:

(a) ist Spezialfall von Bemerkung 1.26.

(b) per Induktion nach $r$:

$$n(n-1)\cdot\ldots\cdot(n-r+1) = \frac{n!}{(n-r)!} = r!\binom{n}{r}$$

(d) $|K^{(r)}|$ ist die Anzahl aller $r$-elementigen Teilmengen von $M$. Wegen $|M^{(r)}| = r!\cdot|K^{(r)}|$ ist nach (b): $|K^{(r)}| = \binom{n}{r}$.

(c) Ohne Einschränkung ist $M = \{1,\ldots,r\}$, $a = (a_1,\ldots,a_r) \in K^r$, d.h. $a_1 \le \ldots \le a_r$. Definiere $A := \{1,2,\ldots,n+r-1\}$ und dann

$$\mathcal{P}_r(A) := \{B : B \subset A,\ |B| = r\}, \qquad f : K^r \to \mathcal{P}_r(A), \quad (a_1,\ldots,a_r) \mapsto \{a_1, a_2+1, \ldots, a_r+r-1\}$$

Es folgt, daß $f$ bijektiv ist, d.h. es gilt bezüglich der Elementanzahl Gleichheit:

$$|K^r| = |\mathcal{P}_r(A)| \overset{(d)}{=} \binom{n+r-1}{r}$$

(e) per Induktion nach $r$ zu zeigen.

**Bemerkung 1.27:**
Seien $A, B$ Mengen mit $|A| = r$, $|B| = n$ mit $r \le n$ und $r, n \in \mathbb{N}$. Dann gilt

(a) $|\{f : A \to A,\ f \text{ bijektiv}\}| = r!$

(b) $|\{f : A \to B,\ f \text{ injektiv}\}| = \frac{n!}{(n-r)!}$

(c) $|\{f : A \to B\}| = n^r$ und

(d) $|\{f : B \to A,\ f \text{ surjektiv}\}| = \sum_{k=1}^{r} (-1)^k \binom{r}{k}(r-k)^n$

**Beispiel 1.28 (Geburtstage):**
Wie groß ist die Wahrscheinlichkeit, daß mindestens zwei von $r$ Personen am gleichen Tag Geburtstag haben. Dabei sollen Mehrlingsgeburten und Schaltjahre unberücksichtigt bleiben. Es wird angenommen, daß alle möglichen Geburtstage eines Jahres gleichwahrscheinlich sind.

Ansatz: $\Omega_0 = \{1,\ldots,n\}$ mit $n = 365$. Ein $i \in \Omega_0$ entspricht dem $i$-ten Tag des Jahres. Wir setzen $\Omega = \Omega_0^r$. Es entspricht $\omega = (\omega_1,\ldots,\omega_r)$, daß die $i$-te Person am Tag $i$ Geburtstag hat. Dann entspricht $P_{n,r}$ der Laplace-Verteilung auf $\Omega$.

$$E := \{(\omega_1,\ldots,\omega_r) \in \Omega : \exists\, i,j \in \{1,\ldots,r\} \text{ mit } i \ne j \text{ und } \omega_i = \omega_j\}$$

Gesucht ist $P_{n,r}(E)$.

$$\begin{aligned} E^c &= \{(\omega_1,\ldots,\omega_r) \in \Omega : \forall\, 1 \le i < j \le r : \omega_i = \omega_j\} \\ &\hat{=} \text{ Ziehen ohne Zurücklegen aus } \Omega_0 \text{ mit Berücksichtigung der Reihenfolge} \\ &\hat{=} \text{ alle } i\text{-Permutationen ohne Wiederholung von } \{\omega_1,\ldots,\omega_r\} \end{aligned}$$

Somit folgt

$$\begin{aligned} P_{n,r}(E^c) &= \frac{|E^c|}{|\Omega|} = \frac{n!}{(n-1)!\cdot n^r} = \frac{n(n-1)\cdots(n-r+1)}{n\cdot n\cdots n} = \\ &= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{r-1}{n}\right) = e^{\log\left(1-\frac{1}{n}\right)}\cdot\ldots\cdot e^{\log\left(1-\frac{r-1}{n}\right)} \le \\ &\le e^{-\frac{1}{n}}\cdot\ldots\cdot e^{-\frac{r-1}{n}} = \exp\left(-\sum_{i=1}^{r-1}\frac{i}{n}\right) = \exp\left(-\frac{1}{2n}r(r-1)\right) \end{aligned}$$

und damit insgesamt

$$P_{n,r}(E) = 1 - P_{n,r}(E^c) \geq 1 - \exp\left(-\frac{r(r-1)}{2n}\right)$$

So sieht man zum Beispiel, daß $P_{n,r}(E) > \frac{1}{2}$ für $r > 23$, oder auch

| $r$ | $P_{n,r}(E)$ |
|---|---|
| 30 | $\geq 0.696$ |
| 60 | $\geq 0.992$ |
| 183 | $\geq 1 - 4.78 \cdot 10^{-27}$ |

Literatur: Plachky, Baringshaus, Schmitz: Stochastik I, Akad. Verlagsgesellschaft Wiesbaden.

## 1.4 Elementare bedingte Wahrscheinlichkeiten

**Beispiel 1.29:**
Frau W. hat gewürfelt und teilt uns nur mit, daß sie eine gerade Zahl gewürfelt hat. Wie groß ist die Wahrscheinlichkeit, daß es eine „6“ ist, bzw. daß es eine Zahl größer gleich „4“ ist?

Ansatz: $\Omega = \{1, \ldots, 6\}$, $P = G(\Omega)$, $A_1 = \{6\}$, $A_2 = \{4, 5, 6\}$ und $B = \{2, 4, 6\}$.

$$P(A_1) = \frac{1}{6}, \qquad P(A_1|B) = \frac{1}{3} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{P(A_1 B)}{P(B)}$$

beziehungsweise

$$P(A_2) = \frac{1}{2}, \qquad P(A_2|B) = \frac{2}{3} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{P(A_2 B)}{P(B)}$$

**Definition 1.30:**
Gegeben sie ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $A, B \in \mathcal{A}$ und $P(B) > 0$. Dann heißt

$$P(\cdot|B) : \mathcal{A} \to [0, 1], \text{ definiert durch } P(A|B) := \frac{P(AB)}{P(B)},$$

die (elementare) bedingte Wahrscheinlichkeit von $A$ gegeben $B$.

**Beispiel (Augen- und Haarfarbe, $2 \times 2$-Tafel):**
Gegeben sei folgende Tabelle der Dortmunder Statistikstudenten (1. Sem) von 1979/80:

| Haare $\downarrow$ \ Augen $\rightarrow$ | blau | ¬blau | $\sum$ |
|---|---|---|---|
| schwarz | 1 | 5 | 6 |
| ¬schwarz | 17 | 17 | 34 |
| $\sum$ | 18 | 22 | 40 |

Wie groß ist die Wahrscheinlichkeit, daß ein zufällig getroffener Dortmunder Statistikstudent mit schwarzen Haaren auch blaue Augen hat? Wir setzen dazu $A \mathrel{\hat{=}}$ blaue Augen und $B \mathrel{\hat{=}}$ schwarze Haare. Dann sehen die relativen Häufigkeiten folgendermaßen aus:

| | $A$ | $A^c$ | |
|---|---|---|---|
| $B$ | $\frac{1}{40}$ | $\frac{5}{40}$ | $\frac{6}{40}$ |
| $B^c$ | $\frac{17}{40}$ | $\frac{17}{40}$ | $\frac{34}{40}$ |
| | $\frac{18}{40}$ | $\frac{22}{40}$ | 1 |

Dann ist

$$P(\{\text{Augen blau}\}|\{\text{Haare schwarz}\}) = P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{40}}{\frac{6}{40}} = \frac{1}{6}.$$

**Lemma 1.31:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $B \in \mathcal{A}$ mit $P(B) > 0$. Dann ist $\big(\Omega, \mathcal{A}, P(\cdot|B)\big)$ wieder ein Wahrscheinlichkeitsraum mit

$$P(\cdot|B) : \mathcal{A} \to [0,1], \qquad A \mapsto P(A|B) := \frac{P(AB)}{P(B)}.$$

Beweis: Sei $Q := P(\cdot|B)$. Dann ist $Q(\emptyset) = 0$ und $Q(A) \geq 0$ für alle $A \in \mathcal{A}$. Sei weiter $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ paarweise disjunkt. Dann gilt

$$Q\Big(\sum_{n\in\mathbb{N}} A_n\Big) = \frac{P\big(\sum\limits_{n\in\mathbb{N}} A_n B\big)}{P(B)} = \sum_{n\in\mathbb{N}} \frac{P(A_n B)}{P(B)} = \sum_{n\in\mathbb{N}} P(A|B) = \sum_{n\in\mathbb{N}} Q(A_n) \quad \Rightarrow \quad (m3)$$

Es folgt, daß $Q$ ein Wahrscheinlichkeitsmaß über $(\Omega, \mathcal{A})$ ist.

**Satz 1.32 (Multiplikationssatz für bedingte Ereignisse):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $A_1, \ldots, A_n \in \mathcal{A}$ mit $P(\bigcap\limits_{i=1}^{n} A_i) > 0$. Dann gilt

$$P\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{j=1}^{n} P\Big(A_j|\bigcap_{i=1}^{j-1} A_i\Big) \qquad \text{mit } \bigcap_{i=1}^{0} A_i := \Omega$$

Beweis per Induktion: Sei $n = 1$:

$$P\Big(A_1|\bigcap_{i=1}^{0} A_i\Big) = P(A_1|\Omega) = \frac{P(A_1\Omega)}{P(\Omega)} = P(A_1)$$

Und für $n = 2$ mit $0 < P(A_1A_2) \leq P(A_1)$ folgt

$$P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)}$$

und durch Multiplikation mit $P(A_1)$ und Anwendung der Definition folgt die Behauptung.

Induktionsschritt $n \mapsto n+1$:

$$P\Big(\bigcap_{j=1}^{n+1} A_j\Big) = P\Big(\bigcap_{j=1}^{n} A_j \cap A_{n+1}\Big) \overset{(n=2)}{=} P\Big(\bigcap_{j=1}^{n} A_j\Big) \cdot P\Big(A_{n+1}|\bigcap_{j=1}^{n} A_j\Big) \overset{\text{(IV)}}{=} \prod_{j=1}^{n+1} P\Big(A_j|\bigcap_{i=1}^{n+1} A_i\Big)$$

**Beispiel 1.33:**

(a) Es sei ein Urne mit $r \geq 1$ roten und $s \geq 1$ schwarzen Kugeln gegeben. Wir ziehen zweimal ohne Zurücklegen. Wie groß ist die Wahrscheinlichkeit, daß beide gezogene Kugeln schwarz sind? Wir gehen das Problem heuristisch, d.h. ohne Spezifierung des Modells, an:

$$A_i \mathrel{\hat{=}} i\text{-te gezogenen Kugel ist schwarz}$$

Es ist somit $A = A_1A_2$ und

$$P(A) = P(A_1)P(A_2|A_1) = \frac{s}{r+s} \cdot \frac{s-1}{r+s-1}$$

(b) Reconte-Problem (vgl. Beispiel 1.19):

$$\begin{aligned} B_k &:= \{\text{genau } k \text{ Briefe im richtigen Umschlag}\} \\ A_i &:= \{i\text{-ter Brief im richtigen Umschlag}\} \end{aligned}$$

Dann ist

$$P_n(B_0) = \sum_{l=0}^{n} (-1)^l \frac{1}{l!}, \qquad P_n(A_{i_1} \cdots A_{i_k}) = \frac{(n-k)!}{n!}$$

Eine Lösungsmethode für $P(B_k)$ ist mit Hilfe des Multiplikationssatzes:

$$\begin{aligned} P_n(B_k) &= \sum_{i \le i_1 < \ldots < i_k \le n} P_n(A_{1_1} \cdots A_{i_k} A^c_{i_{k+1}} \cdots A^c_{i_n}) = \\ &= \binom{n}{k} P_n(A_1 \cdots A_k) \underbrace{P_n(A^c_{k+1} \cdots A^c_n | A_1 \cdots A_k)}_{P_{n-k}(C)} = \\ &= \binom{n}{k} \frac{(n-k)!}{n!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!} = \frac{1}{k!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!} \end{aligned}$$

mit $C = \{\text{kein Brief unter } n-k \text{ landet im richtigen Umschlag}\}$.

**Satz 1.34 (Satz von der totalen Wahrscheinlichkeit/totalen Zerlegung):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $A \in \mathcal{A}$, $(B_n)_{n \in \mathbb{N}} \subset \mathcal{A}$ mit $(B_n)$ paarweise diskunkt sowie $A \subset \sum_{n \in \mathbb{N}} B_n$. Dann gilt:

$$P(A) = \sum_{n \in \mathbb{N}:\, P(B_n) > 0} P(A|B_n) P(B_n).$$

Beweis: Es ist $A \subset \sum_{n \in \mathbb{N}} B_n = A = \sum_{n \in \mathbb{N}} AB_n$. Und es folgt

$$P(A) = P(\sum_{n \in \mathbb{N}} AB_n) = \sum_{n \in \mathbb{N}} P(AB_n) = \sum_{n \in \mathbb{N}:\, P(B_n) > 0} P(A|B_n) P(B_n)$$

**Satz 1.35 (Bayes'sche Formel):**
Seien die Voraussetzungen wie im vorherigen Satz. Sei zusätzlich $P(A) > 0$. Dann gilt:

$$P(B_k|A) = \frac{P(A|B_k) P(B_k)}{\sum_{n \in \mathbb{N}:\, P(B_n) > 0} P(A|B_n) P(B_n)} \qquad \forall\, k \in \mathbb{N}.$$

Häufig wir die Bayes'sche Formel nur für $\sum_{n \in \mathbb{N}} B_n = \Omega$ formuliert.

Beweis:

$$P(B_k|A) = \frac{P(B_k A)}{P(A)} \overset{(*)}{=} \frac{P(AB_k) P(B_k)}{\sum_{n \in \mathbb{N}:\, P(B_n) > 0} P(AB_n) P(B_n)}$$

$(*)$ = Definition von $P(\cdot|\cdot)$ und Satz 1.34.

**Beispiel 1.36:**
Eine Urne enthält mit einer Wahrscheinlichkeit von $\frac{1}{2^n}$, $n \in \mathbb{N}$, $2^n - 1$ rote und eine schwarze Kugel. Die schwarze Kugel ist gezogen worden. Wie groß ist die Wahrscheinlichkeit, daß die Urne $k$ rote Kugeln enthält?

Ansatz: $\Omega = \{2^n - 1 : n \in \mathbb{N}\} \times \{S, R\}$. Dann sei

$$B_n \hat{=} \{\text{Urne enthält } 2^n - 1 \text{ rote Kugeln}\} = \big\{(2^n - 1, \omega_2) : \omega_2 \in \{S, R\}\big\}$$

Es ist $P(B_n) = \dfrac{1}{2^n}$. Wir betrachten $A \hat{=} \{\text{schwarze Kugel gezogen}\}$. Dann ist $P(A|B_n) = \dfrac{1}{2^n}$.

$$\Rightarrow \quad P(B_n A) = P\big(\{(2^n - 1, S)\}\big) = P(A|B_n)P(B_n) = \frac{1}{2^{2n}}$$

Zur Kontrolle muß $P(\Omega) = 1$ sein:

$$P(\Omega) = \sum_{n \in \mathbb{N}} \big[P(B_n A) + P(B_n A^c)\big] = \sum_{n \in \mathbb{N}} \Big[\frac{1}{2^{2n}} + \frac{2^n - 1}{2^{2n}}\Big] = 1$$

Also ist $P$ ein Wahrscheinlichkeitsmaß über $(\Omega, 2^\Omega)$. Es ist

$$\begin{aligned} P(B_n|A) &= P(\{\text{Urne enthält } 2^n - 1 \text{ rote} \mid \text{schwarze Kugel gezogen}\}) = \\ &= \frac{P(A|B_n)P(B_n)}{\sum\limits_{k \in \mathbb{N}} P(A|B_k)P(B_k)} = \frac{\dfrac{1}{2^{2n}}}{\sum\limits_{k \in \mathbb{N}} \dfrac{1}{2^{2k}}} = \frac{3}{2^{2n}} \end{aligned}$$

So ist zum Beispiel $P(B_1|A) = \frac{3}{4}$, $P(B_2|A) = \frac{3}{16}$ und $P(B_3|A) = \frac{3}{64}$. Es heißt $P(B_n)$ die a-priori Wahrscheinlichkeit (Vorinformation). Dann heißt $P(B_n|A)$ die a-posteriori Wahrscheinlichkeit, d.h. das ist der Kenntnisstand über die Urne nach einmaligem Ziehen mit Ergebnis „schwarze Kugel gezogen“.

## 1.5 Stochastische Unabhängigkeit von Ereignissen

Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit $A, B \in \mathcal{A}$. Weiß man schon, daß $A$ eingetreten ist, so kann $A$ das Eintreten von $B$ beeinflussen, d.h. begünstigen oder benachteiligen (stochastige Abhängigkeit) oder aber auch nicht (stochastische Unabhängigkeit).

**Beispiel 1.37:**
Es seien 16 schwarze und 16 weiße Schachfiguren gegeben (vgl. Beispiel 1.33(a)).

(a) Zweimal Ziehen ohne Zurücklegen.

$$A_i \hat{=} \{i\text{-te gezogene Figur ist schwarz}\}, \qquad i = 1, 2.$$

Es ist $P(A_1) = P(A_2) = \frac{1}{2}$. Dann ergibt sich

$$P(A_2|A_1) = \frac{15}{31} < P(A_2), \qquad P(A_2|A_1^c) = \frac{16}{31} > P(A_2).$$

(b) Zweimal Ziehen mit Zurücklegen.

$$P(A_2|A_1) = \frac{1}{2} = P(A_2), \qquad P(A_2|A_1^c) = \frac{1}{2} = P(A_2).$$

Es ist

$$P(A_2|A_1) = \frac{P(A_2 A_1)}{P(A_1)} \quad \Rightarrow \quad P(A_1 A_2) = P(A_1)P(A_2).$$

**Definition 1.38:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $A, B \in \mathcal{A}$. Dann heißen $A$ und $B$ stochastisch unabhängig (unter $P$) genau dann, wenn

$$P(AB) = P(A)P(B).$$

**Beispiel 1.39:**

(i) Falls $P(B) > 0$, dann gilt:

$$P(AB) = P(A)P(B) \quad \Longleftrightarrow \quad P(A|B) = P(A).$$

Entsprechendes gilt für $P(A) > 0$.

(ii) Ist $P(B) \in \{0, 1\}$, so folgt $P(AB) = P(A)P(B)$, denn

$$P(B) = 0 \quad \Rightarrow \quad 0 \leq P(AB) \leq P(B) = 0 = P(A)P(B)$$
$$P(B) = 1 \quad \Rightarrow \quad P(B^c) = 0 \Rightarrow P(AB^c) = 0 \Rightarrow P(AB) = P(AB) + P(AB^c) = P(A) = P(A)P(B)$$

(iii) Sind $A$ und $B$ stochastisch unabhängig, so sind auch $A, B^c$; $A^c, B$ und $A^c, B^c$ stochastisch unabhängig, denn

$$P(AB^c) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)\big(1 - P(B)\big) = P(A)P(B^c)$$

Die anderen Fälle sind damit auch gezeigt.

Eine Frage ist jetzt, wie man für endlich (abzählbar) viele Ereignisse die stochastische Unabhängigkeit definiert.

**Beispiel 1.40:**
Wir ziehen einmal aus einer Urne mit den Kugeln mit Nummern $1, \ldots, 4$. Wir setzen $\Omega = \{1, \ldots, 4\}$ und $P = G(\Omega)$. Seien $A = \{1, 2\}$, $B = \{1, 3\}$ und $C = \{1, 4\}$. Wie sofort ersichtlich, sind $A, B$ und $C$ paarweise stochastisch unabhängig, d.h.

$$P(AB) = P(A)P(B) = \frac{1}{4}, \qquad \text{(etc.)}$$

Aber es ist

$$P(C|AB) = \frac{P(ABC)}{P(AB)} = \frac{1}{4}\left(\frac{1}{4}\right)^{-1} = 1 > \frac{1}{2} = P(C),$$

d.h. $C$ ist nicht stochastisch unabhängig von $A$ und $B$, d.h. $A \cap B$:

$$\frac{1}{4} = P(ABC) \neq P(A)P(B)P(C) = \frac{1}{8}$$

**Definition 1.41 (Stochastische Unabhängigkeit von endlich (abzählbar) vielen Ereignissen):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Weiter seien $A_1, \ldots, A_n \in \mathcal{A}$ mit $(A_i)_{i \in I} \in \mathcal{A}$, wobei $I$ eine abzählbare Indexmenge ist.

(a) $A_1, \ldots, A_n$ heißen stochastisch unabhängig (unter $P$) genau dann, wenn

$$\forall\, J \subseteq \{1, \ldots, n\},\ |J| \geq 2: \qquad P\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} P(A_j)$$

(b) $A_i,\ i \in I$, heißen stochastisch unabhängig (unter $P$) genau dann, wenn

$$\forall\, J \subseteq I,\ 2 \leq |J| < \infty: \qquad P\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} P(A_j)$$

**Beispiel 1.42:**

(a) Eine abzählbare Familie von Ereignissen ist genau dann stochastisch unabhängig, wenn jede endliche Teilfamilie stochastisch unabhängig ist.

(b) Sind $(A_i)_{i \in I}$ stochastisch unabhängig, dann sind $(A_i)_{i \in I}$ paarweise stochastisch unabhängig.

**Lemma 1.43:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $A_i \in \mathcal{A}$, $D_i \in \{A_i, A_i^c\}$ für $i = 1, \ldots, n$. Dann sind äquivalent:

(a) $A_1, \ldots, A_n$ stochastisch unabhängig,

(b) $D_1, \ldots, D_n$ stochastisch unabhängig,

(c) Für alle $(E_1, \ldots, E_n) \in \times_{i=1}^n \{A_i, A_i^c\}$ gilt:

$$P\Big(\bigcap_{i=1}^n E_i\Big) = \prod_{i=1}^n P(E_i)$$

**Lemma 1.44:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Es seien $A_1, \ldots, A_n \in \mathcal{A}$ stochastisch unabhängig. Dann gilt:

$$P\Big(\bigcup_{i=1}^n A_i\Big) = 1 - \prod_{i=1}^n \big(1 - P(A_i)\big).$$

Beweis:

$$P\Big(\bigcup_{i=1}^n A_i\Big) = 1 - P\Big(\bigcap_{i=1}^n A_i^c\Big) = 1 - \prod_{i=1}^n P(A_i^c) = 1 - \prod_{i=1}^n \big(1 - P(A_i)\big)$$

**Beispiel 1.45 (n-facher Münzwurf, Binominalverteilung):**
Eine Münze wird $n$-mal geworfen, das Ergebnis eines Wurfes ist 1 bzw. 0 mit der Wahrscheinlichkeit $p$ bzw. $1-p$. Wir setzen an: $\Omega = \{0,1\}^n$, $\mathcal{A} = 2^\Omega$, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$. Es sei $A_i = \{\omega \in \Omega : \ \omega_i = 1\}$. Also ist $P(A_i) = p$ und $P(A_i^c) = 1 - p$. Sei $\omega \in \Omega$ mit $\omega_{i_1} = \ldots = \omega_{i_k} = 1$ und $\omega_{j_1} = \ldots = \omega_{j_{n-k}} = 0$ mit $\{i_1, \ldots, i_k\} + \{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\}$.

Ansatz:

$$P(\{\omega\}) = P(A_{i_1} \cdots A_{i_k} A_{j_1}^c \cdots A_{j_{n-k}}^c) := \prod_{r=1}^k P(A_{i_r}) \prod_{s=1}^{n-k} P(A_{j_s}^c) = p^k (1-p)^{n-k}$$

d.h. es ist unabhängig von der Reihenfolge der Nullen und Einsen. Es gibt $\binom{n}{k}$ $\omega \in \Omega$ mit $k$ Einsen und $(n-k)$ Nullen. Wir setzen

$$B_k \mathrel{\hat{=}} \{k \text{ mal die 1, } (n-k) \text{ mal die 0}\}$$

Somit folgt:

$$P(B_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Zur Sicherheit nochmal eine Kontrollrechnung:

$$P(\Omega) = P\Big(\sum_{k=0}^n B_k\Big) = \sum_{k=0}^n P(B_k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = 1.$$

$P$ heißt Binominalverteilung mit Parameter $n$ und $p$, kurz $P = B(n,p)$, mit $n \in \mathbb{N}$ und $p \in [0,1]$.

**Anwendung:**
Eine Maschine produziert Schrauben. Die Ausschußquote der Maschine sei 10 Prozent (Erfahrungswert, empirische Wahrscheinlichkeit). Gesucht ist die Wahrscheinlichkeit, daß unter drei willkürlich gezogenen Schrauben höchstens eine defekt ist.

$$\begin{aligned} A_i &\mathrel{\hat{=}} i\text{-te Schraube ist defekt}, \qquad P(A_i) = \frac{1}{10} \text{ für } i = 1,2,3 \\ B_0 &\mathrel{\hat{=}} \text{keine Schraube ist defekt} \\ B_1 &\mathrel{\hat{=}} \text{genau eine Schraube ist defekt} \end{aligned}$$

$$P(B_0) + P(B_1) = \binom{3}{0} p^0 (1-p)^{3-0} + \binom{3}{1} p^1 (1-p)^{3-1} \approx 97,2\,\%$$

## 1.6 Diskrete Wahrscheinlichkeitsverteilungen

**Definition 1.46:**
$(\Omega, 2^\Omega, P)$ bzw. $(\Omega, P)$ heißt diskreter Wahrscheinlichkeitsraum genau dann, wenn

(i) Es existiert ein $p : \Omega \to [0,1]$ mit

$$\Omega_0 := \{\omega \in \Omega : \; p(\omega) > 0\} \neq 0 \qquad \text{und } |\Omega_0| \leq |\mathbb{N}|,$$

d.h. $\Omega_0$ ist höchstens abzählbar unendlich.

(ii) Es ist definiert

$$\sum_{\omega \in \Omega} p(\omega) := \sum_{\omega \in \Omega_0} p(\omega) = 1$$

(iii) Für alle $E \subseteq \Omega$ gilt

$$P(E) = \sum_{\omega \in E} p(\omega) := \sum_{\omega \in E \Omega_0} p(\omega)$$

$P$ heißt diskretes Wahrscheinlichkeitsmaß (Wahrscheinlichkeitsverteilung, Verteilung) auf $\Omega$. $p$ heißt diskrete Wahrscheinlichkeitsdichte (Zähldichte) von $P$. $\Omega_0$ heißt Träger von $P$.

**Beispiel 1.47:**

(a) Bernoulli-Verteilung: $B(1,p)$, $p \in [0,1]$, $\Omega \supseteq \tilde{\Omega} = \{0,1\} \supseteq \Omega_0$. Es ist $p(0) = 1-p$, $p(1) = p$ und $p(\omega) = 0$ für alle $\omega \in \Omega \setminus \Omega_0$. So sei zum Beispiel das Geschlecht eines Neugeborenen 1 für männlich und 0 für weiblich, dann ist $p \approx 0.513$.

(b) $G(\Omega_0)$, $|\Omega_0| < \infty$, Gleichverteilung auf $\Omega_0$ (vgl. 1.2).

(c) Verteilung der Augensumme von 2 (bzw. $n$) Würfeln.

$$\Omega_0 = \{2, 3, \ldots, 12\}, \qquad p(2) = \frac{1}{36}, \quad p(3) = \frac{2}{36}, \quad \text{(etc.)}$$

(d) Binominalverteilung $B(n,p)$, $n \in \mathbb{N}$, $p \in [0,1]$ (vgl. Bsp. 1.45):

$$\Omega_0 \subseteq \tilde{\Omega} = \{0, 1, \ldots, n\}, \qquad p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, \ldots, n\}$$

Beachte: Für $p = 0$ ist $\Omega_0 = \{0\}$, für $p = 1$ ist $\Omega_0 = \{n\}$.

(e) Geometrische Verteilung: $NB(1,p)$ mit $p \in (0,1]$, $\Omega_0 = \mathbb{N}_0$ und $p(i) = p(1-p)^i$ mit $i \in \mathbb{N}_0$.
Beispiel: Würfle so lange, bis die erste „6“ erscheint. Notiere als Ergebnis die Anzahl der Würfe ohne eine „6“.

$$\omega \mathrel{\hat{=}} (\text{Anzahl der Würfe bis zur ersten sechs}) - 1$$

Annahme: Die Wahrscheinlichkeit für eine „6“ ist $p \in (0,1]$

$$\Rightarrow \quad P(\{i\}) = p(1-p)^i, \qquad i \in \mathbb{N}_0$$

(f) Negative Binominalverteilung $NB(m,p)$ mit $m \in \mathbb{N}$, $p \in (0,1]$ und $\Omega_0 \subseteq \mathbb{N}_0$. Es ist für alle $i \in \mathbb{N}_0$:

$$p(i) = \binom{m-1+i}{m-1} p^m (1-p)^i \qquad \text{(Verallgemeinerung von (e))}$$

Beispiel: Würfel so lange, bis die $m$-te 6 erscheint. Es bezeichne $\omega$ die Anzahl der Würfe, bis die $m$-te 6 erscheint, minus 1. Ferner ist $P(\{i\}) = p(i)$.

(g) Hypergeometrische Verteilung $H(n,m,k)$ mit $m \leq n$, $k \leq n$ und $\Omega_0 \subseteq \tilde{\Omega} = \{0,1,\ldots,n\}$. Dann ist

$$p(i) = \begin{cases} \dfrac{\binom{m}{i}\binom{n-m}{k-i}}{\binom{n}{k}} & \text{falls } \max\{0, k+m-n\} \leq i \leq \min\{m,k\} \\ 0 & \text{sonst} \end{cases}$$

Beispiel (Ziehen ohne Zurücklegen): Eine Urne enthält $n$ Kugeln, davon $m$ rote und $n-m$ schwarze. Ziehe eine Kugel ohne Zurücklegen. Dann ist $p(i)$ die Wahrscheinlichkeit dafür, $i$ rote und $k-i$ schwarze Kugeln zu ziehen.

Beispiel (Qualitätskontrolle): Warensendung mit $n$ Stück, davon $m$ schlechte und $n-m$ gute Stücke.

(h) Multinominalverteilung $M(n, p_1, \ldots, p_k)$ mit $n, k \in \mathbb{N}$, $k \geq 2$, $p_j \in [0,1]$ und $\sum_{j=1}^{k} p_j = 1$. Es sei $\Omega_0 \subseteq \tilde{\Omega} = \{0,1,\ldots,n\}^k$. Dann ist

$$p(i_1,\ldots,i_k) = \begin{cases} \binom{n}{i_1 \cdots i_k} \prod_{j=1}^{k} p_j^{i_j} & \text{falls } \sum_{j=1}^{k} i_j = n \\ 0 & \text{sonst} \end{cases}$$

wobei $\binom{n}{i_1 \cdots i_k} := \dfrac{n!}{\prod_{j=1}^{k} i_j!}$ Multinominalkoeffizient heißt.

Beispiel: $n$-mal Würfeln mit $p(i)$ für die Wahrscheinlichkeit, eine $i$ zu würfeln. Es sei $\sum_{i=1}^{6} p_i = 1$. Dann ist $p(i_1,\ldots,i_6)$ die Wahrscheinlichkeit dafür, daß $i_1$ mal die 1, …, $i_6$ mal die 6 gewürfelt wird.

(i) Poisson-Verteilung $Po(\lambda)$ mit $\lambda \in [0,\infty)$, $\Omega_0 \subseteq \tilde{\Omega} = \mathbb{N}_0$. Dann ist

$$p(i) = e^{-\lambda} \frac{\lambda^2}{i!} \qquad (i \in \mathbb{N}_0) \qquad \text{(Verteilung seltener Ereignisse)}$$

**Satz 1.47 (Poisson'scher Grenzwertsatz, Poisson 1832):**
Seien $P_n = B(n,p_n)$ mit $p_n = \frac{\lambda}{n}$, $\lambda > 0$, $n \in \mathbb{N}$ und sei $P = Po(\lambda)$. Dann gilt für alle $k \in \mathbb{N}_0$:

$$\lim_{n\to\infty} P_n(\{k\}) = P(\{k\}).$$

Beweis: Wegen $p_n \to 0$ folgt, daß $k \in \mathbb{N}$ ist.

$$\begin{aligned} \lim_{n\to\infty} P_n(\{k\}) &= \lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!} \lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^{n-k} \prod_{j=1}^{k} \lim_{n\to\infty} (n-k+j)p_n = \\ &= \frac{1}{k!} e^{-\lambda} \lambda^k = P(\{k\}) \end{aligned}$$

# 2 Zufallsvariable und Integrationstheorie

## 2.1 Zufallsvariable und stochastische Unabhängigkeit

Ereignisse im Zufallsexperiment lassen sich oft einfacher darstellen, wenn man Funktionen zu Hilfe nimmt. Betrachte zum Beispiel den $n$-fachen Münzwurf: $\Omega = \{0,1\}^n$, $p(\omega) = \frac{1}{2^n}$. Es sei

$$A_k \hat{=} \{\text{mindestens } k \text{ mal die } 1\} = \Big\{\omega \in \Omega : \underbrace{\sum_{i=1}^{n} \omega_i}_{=:X(\omega)} \geq k\Big\}$$

mit $X : \Omega \to \{0, 1, \ldots, n\}$. Schon beim Kodieren wurden Funktionen eingesetzt, zum Beispiel $\{W, Z\} \to \{0, 1\}$.

**Definition 2.1:**
Ist $(\Omega, 2^\Omega, P)$ ein diskreter Wahrscheinlichkeitsraum, so heißt jede Abbildung $X : \Omega \to \mathcal{X}$ eine $\mathcal{X}$-wertige Zufallsvariable auf $(\Omega, 2^\Omega, P)$. Dabei ist $\mathcal{X}$ eine Menge mit $\mathcal{X} \neq \emptyset$. Jedes Ereignis der Gestalt

$$\{\omega \in \Omega : \ X(\omega) \in B\} =: X^{-1}(B) =: \{X \in B\}$$

mit $B \in 2^\mathcal{X}$ heißt ein durch $X$-beschriebenes Ereignis. Wahrscheinlichkeiten schreibt man kurz

$$P(\{\omega \in \Omega : \ X(\omega) \in B\}) = P\big(X^{-1}(B)\big) =: P(X \in B)$$

**Bemerkung 2.2:**
Seien $\Omega, \mathcal{X} \neq \emptyset$ und $X : \Omega \to \mathcal{X}$. Dann heißen $X(A) := \{X(\omega) : \ \omega \in A\}$ Bild von $A \subseteq \Omega$ und $X^{-1}(B) := \{\omega \in \Omega : \ X(\omega) \in B\}$ Urbild von $B \subseteq \mathcal{X}$. Die zugehörige Abbildungen $X : 2^\Omega \to 2^\mathcal{X}$ und $X^{-1} : 2^\mathcal{X} \to 2^\Omega$ heißen Fortsetzung von $X|_\Omega$ zu $X|_{2^\Omega}$ bzw. Umkehrabbildung (Urbildabbildung, Urbildfunktion). Es gilt

(i) Für alle $B \subseteq \mathcal{X}$ gilt $X\big(X^{-1}(B)\big) \subseteq B$. Gleichheit gilt genau dann, wenn $X$ surjektiv ist.

(ii) Für alle $A \subseteq \Omega$ gilt $X^{-1}\big(X(A)\big) \supseteq A$. Gleichheit gilt genau dann, wenn $X|_\Omega$ injektiv ist.

WARNUNG: Verwechsle niemals $X^{-1} : 2^\mathcal{X} \to 2^\Omega$ mit der inversen Abbildung $f^{-1} : \mathcal{X} \to \Omega$ für bijektive Abbildungen $f : \Omega \to \mathcal{X}$.

**Lemma 2.3 (Operationstreue der Umkehrabbildung):**
Seien $\Omega, \mathcal{X} \neq \emptyset$, $X : \Omega \to \mathcal{X}$ und $B_i \subseteq \mathcal{X}$ für $i \in I$ (beliebige Indexmenge). Dann gilt

(a) $X^{-1}\Big(\bigcup\limits_{i\in I} B_i\Big) = \bigcup\limits_{i\in I} X^{-1}(B_i)$

(b) $X^{-1}(B_i^c) = \big(X^{-1}(B_i)\big)^c$

(c) $X^{-1}(\emptyset) = \emptyset$ und $X^{-1}(\mathcal{X}) = \Omega$

(d) Ist $B_i \subseteq B_j$, so folgt $X^{-1}(B_i) \subseteq X^{-1}(B_j)$

(e) $X^{-1}\Big(\bigcap\limits_{i\in I} B_i\Big) = \bigcap\limits_{i\in I} X^{-1}(B_i)$

(f) $X^{-1}\Big(\sum\limits_{i\in I} B_i\Big) = \sum\limits_{i\in I} X^{-1}(B_i)$, falls $B_i$ paarweise disjunkt sind.

**Lemma 2.4 (Urbildabbildung einer Komposition):**
Seien die Voraussetzungen des Lemmas 2.3 erfüllt sowie $T : \mathcal{X} \to \mathcal{T}$. Dann gilt

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

**Lemma 2.5:**
Sei $X$ eine $\mathcal{X}$-wertige Zufallsvariable auf einem diskreten Wahrscheinlichkeitsraum $(\Omega, 2^{\Omega}, P)$. Dann ist $P^X$ diskretes Wahrscheinlichkeitsmaß auf $(\mathcal{X}, 2^{\mathcal{X}})$, wobei für alle $B \in 2^{\mathcal{X}}$ gilt

$$P^X(B) := P\big(X^{-1}(B)\big).$$

**Bemerkung:**
Ist $(\Omega, \mathcal{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum, also nicht notwendig diskret, so läßt sich $P\big(X^{-1}(B)\big)$ nur dann berechnen, wenn $X^{-1}(B) \in \mathcal{A}$ ist, da $P(\cdot)$ nur auf $\mathcal{A}$ definiert ist. Dies führt zum Begriff der meßbaren Abbildungen.

**Definition 2.6:**
Gegeben seien die Maßräume $(\Omega, \mathcal{A})$ und $(\mathcal{X}, \mathcal{B})$.

(a) Jede Funktion $f : \Omega \to \mathcal{X}$ heißt $\mathcal{A}$-$\mathcal{B}$-meßbar (kurz: meßbar) genau dann, wenn für alle $B \in \mathcal{B}$ gilt:

$$f^{-1}(B) \in \mathcal{A}.$$

Kurz: $f : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B})$.

(b) Ist zusätzlich $P$ ein Wahrscheinlichkeitsmaß über $(\Omega, \mathcal{A})$, so heißt dann jede meßbare Abbildung $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B})$ eine ($\mathcal{X}$-wertige) Zufallsvariable.

(c) Seien die Voraussetzungen von (b) erfüllt. Es sei weiter $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B})$ gegeben. Dann heißt $P^X : \mathcal{B} \to [0, 1]$, definiert durch

$$\forall\, B \in \mathcal{B} \qquad P^X(B) := P\big(X^{-1}(B)\big),$$

Bildmaß oder Verteilung von $X$.

**Lemma 2.7:**
Mit den Bezeichnungen von Definition 2.6 ist $(\mathcal{X}, \mathcal{B}, P^X)$ ein Wahrscheinlichkeitsraum.

**Bemerkung:**
Ist $\mathcal{X}$ Teilmenge von ..., so heißt $X$ ... Zufallsvariable.

$$\mathcal{X} \subseteq \begin{cases} \mathbb{Z} \\ \mathbb{R} \\ \overline{\mathbb{R}} := [-\infty, +\infty] \\ \mathbb{R}^k \\ I \text{ abz.} \end{cases} \qquad \longrightarrow \qquad X \begin{cases} \text{ganzzahlige} \\ \text{reelle} \\ \text{numerische} \\ \text{vektorwertige, } k\text{-dimensionale} \\ \text{diskrete} \end{cases}$$

**Beispiel 2.8:**

(a) Wir würfeln zweimal. Angenommen $\Omega = \{1, \ldots, 6\}^n$, $P = \mathrm{G}(\Omega)$. Wir setzen $X(\omega) = \omega_1 + \omega_2$. Dann ist $\mathcal{X} = \{2, \ldots, 12\}$, $X : \Omega \to \mathcal{X}$ und

$$P^X(\{x\}) = \begin{cases} \frac{i}{36} & \text{falls } i \in \{2, \ldots, 7\} \\ \frac{13-i}{36} & \text{falls } i \in \{8, \ldots, 12\} \end{cases}$$

(b) (Indikatorfunktion) Sei $(\Omega, \mathcal{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum, $A \in \mathcal{A}$ und $X := I_A$, also ist $X : (\Omega, \mathcal{A}) \to (\{0,1\}, 2^{\{0,1\}})$ und

$$P^X = \mathrm{B}(1,p) \qquad \text{mit} \quad p := P(A) = P\big(X^{-1}(\{1\})\big).$$

**Bemerkung:**
Es ist $I_A$ $\mathcal{A}$-$2^{\{0,1\}}$-meßbar genau dann, wenn $A \in \mathcal{A}$. $I_A$ ist $\sigma(\{A\})$-$2^{\{0,1\}}$-meßbar. Dabei ist $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.

**Bemerkung 2.9 (Meßbarkeit):**

(a) Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ Meßräume und $\mathcal{E} \subseteq 2^{\Omega'}$ mit $\mathcal{A}' = \sigma(\mathcal{E}')$. Dann ist $f$ genau dann $\mathcal{A}$-$\mathcal{A}'$-meßbar, wenn für alle $A' \in \mathcal{E}'$ gilt: $f^{-1}(A') \in \mathcal{A}$.

(b) Seien $f_1 : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ und $f_2 : (\Omega_2, \mathcal{A}_2) \to (\Omega_3, \mathcal{A}_3)$. Dann ist $f_2 \circ f_1$ $\mathcal{A}_1$-$\mathcal{A}_3$-meßbar.

(c) Jede stetige Abbildung $f : \mathbb{R}^p \to \mathbb{R}^q$ ist $\mathbb{B}^p$-$\mathbb{B}^q$-meßbar (kurz: Borel-meßbar).

(d) Ist $(\Omega, \mathcal{A})$ ein Meßraum, $f : \Omega \to \overline{\mathbb{R}}$, und sei $\overline{\mathbb{B}} = \sigma\big(\mathbb{B} \cup \big\{\{-\infty\}, \{+\infty\}\big\}\big)$. Dann ist die $\mathcal{A}$-$\overline{\mathbb{B}}$-Meßbarkeit von $f$ äquivalent mit jeder der Bedingungen

(i) Für alle $\alpha \in \mathbb{R}$ ist $\{f \geq \alpha\} \subseteq \mathcal{A}$. Dabei ist $\{f \geq \alpha\} := \{\omega : f(\omega) \geq \alpha\}$.

(ii) Für alle $\alpha \in \mathbb{R}$ ist $\{f > \alpha\} \subseteq \mathcal{A}$.

(iii) Für alle $\alpha \in \mathbb{R}$ ist $\{f \leq \alpha\} \subseteq \mathcal{A}$.

(iv) Für alle $\alpha \in \mathbb{R}$ ist $\{f < \alpha\} \subseteq \mathcal{A}$.

(e) Seien $f, g : \Omega \to \overline{\mathbb{R}}$ $\mathcal{A}$-$\overline{\mathbb{B}}$-meßbar, dann gilt

(i) $\{f < g\},\ \{f \leq g\},\ \{f > g\},\ \{f \geq g\} \in \mathcal{A}$.

(ii) $f \cdot g$ und $f + g$ sind $\mathcal{A}$-$\overline{\mathbb{B}}$-meßbar.

(f) Ist $f_n : \Omega \to \overline{\mathbb{R}}$ $\mathcal{A}$-$\overline{\mathbb{B}}$-meßbar für alle $n \in \mathbb{N}$. Dann sind auch

$$\sup_{n\to\infty} f_n, \quad \inf_{n\to\infty} f_n, \quad \lim_{n\to\infty} \sup f_n \quad \text{und} \quad \lim_{n\to\infty} \inf f_n \quad \mathcal{A}\text{-}\overline{\mathbb{B}}\text{-meßbar.}$$

(g) Gegeben seien die Meßräume $(\Omega_i, \mathcal{A}_i)$, $i \in I$, und $f_i : \Omega \to \Omega_i$. Dann gilt:

$$\mathcal{A} := \mathcal{A}(f_i : i \in I) := \sigma\Big(\bigcup_{i\in I} f_i^{-1}(\mathcal{A}_i)\Big)$$

ist die kleinste $\sigma$-Algebra, für die jedes $f_i$ $\mathcal{A}$-$\mathcal{A}_i$-meßbar ist. Es heißt $\mathcal{A}(f_i : i \in I)$ die von $f_i$, $i \in I$, erzeugte $\sigma$-Algebra. Dabei ist

$$f_i^{-1}(\mathcal{A}_i) = \{f_i^{-1}(B) : B \in \mathcal{A}\}.$$

Um stochastische Unabhängigkeit von Zufallsvariablen definieren zu können, sind noch einige Vorbereitungen notwendig.

**Definition 2.10 (Stochastische Unabhängigkeit von Mengensystemen):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, eine beliebige Indexmenge $I$ und weiter $\mathcal{E}_i \subseteq \mathcal{A}$ für $i \in I$. Dann heißen die Mengensysteme $(\mathcal{E}_i)_{i\in I}$ stochastisch unabhängig, falls für alle endlichen $I_0 \subseteq I$ gilt:

$$\forall\, i \in I_0 : \forall\, A_i \in \mathcal{E}_i : \qquad P\Big(\bigcap_{i\in I_0} A_i\Big) = \prod_{i\in I_0} P(A_i).$$

**Bemerkung 2.11:**
Definition 2.10 ist eine Verallgemeinerung von Definition 1.41. Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $A_1, \ldots, A_n \in \mathcal{A}$. Dann gilt:

$$\begin{aligned} &A_1, \ldots, A_n \text{ stoch. unabh. im Sinne von Def. 1.41 (a)} \iff \\ &\iff \quad \{A_1\}, \ldots, \{A_n\} \text{ stoch. unabh. im Sinne von Def. 2.10} \\ &\overset{\text{Ü}}{\iff} \quad \{\emptyset, A_1, A_1^c, \Omega\}, \ldots, \{\emptyset, A_n, A_n^c, \Omega\} \text{ stoch. unabh. im Sinne von Def. 2.10} \end{aligned}$$

**Definition 2.12 (von $X$ erzeugte $\sigma$-Algebra):**
Gegeben seien die Meßräume $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ sowie $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathbb{B})$. Dann heißt

$$\sigma(X) := X^{-1}(\mathbb{B}) := \{X^{-1}(B) : B \in \mathbb{B}\}$$

die von $X$ erzeugte $\sigma$-Algebra oder Urbild-$\sigma$-Algebra.

**Bemerkung:**

(a) $\sigma(X)$ ist eine $\sigma$-Algebra über $\Omega$:

($\sigma$1) $X^{-1}(\mathcal{X}) = \Omega \in \sigma(X)$

($\sigma$2) Sei $B \in \mathbb{B}$ mit $X^{-1}(B) \in \sigma(X) \quad \Rightarrow \quad \Omega - X^{-1}(B) = X^{-1}(\underbrace{\Omega - B}_{\in \mathbb{B}}) \in \sigma(X)$

($\sigma$3) Sei $(B_n)_{n \in \mathbb{N}} \subset B$. Dann folgt

$$\bigcup_{n \in \mathbb{N}} \underbrace{X^{-1}(B_n)}_{\in \sigma(X)} = X^{-1}\Big( \underbrace{\bigcup_{n \in \mathbb{N}} B_n}_{\in \mathbb{B}} \Big) \in \sigma(X)$$

(b) Es gilt $\sigma(X) \subseteq \mathcal{A}$, da $X$ $\mathcal{A}$-$\mathcal{B}$-meßbar ist nach Voraussetzung.

**Definition 2.14:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $I$ eine beliebige Indexmenge und $X_i : (\Omega, \mathcal{A}) \to (\mathcal{X}_i, \mathbb{B}_i)$ für $i \in I$. Dann heißt die Familie $(X_i)_{i \in I}$ von Zufallsvariablen stochastisch unabhängig, falls die Familie der von ihnen erzeugten $\sigma$-Algebren $\big(\sigma(X_i)\big)_{i \in I}$ stochastisch unabhängig ist.

**Beispiel 2.15:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $A_1, \ldots, A_n \in \mathcal{A}$ und $X_i = I_{A_i}$. Dann gilt:

$$\begin{aligned} X_1, \ldots, X_n \text{ stochastisch unabhängig} \quad &\iff \quad \{\emptyset, A_1, A_1^c, \Omega\}, \ldots, \{\emptyset, A_n, A_n^c, \Omega\} \text{ stochastisch unabhängig} \\ &\iff \quad A_1, \ldots, A_n \text{ stochastisch unabhängig} \end{aligned}$$

**Bemerkung 2.16:**
Unter den Voraussetzungen von Definition 2.14 sind die $X_i$, $i \in I$, also genau dann stochastisch unabhängig, falls für alle endlichen $I_0 \subseteq I$ gilt:

$$\forall\, i \in I_0,\ \forall\, B_i \in \mathbb{B}_i : \qquad P\Big( \bigcap_{i \in I_0} X_i^{-1}(B_i) \Big) = \prod_{i \in I_0} P\big(X_i^{-1}(B_i)\big). \tag{1}$$

bzw. äquivalent

$$P(X_i \in B_i,\ i = 1, \ldots, n) = \prod_{i \in I_0} P(X_i \in B_i). \tag{2}$$

Der folgende Satz zeigt, daß (1) bzw. (2) nicht für alle $B_i \in \mathbb{B}_i$ überprüft werden muß.

**Satz 2.17:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \to (\mathcal{X}_i, \mathbb{B}_i)$ für alle $i = 1, \ldots, n$. Weiter sei $\mathcal{D}_i$ ein durchschnittsstabiler Erzeuger von $\mathbb{B}_i$, $i = 1, \ldots, n$ (d.h. $\mathbb{B}_i = \sigma(\mathcal{D}_i)$ und mit $D_r, D_s \in \mathcal{D}_i$ ist auch $D_r D_s \in \mathcal{D}_i$) mit $\Omega_i \in \mathcal{D}_i$, $i = 1, \ldots, n$. Dann sind äquivalent

(a) $X_1, \ldots, X_n$ stochastisch unabhängig

(b) Für alle $i = 1, \ldots, n$ und alle $D_i \in \mathcal{D}_i$ gilt:

$$P\Big(\bigcap_{i=1}^{n} X_i^{-1}(D_i)\Big) = \prod_{i=1}^{n} P\big(X_i^{-1}(D_i)\big).$$

**Bemerkung 2.18:**
Wird ein statistisches Experiment $n$ mal wiederholt, so geht man häufig davon aus, daß sich die Ergebnisse bzw. Wiederholungen nicht gegenseitig beeinflussen (Beispiel $n$-facher Münzwurf). Man nimmt dann an, daß die Meßwerte Realisierungen von stochastisch unabhängigen Zufallsvariablen $X_1, \ldots, X_n$ sind, die einem bestimmten Verteilungsgesetzt gehorchen. Beim $n$-fachen Münzwurf wäre zum Beispiel das Ergebnis des $i$-tes Wurfes die Realisation $X_i(\omega) = \{0, 1\}$ einer Zufallsvariable $X_i$ mit

$$P^{X_i} = \mathrm{B}(1, p)$$

(kurz $X_i \sim \mathrm{B}(1, p)$). Die Annahme der stochastischen Unabhängigkeit der $X_i$ führt zu

$$\begin{aligned} P(X_j = i_j,\ j = 1, \ldots, n) &:= P\Big(\bigcap_{j=1}^{n} \{\omega :\ X_j(\omega) = j\}\Big) = \prod_{j=1}^{n} P\big(\{\omega :\ X_j(\omega) = i_j\}\big) = \\ &= \prod_{j=1}^{n} P^{X_j}(\{i_j\}) = p^{\sum_{j=1}^{n} i_j} (1-p)^{n - \sum_{j=1}^{n} i_j} \end{aligned}$$

für $(i_1, \ldots, i_n) \in \{0, 1\}^n$. Dabei wird der ursprüngliche Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ nicht weiter spezifiziert. Man fordert einfach, daß die $X_i$ stochastisch unabhängig und $\mathcal{A}$-$2^{\{0,1\}}$-meßbar sind, d.h.

$$X_i : (\Omega, \mathcal{A}) \to (\{0, 1\}, 2^{\{0,1\}}).$$

Ist man zum Beispiel an der Anzahl der geworfenen Einsen interessiert, so definiert man sich eine neue Zufallsvariable $Y = \sum_{i=1}^{n} X_i$. Man erhält für $k = 0, 1, \ldots, n$:

$$\begin{aligned} P(Y = k) &= P\Big(\sum_{i=1}^{n} X_i = k\Big) = P\Big(\sum_{\substack{(j_1, \ldots, j_n) \in \{0,1\}^n \\ j_1 + \ldots + j_n = 1}} \{X_i = j_i :\ i = 1, \ldots, n\}\Big) = \\ &= \sum_{\substack{(j_1, \ldots, j_n) \in \{0,1\}^n \\ j_1 + \ldots + j_n = 1}} P(X_i = j_i :\ i = 1, \ldots, n) = \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

d.h. $Y$ ist eine Binominalverteilung, $Y \sim \mathrm{B}(n, p)$. Setzt man $X = (X_1, \ldots, X_n)$, so ist $X$ eine meßbare Abbildung von $(\Omega, \mathcal{A}, P)$ nach $\mathcal{X} = \{0, 1\}^n$, d.h. $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, 2^{\mathcal{X}})$. Es heißt dann $P^X$ über $(\mathcal{X}, 2^{\mathcal{X}})$ gemeinsame Verteilung der $X_i$, $i = 1, \ldots, n$; sie ist gegeben durch

$$P^X\big(\{(i_1, \ldots, i_n)\}\big) = \prod_{j=1}^{n} P^{X_j}\big(\{i_j\}\big)$$

(da die $X_i$ stochastisch unabhängig sind). $P^X$ heißt auch Produktmaß und man schreibt

$$P^X = \bigotimes_{i=1}^{n} P^{X_i}.$$

$(\mathcal{X}, 2^{\mathcal{X}}, P^X)$ heißt Produktraum mit $\mathcal{X} = \underbrace{\{0,1\} \times \ldots \times \{0,1\}}_{n \text{ mal}}$ und

$$2^{\mathcal{X}} = \sigma\big(\{A_1 \times \ldots \times A_n : \ A_i \in 2^{\{0,1\}} \quad \forall\, i = 1, \ldots, n\}\big).$$

Letztlich kann man sagen: Bei gegebenen diskreten Wahrscheinlichkeitsräumen $(\Omega_i, 2^{\Omega_i}, P_i)$ gibt es also keine Schwierigkeit, einen Produktraum zu definieren. Manchmal ist die Situation schwieriger. Wie ist zum Beispiel ein Produktmaß zu definieren?

**Definition 2.19:**
Gegeben seien die Maßräume $(\Omega_i, \mathcal{A}_i)$, $i = 1, \ldots, n$ und $\Omega = \times_{i=1}^n \Omega_i$. $pr_i : \Omega \to \Omega_I$ definiert durch $pr_i(\omega) = \omega_i$ (Projektionsabbildung). Dann heißt

$$\mathcal{A} := \bigotimes_{i=1}^n \mathcal{A}_i := \sigma(pr_i : i = 1, \ldots, n)$$

Produkt-$\sigma$-Algebra.

**Bemerkung 2.20:**

(a) $\bigotimes_{i=1}^n \mathcal{A}_i$ ist also die kleinste $\sigma$-Algebra, bzgl. der die Projektionsabbildungen $pr_i : \Omega \to \Omega_i$ $\mathcal{A}$-$\mathcal{A}_i$-meßbar sind.

(b) Falls $\mathcal{E}_i \subseteq 2^{\Omega_i}$ mit $\sigma(\mathcal{E}_i) = \mathcal{A}_i$ und es existiert $(E_{ij})_{j \in \mathbb{N}} \subset \mathcal{E}_i$ mit $E_{ij} \uparrow \Omega_i$. Dann gilt

$$\bigotimes_{i=1}^n \mathcal{A}_i = \sigma\big(\{E_1, \ldots, E_n : \ E_i \in \mathcal{E}_i \ \forall\, i = 1, \ldots, n\}\big)$$

**Resultat aus der Maßtheorie:**

**Satz 2.21:**
Gegeben seien die $\sigma$-endlichen Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i)$ für $i = 1, \ldots, n$ (d.h. $\mu_i$ $\sigma$-endlich, also $\exists\ (A_{ij})_{j \in \mathbb{N}}$ mit $A_{ij} \uparrow \Omega_i$ und $\mu_i(A_{ij}) < \infty$). Dann gibt es genau ein Maß

$$\mu := \bigotimes_{i=1}^n \mu_i = \mu_1 \otimes \mu_2 \otimes \ldots \otimes \mu_n \quad \text{auf} \quad (\Omega, \mathcal{A}) = \Big( \times_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i \Big),$$

das sogenannte Produktmaß der $\mu_i$, so daß gilt

$$\mu(A_1 \times \ldots \times A_n) = \prod_{j=1}^n \mu_j(A_j) \qquad \forall\, i = 1, \ldots, n, \quad \forall\, A_i \in \mathcal{A}_i.$$

**Definition 2.22:**
Für je endlich viele $\sigma$-endliche Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, \ldots, n$, heißt

$$\Big( \times_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i, \bigotimes_{i=1}^n \mu_i \Big)$$

das Produkt dieser Maßräume. Bezeichnung $\bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mu_i)$.

**Bemerkung 2.23:**
Sind die $\mu_i$ in Satz 2.21 Wahrscheinlichkeitsmaße $P_i$, also insbesondere $\sigma$-endlich (da $P(\Omega_i) = 1 < \infty$),

so erhält man das eindeutig bestimmte Produktwahrscheinlichkeitsmaß

$$P := \bigotimes_{i=1}^{n} P_i.$$

Damit zurück zu den Zufallsvariablen. Wir machen eine Vorbetrachtung: Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $X_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$, $i = 1, \ldots, n$. Definiere $X = (X_1, \ldots, X_n) : \Omega \to \times_{i=1}^{n} \Omega_i$ vermöge $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$. Dann gilt für jede Menge $B_1 \times \ldots \times B_n$ mit $B_i \in \mathcal{A}_i$ für alle $i$:

$$X^{-1}(B_1 \times \ldots \times B_n) = X_1^{-1}(B_1) \cap \ldots \cap X_n^{-1}(B_n) \in \mathcal{A} \tag{3}$$

d.h. wegen Bemerkung 2.9 (a) ist $X$ $\mathcal{A}$-$\bigotimes_{i=1}^{n} \mathcal{A}_i$-meßbar. Insbesondere sind damit die Verteilungen $P^{X_i}$ der $X_i$ und $P^X$ von $X$ definiert. $P^X = P^{(X_1, \ldots, X_n)}$ heißt die gemeinsame Verteilung der $X_1, \ldots, X_n$.

Der folgende Satz gibt Auskunft darüber, wann $X_1, \ldots, X_n$ stochastisch unabhängig sind.

**Satz 2.24:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \to (\mathcal{X}_i, \mathcal{B}_i)$ für $i = 1, \ldots, n$. Dann gilt:

$$X_1, \ldots, X_n \text{ stochastisch unabhängig} \iff P^{(X_1, \ldots, X_n)} = \bigotimes_{i=1}^{n} P^{X_i}$$

Beweis: Für $i = 1, \ldots, n$ sei $B_i \in \mathcal{B}_i$. Dann folgt mit $X = (X_1, \ldots, X_n)$ wegen (2.3)

$$P^X(B_1 \times \ldots \times B_n) = P\big(X^{-1}(B_1 \times \ldots \times B_n)\big) = P\Big(\bigcap_{i=1}^{n} X_i^{-1}(B_i)\Big)$$

sowie $P^{X_i}(B_i) = P\big(X_i^{-1}(B_i)\big)$. $P^X$ ist genau dann das Produktmaß der $P^{X_i}$ (nach Satz 2.20), wenn für beliebige $B_i \in \mathcal{B}_i$ gilt:

$$P^X(B_1 \times \ldots \times B_n) = P^{X_1}(B_1) \cdot \ldots \cdot P^{X_n}(B_n).$$

Äquivalent dazu ist die Forderung

$$\forall\, i = 1, \ldots, n, \quad \forall\, B_i \in \mathcal{B}_i : \qquad P\Big(\bigcap_{i=1}^{n} X_i^{-1}(B_i)\Big) = \prod_{i=1}^{n} P\big(X_i^{-1}(B_i)\big).$$

Dies ist aber wegen Satz 2.17 (oder auch Bemerkung 2.15) äquivalent zur stochastischen Unabhängigkeit von $X_1, \ldots, X_n$.

**Bemerkung 2.25:**

(a) Satz 2.24 zeigt auch, wie man zu gegebenen Wahrscheinlichkeitsräumen $(\Omega_i, \mathcal{A}_i, P_i)$ einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und darauf definierte Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$ definiert, die stochastisch unabhängig sind und vorgegebenen Verteilungen $P_i$ haben: Wähle

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

und $X_i := pr_i$.

(b) (vgl. Bemerkung 2.17) Beschreibt jeder der Wahrscheinlichkeitsräume $(\Omega_i, \mathcal{A}_i, P_i)$ ein Zufallsexperiment $\mathcal{Z}_i$ mit zufälligem Ausgang, so sollte $\bigotimes_{i=1}^{n} (\Omega_i, \mathcal{A}_i, P_i)$ dasjenige Zufallsexperiment $\mathcal{Z}$ beschreiben, welches darin besteht, daß die einzelnen $\mathcal{Z}_i$ „ohne gegenseitige Beeinflussung“ nacheinander oder nebeneinander ablaufen. Die Sprechweise „ohne gegenseitige Beeinflussung“ wird von nun an ersetzt durch „(stoch.) unabhängig“. Die oben definierten $X_i$ beschreiben den Ausgang des Experimentes $\mathcal{Z}_i$ im Gesamtexperiment $\mathcal{Z}$.

**Beispiel 2.26 ($k$-Stichprobenmodell):**
Wir wollen $k \geq 2$ Behandlungsmethoden (Kreislaufmittel) vergleichen. Dazu werden $n$ Patienten streng zufällig in Gruppen vom Umfang $n_i$ eingeteilt mit

$$\sum_{i=1}^{k} n_i = n.$$

Nach der Behandlung werden die interessanten Merkmale (Blutdruck, Veränderung des Blutdrucks, ...) gemessen. Nehmen wir an, es wird nur geprüft, ob das jeweilige Medikament wirkt oder nicht wirkt. Die Ergebnisse kann man dann als Realisierung von stochastisch unabhängigen $\mathrm{B}(1, p_i)$-verteilten Zufallsvariablen auffassen.

Modell: Zufallsvariablen $X_{ij} \sim \mathrm{B}(1, p_i),\ j = 1, \ldots, n_i,\ i = 1, \ldots, k$ stochastisch unabhängig. Dabei sind $p_i \in [0, 1]$ die unbekannten Parameter ($n_i \in \mathbb{N},\ i = 1, \ldots, k$).

$$X_{ij}(\omega) = \begin{cases} 1 & \text{Behandlung hat bei Patient } j \text{ in Gruppe } n_i \text{ gewirkt} \\ 0 & \text{sonst} \end{cases}$$

Fragestellungen, die in der Statistik behandelt werden, sind zum Beispiel, welche Behandlung am besten ist, ob alle Behandlungen gleich gut sind oder für welche Behandlung $p_i \geq 0.90$ ist.

Ein nützliches Hilfsmittel zur Charakterisierung der Verteilung von reellwertigen (oder $\mathbb{R}^n$-wertigen) Zufallsvariablen wird durch die folgende Definition eingeführt.

**Definition 2.27:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann heißt $F : \mathbb{R} \to [0, 1]$, definiert durch $F(x) := P\big((-\infty, x]\big)$, die Verteilungsfunktion von $P$. Ist $P := P^X$ die Verteilung einer Zufallsvariablen $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$, so heißt $F$ die Verteilungsfunktion von $X$. Dann ist $F(x) = P(X \leq x)$.

**Beispiel 2.28:**

(a) Sei $P = \mathrm{G}([0, 1])$, d.h. stetige Gleichverteilung auf $[0, 1]$. Dann ist

$$F(x) = x \cdot I_{[0,1]}(x) + I_{(1,\infty)}(x) \qquad \forall\, x \in \mathbb{R}.$$

(b) Sei $X \sim \mathrm{B}(1, p),\ p \in (0, 1)$, dann ist

$$F(x) = P^X\big((-\infty, x]\big) = P\big(X \in (-\infty, x)\big) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

**Satz 2.29:**

(a) Die Verteilungsfunktion $F$ eines Wahrscheinlichkeitsmaßes auf $(\mathbb{R}, \mathbb{B})$ hat folgende Eigenschaften:

(i) $F$ ist monoton wachsend,

(ii) $F$ ist rechtsseitig stetig,

(iii) Es gilt $\lim\limits_{x \to -\infty} F(x) = 0$ und $\lim\limits_{x \to \infty} F(x) = 1$.

(b) Ist umgekehrt $F : \mathbb{R} \to [0, 1]$ eine Funktion mit den Eigenschaften (i) bis (iii), so gibt es ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}, \mathbb{B})$ mit der Verteilungsfunktion $F$.

Beweis:

(a) Sei $F$ eine Verteilungsfunktion auf $(\mathbb{R}, \mathbb{B})$.

(i) Sei $x < y$, dann ist $F(x) = P\big((-\infty, x]\big) \leq P\big((-\infty, y]\big) = F(y)$.

(ii) Sei $x_n \downarrow x$, dann ist auch $(-\infty, x_n] \downarrow (-\infty, x]$, d.h. $\bigcap\limits_{n\in\mathbb{N}} (-\infty, x_n] = (-\infty, x]$. Mit dem Stetigkeitssatz (Satz 1.20) folgt

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P\big((-\infty, x_n]\big) = P\big((-\infty, x]\big) = F(x)$$

(iii) Es gilt mit Satz 1.20:

$$\begin{aligned} \lim_{x\to-\infty} F(x) &= \lim_{n\to\infty} P\big((-\infty, -n]\big) = P(\emptyset) = 0 \\ \lim_{x\to\infty} F(x) &= \lim_{n\to\infty} P\big((-\infty, n]\big) = P(\mathbb{R}) = 1 \end{aligned}$$

Anmerkung: Ist $x_n \uparrow x$, so folgt

$$\lim_{n\to\infty} F(x_n) = F(x^-) = P\big((-\infty, x)\big)$$

**Lemma 2.30:**
Für die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes $P$ auf $(\mathbb{R}, \mathbb{B})$ gilt, daß sie höchstens abzählbar viele Sprungstellen hat.

**Definition 2.31:**

(a) Gegeben sei ein Wahrscheinlichkeitsraum $(\mathbb{R}^n, \mathbb{B}^n, P)$. Dann heißt $F : \mathbb{R}^n \to [0,1]$, definiert durch

$$\forall\, x = (x_1, \ldots, x_n) \in \mathbb{R}^n : \qquad F(x_1, \ldots, x_n) := P\Big(\times_{i=1}^n (-\infty, x_i]\Big),$$

die ($n$-dimensionale) Verteilungsfunktion von $P$. Ist $pr_i$ die $i$-te Projektionsabbildung, so heißt $P_i := P^{pr_i}$ die $i$-te Randverteilung von $P$ und $F_i$ bezeichnet die Verteilungsfunktion von $P_i$, die $i$-te Randverteilungsfunktion von $F$. Es ist

$$F_i(x) = P\big(\{\omega : pr_i(\omega) \leq x\}\big).$$

(b) Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und Zufallsvariablen $X_i : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$ und $X = (X_1, \ldots, X_n)$. $F$ sei die Verteilungsfunktion von $P^X$. Dann heißt $F$ die ($n$-dimensionale) Verteilungsfunktion von $X$, d.h. für alle $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt

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

**Bemerkung 2.32:**

(a) Unter den obigen Voraussetzungen gilt für $F$ und die zugehörigen $F_i$:

(i) $F_i(x_i) = \lim\limits_{\substack{x_j \to \infty \\ \forall\, j\neq i}} F(x_1, \ldots, x_n) =: F(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty)$

(ii) ($\alpha$) $F$ ist in jeder Variablen rechtsseitig stetig,

($\beta$) Für alle $a, b \in \mathbb{R}^n$ mit $a \leq b$ (komponentenweise) gilt $\Delta_a^b F \geq 0$ ($\Delta$-isoton), dabei ist

$$\begin{aligned} \Delta_a^b F &:= \Delta_{a_1 b_1}^1 \ldots \Delta_{a_n b_n}^n F \\ \Delta_{a_i b_i}^i F &:= F(\cdot, \ldots, \cdot, b_i, \cdot, \ldots, \cdot) - F(\cdot, \ldots, \cdot, a_i, \cdot, \ldots, \cdot) \end{aligned}$$

($\gamma$) Für alle $x \in \mathbb{R}^n$ gilt: $\lim\limits_{x_i \to -\infty} F(x) = 0 \quad \forall\, i = 1, \ldots, n$

($\delta$) Es ist $\lim\limits_{x\to\infty\cdot 1_n} F(x) = 1$

($\eta$) Es ist $\Delta_a^b F = P\big((a,b]\big)$ mit $(a,b] := \times_{i=1}^n (a_i, b_i]$

Beispiel: Sei $n = 2$, $a_i \le b_i$ für alle $i = 1, 2$. Dann ist

$$\begin{aligned}\Delta_a^b F &= \Delta^1_{a_1 b_1} \Delta^2_{a_2 b_2} F = \Delta^1_{a_1 b_1} \Big( F(\cdot, b_2) - F(\cdot, a_2) \Big) = \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)\end{aligned}$$

(b) Für alle $F : \mathbb{R}^n \to [0,1]$ mit Eigenschaften ($\alpha$) bis ($\delta$) existiert genau ein Wahrscheinlichkeitsmaß $P$ auf $(\mathbb{R}^n, \mathbb{B}^n)$, so daß für alle $x \in \mathbb{R}^n$ gilt

$$F(x) = P\big((-\infty, x]\big),$$

bzw. für alle $a, b \in \mathbb{R}^n$: $\quad P\big((a,b]\big) = \Delta_a^b F$ (vgl. Satz 2.29 b / „Korrespondenzsatz“)

**Satz 2.33:**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und $X_i : (\Omega, \mathcal{A}) \to (\mathbb{R}^1, \mathbb{B}^1)$ für $i = 1, \ldots, n$. Sei weiter $X = (X_1, \ldots, X_n)$, $F$ Verteilungsfunktion von $X$ und $F_i$ Verteilungsfunktionen von $X_i$ für $i = 1, \ldots, n$. Dann sind äquivalent:

(a) $X_1, \ldots, X_n$ sind stochastisch unabhängig,

(b) Für alle $x \in \mathbb{R}^n$ gilt: $\quad F(x) = \prod_{i=1}^n F_i(x)$.

Beweis: Sei $\mathcal{D} = \{(-\infty, x] : x \in \mathbb{R}\} \cup \{\mathbb{R}\}$. Dann ist $\mathcal{D}$ ein durchschnittsstabiler Erzeuger von $\mathbb{B}$, d.h. wegen Satz 2.17 ist (a) äquivalent mit

$$\forall\, i = 1, \ldots, n: \quad \forall\, D_i \in \mathcal{D}: \qquad P\Big( \bigcap_{i=1}^n X_i^{-1}(D_i) \Big) = \prod_{i=1}^n P\big( X_i^{-1}(D_i) \big)$$

bzw. äquivalent mit

$$\forall\, i = 1, \ldots, n: \quad \forall\, x \in \{\mathbb{R}\} \cup \{+\infty\}: \qquad P(X_1 \le x_1, \ldots, X_n \le x_n) = \prod_{i=1}^n P(X_i \le x_i)$$

bzw. äquivalent mit

$$\forall\, i = 1, \ldots, n: \quad \forall\, x_i \in \mathbb{R}: \qquad F(x_1, \ldots, x_n) = \prod_{i=1}^n F(x_i).$$

**Bemerkung 2.34:**

(a) Ist $X$ eine diskrete Zufallsvariable mit Träger $\mathcal{X} \subset \mathbb{R}$ und Zähldichte $f$, so ist die Verteilungsfunktion $F$ von $X$ gegeben durch

$$F(x) = \sum_{\substack{x' \le x \\ x' \in \mathcal{X}}} f(x')$$

Bezeichnet $\kappa$ das Zählmaß, so schreibt man auch

$$F(x) = \int\limits_{(-\infty, x] \cap \mathcal{X}} f(x)\, d\kappa(x) := \sum_{\substack{x' \le x \\ x' \in \mathcal{X}}} f(x')$$

(b) Viele reellwertige (nicht-diskrete) Zufallsvariablen besitzen eine Verteilungsfunktion der Gestalt

$$F(x) = \int_{-\infty}^{x} g(t)\, dt$$

mit $g : \mathbb{R} \to [0, \infty)$ und $\int_{-\infty}^{\infty} g(t)\, dt = 1$. Wahrscheinlichkeitsmaße mit solch einer Verteilungsfunktion heißen absolut stetig, $g$ heißt Dichte. Analog für $\mathbb{R}^n$-wertige Zufallsvariablen:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} g(t_1, \ldots, t_n)\, d(t_1, \ldots, t_n)$$

mit $g : \mathbb{R}^n \to [0, \infty)$ und

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(t_1, \ldots, t_n)\, d(t_1, \ldots, t_n) = 1.$$

Sind nun $X_1, \ldots, X_n$ reellwertig, $P^{X_1}, \ldots, P^{X_n}$ absolut stetig mit Dichten $g_1, \ldots, g_n$, so sind folgende Aussagen äquivalent:

(i) $X_1, \ldots, X_n$ stochastisch unabhängig,

(ii) Für alle $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ gilt:

$$P(X_1 \le x_1, \ldots, X_n \le x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod_{i=1}^{n} g_i(t_i)\, dt_1 \cdots dt_n,$$

(iii) Für alle $i = 1, \ldots, n$ und $B_i \in \mathbb{B}$ gilt:

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = \int_{B_1} \cdots \int_{B_n} \prod_{i=1}^{n} g_i(t_i)\, dt_1 \cdots dt_n.$$

Es heißt $g(x) := \prod_{i=1}^{n} g_i(x_i)$ Produktdichte für $x \in \mathbb{R}^n$.

(c) Sind die $X_i$ diskret verteilt mit Träger $\mathcal{X}_i$ und Zähldichte $f_i$ $(i = 1, \ldots, n)$, so sind äquivalent:

(i) $X_1, \ldots, X_n$ stochastisch unabhängig,

(ii) Für alle $i = 1, \ldots, n$ und für alle $x_i \in \mathcal{X}_i$ gilt

$$P(X_i = x_i,\ i = 1, \ldots, n) = \prod_{i=1}^{n} f_i(x_i),$$

(iii) Für alle $i = 1, \ldots, n$ und für alle $B_i \subseteq \mathcal{X}_i$ gilt

$$P(X_i \in B_i,\ i = 1, \ldots, n) = \prod_{i=1}^{n} \sum_{x_i \in B_i} f_i(x_i) = \prod_{i=1}^{n} \int_{B_i} f_i(x_i)\, d\kappa(x_i).$$

Für $x \in \times_{i=1}^{n} \mathcal{X}_i$ heißt $f(x) := \prod_{i=1}^{n} f_i(x_i)$ ebenfalls Produktdichte und ist wiederum eine Zähldichte.

## 2.2 Integrationstheorie in Kürze

Eine wichtige Rolle in der Wahrscheinlichkeitstheorie und Statistik spielen sogenannte Erwartungswerte (Mittelwerte) bzw. Momente von Zufallsvariablen bzw. Verteilungen.

**Beispiel ($n$-facher Münzwurf):**
Wieviele Einsen kann man im Mittel erwarten? Bei diskreten Zufallsvariablen ist dies noch einfach: Man definiert

$$EX := \int_{\mathcal{X}_0} x \cdot f(x)\, d\kappa(x) := \sum_{x \in \mathcal{X}_0} x \cdot f(x)$$

mit Zähldichte $f$ und $\mathcal{X}_0$ als Träger von $P^X$. Für $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, \ldots, n$, erhält man

$$\begin{aligned} EX &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-1-(x-1)} = \\ &= np \underbrace{\sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-1-x}}_{=1} = np \end{aligned}$$

Ist $P$ jedoch beliebig, so ist ein anderer Integralbegriff notwendig. Was ist zum Beispiel $EX = \int X\, dP =$? bzw. für beliebige Maße $\mu$ und meßbare (numerische) Funktionen $f$.

**1. Schritt:** Approximation meßbarer numerischer Funktionen $f : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$ durch primitive Funktionen. Im folgenden sei immer $(\Omega, \mathcal{A})$ und

$$\mathcal{M} := \{f : f \text{ meßbare numerische Funktion}\}$$

gegeben.

(1) Indikatorfunktion:

$$I_A \in \mathcal{M} \quad \Longleftrightarrow \quad A \in \mathcal{A}.$$

(2) Primitive Funktionen:

**Definition:** Für $A_i \in \mathcal{A}$, $i = 1, \ldots, n$, paarweise disjunkt mit $\sum_{i=1}^{n} A_i = \Omega$ und $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ heißt

$$g := \sum_{i=1}^{n} \alpha_i I_{A_i} : \Omega \to \mathbb{R}$$

primitive Funktion.

(3) Approximation nicht negativer Funktionen $f \in \mathcal{M}$ durch primitive Funktionen.

**Satz 2.35:**
Es gilt:

$$f \in \mathcal{M},\ f \geq 0 \quad \Longleftrightarrow \quad \exists\, (f_n)_{n \in \mathbb{N}} : \forall\, n \in \mathbb{N} :\ f_n \geq 0,\ \text{primitiv und } f_n \uparrow f,$$

d.h. jede nicht negative meßbare numerische Funktion ist isotoner Limes einer Folge nicht negativer primitiver Funktionen.

Beweis (in Kurzform):

„$\Leftarrow$“ z.z. $\lim f_n \in \mathcal{M}$, und das gilt (Bem. 2.9 Teil f).

„⇒“ Die primitiven Funktionen

$$f_n := \sum_{k=1}^{n2^n} \frac{k-1}{n} I_{\{\frac{k-1}{2^n} \le f < \frac{k}{2^n}\}} + n I_{\{f \ge n\}} \qquad (n \in \mathbb{N})$$

leisten das Gewünschte ...

(4) Beliebige meßbare Funktionen $f \in \mathcal{M}$: Definiere dazu

$$f^+ := \max\{f, 0\} \quad \text{und} \quad f^- := -\min\{f, 0\}.$$

Dann ist $f^+ - f^- = f$ und es sind $f^+, f^- \ge 0$. Weiter gilt

$$f \in \mathcal{M} \iff f^+, f^- \in \mathcal{M}.$$

Damit ist die Approximation von beliebigen $f \in \mathcal{M}$ durch primitive Funktionen klar.

**2. Schritt:** Definition des $\mu$-Integrals in vier Schritten (entsprechend obigen vier Approximationsschritten von $f \in \mathcal{M}$ durch primitive Funktionen).

(1) Für $f = I_A \in \mathcal{M}$ sei $\int I_A \, d\mu := \mu(A)$. Andere Schreibweise

$$\int_A d\mu := \int I_A \, d\mu =: \int I_A(\omega) \, d\mu(\omega).$$

Es folgt sofort die Nulltreue $\int 0 \, d\mu = \int I_\emptyset \, d\mu = \mu(\emptyset) = 0$.

(2) Für primitives $f = \sum_{i=1}^{n} \alpha_i I_{A_i} \in \mathcal{M}$ sei

$$\int f \, d\mu = \int \sum_{i=1}^{n} \alpha_i I_{A_i} \, d\mu := \sum_{i=1}^{n} \alpha_i \int I_{A_i} \, d\mu \overset{(1)}{=} \sum_{i=1}^{n} \alpha_i \mu(A_i).$$

**Bemerkung:**
Die Darstellung einer primitiven Funktion ist nicht eindeutig, zum Beispiel

$$f = I_{\mathbb{R}} - \frac{1}{2} I_{(0,1]}.$$

Obige Definition ist dennoch eindeutig, denn sei

$$f = \sum_{i=1}^{n} \alpha I_{A_i} = \sum_{j=1}^{m} \beta_j I_{B_j}$$

mit $\sum_{i=1}^{n} A_i = \sum_{j=1}^{m} B_j = \Omega$. Es folgt sofort $\alpha_i = \beta_j$, falls $A_i B_j \neq \emptyset$. Weiter gilt

$$\begin{aligned} \int \sum_{i=1}^{n} \alpha_i I_{A_i} \, d\mu &= \sum_{i=1}^{n} \alpha_i \mu(A_i) = \sum_{i=1}^{n} \alpha_i \mu(A_i \sum_{j=1}^{m} B_j) = \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \mu(A_i B_j) = \sum_{j=1}^{m} \sum_{i=1}^{n} \beta_j \mu(B_j A_i) = \ldots = \\ &= \int \sum_{j=1}^{m} \beta_j I_{B_j} \, d\mu \end{aligned}$$

**Lemma 2.36:**
Seien $f, g \geq 0$ primitiv, $f, g \in \mathcal{M}$ und $\lambda \geq 0$. Dann gilt:

(i) Linearität:
$$\int (f + \lambda g)\, d\mu = \int f\, d\mu + \lambda \cdot \int g\, d\mu.$$

(ii) Isotonie: Seien $0 \leq f \leq g$. Dann gilt: $0 \leq \int f\, d\mu \leq \int g\, d\mu$

**Beispiel 2.37:**

(i) Sei $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathbb{B}, \lambda)$ und $f = I_{[0,1]\cap\mathbb{Q}}$. Dann gilt
$$\int f\, d\lambda = \lambda\big([0,1] \cap \mathbb{Q}\big) = 0,$$
aber das Riemann-Integral ist nicht erklärt.

(ii) Sei $f_n = \frac{1}{n} I_{[0,n]}$ mit $n \in \mathbb{N}$. Dann ist
$$\int f_n\, d\lambda = \frac{1}{n}\lambda\big([0,n]\big) = 1 = \lim_{n\to\infty} \int f_n\, d\lambda,$$
aber es ist
$$\int \lim_{n\to\infty} f_n\, d\lambda = \int 0 \cdot I_{[0,\infty)}\, d\lambda = 0 \cdot \lambda\big([0,\infty)\big) = 0 \cdot \infty = 0,$$
d.h. man kann i.A. nicht Limes und Integral vertauschen.

(3) Für $\mathcal{M} \ni f \geq 0$ seien $f_n \in \mathcal{M}$ primitiv mit $f_n \geq 0$ und $f_n \uparrow f$. Definiere
$$\int f\, d\mu := \lim_{n\to\infty} \int f_n\, d\mu.$$

Diese Definition ist eindeutig wegen (ii) im folgenden Lemma.

**Lemma 2.38:**
Seien $g, f_n, g_n \geq 0$, primitiv mit $g, f_n, g_n \in \mathcal{M}$. Dann gilt:

(i) Falls $f_n \uparrow$ und $g \leq \lim f_n$, so ist $\int g\, d\mu \leq \lim \int f_n\, d\mu$.

(ii) Falls $f_n \uparrow$, $g_n \uparrow$ und $\lim f_n = \lim g_n$, so ist $\lim \int f_n\, d\mu = \lim \int g_n\, d\mu$.

Beweis:

(i) Beweis aufwendig, benutze u.a. Lemma 2.36.

(ii) folgt aus (i), denn für alle $m \in \mathbb{N}$ gilt
$$f_m \leq \lim f_n = \lim g_n \quad \overset{(i)}{\Rightarrow} \quad \int f_m\, d\mu \leq \lim \int g_n\, d\mu.$$
Somit folgt
$$\lim \int f_n\, d\mu \leq \lim \int g_n\, d\mu.$$
Die Gleichheit „=“ folgt aus Symmetriegründen.

**Lemma 2.39 (Verallg. von Lemma 2.36):**
Sind $f, g \geq 0$, $f, g \in \mathcal{M}$ und $\lambda \geq 0$. Dann folgen (i) und (ii) aus Lemma 2.36.

(4) Sei $f \in \mathcal{M}$ und $f^+, f^-$ wie oben definiert. Dann

(i) heißt $f$ $\mu$-integrierbar ($\mu$-integrabel)

$$:\Longleftrightarrow \quad \int f^+ \, d\mu < \infty \quad \text{und} \quad \int f^- \, d\mu < \infty.$$

(ii) heißt $f$ $\mu$-quasiintegrierbar

$$:\Longleftrightarrow \quad \int f^+ \, d\mu < \infty \quad \text{oder} \quad \int f^- \, d\mu < \infty.$$

In beiden Fällen heißt $\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu$ das $\mu$-Integral von $f$.

**Satz 2.40:**
Für $f \in \mathcal{M}$ sind äquivalent

(i) $f$ ist $\mu$-integrierbar,

(ii) $f^+$ und $f^-$ sind $\mu$-integrierbar,

(iii) Es existieren $g, h \in \mathcal{M}$ $\mu$-integrierbar, $g, h \geq 0$ mit $f = g - h$,

(iv) Es existiert $\tilde{g} \in \mathcal{M}$ $\mu$-integrierbar mit $|f| \leq \tilde{g}$,

(v) $|f|$ ist $\mu$-integrierbar.

**Satz 2.41 (Verallg. von 2.39):**
Sind $f, g$ $\mu$-integrierbar, so auch $f + \lambda g$ (falls definiert) mit $\lambda \in \mathbb{R}$ und $\max\{f, g\} < \infty$.

(i) Linearität:

$$\int (f + \lambda g) \, d\mu = \int f \, d\mu + \lambda \cdot \int g \, d\mu.$$

(ii) Isotonie: Ist $f \leq g$, so ist $\int f \, d\mu \leq \int g \, d\mu$.

(iii) Es gilt: $\left|\int f \, d\mu\right| \leq \int |f| \, d\mu$.

**Bemerkung 2.42:**

(i) Satz 2.41 gilt auch für quasiintegrierbare Funktionen $f, g$, wenn man in (i) auch noch voraussetzt, daß die rechte Seite definiert ist.

(ii) Für $f, g \in \mathcal{M}_{\mathbb{R}} := \{h \in \mathcal{M} : \, h(\Omega) \subseteq \mathbb{R}\}$ ist $f + \lambda g$ in Satz 2.41 immer definiert.

(iii) Es ist $\mathcal{L}^1(\mu) := \{f \in \mathcal{M}_R : \, f \ \mu\text{-integrierbar}\}$ ein Vektorraum über $\mathbb{R}$ bezüglich

$$(f + g)(\omega) := f(\omega) + g(\omega), \qquad (\lambda f)(\omega) := \lambda \cdot f(\omega).$$

(iv) Für $A \in \mathcal{A}$ und $f \in \mathcal{L}^1(\mu)$ definiert man

$$\int\limits_A f \, d\mu := \int f \cdot I_A \, d\mu.$$

Es folgt eine Zusammenstellung wichtiger Konvergenz- und Grenzvertauschungssätze. Zunächst noch eine wichtige Definition.

**Definition 2.43:**
Gegeben sei ein Maßraum $(\Omega, \mathcal{A}, \mu)$. Ist $E$ eine Eigenschaft, die ein jedes $\omega \in \Omega$ besitzt oder nicht, dann sagt man „$E$ gilt $\mu$-fast sicher (oder $\mu$-fast überall)“ und schreibt

$$\begin{aligned} &E \text{ } \mu\text{-fast sicher oder } E \text{ } \mu\text{-fast überall oder } E[\mu] :\Longleftrightarrow \\ &:\Longleftrightarrow \quad \underbrace{\exists\, N \in \mathcal{A} :\ \mu(N) = 0}_{\exists\ \mu\text{-Nullmenge}} \text{ und } \forall\, \omega \in N^c :\ E \text{ gilt für } \omega. \end{aligned}$$

Beachte: Die Menge $A := \{\omega \in \Omega : E \text{ gilt nicht für } \omega\}$ ist nicht notwendig in $\mathcal{A}$, sondern nur $A \subseteq N$ und $N \in \mathcal{A}$.

**Beispiel:**

(i) Es ist

$$\begin{aligned} f \le g[\mu] \quad &\Longleftrightarrow \quad \exists\, N \in \mathcal{A} :\ \mu(N) = 0 \text{ und } \forall\, \omega \in N^c :\ f(\omega) \le g(\omega) \\ &\Longleftrightarrow \quad \exists\ \mu\text{-Nullmenge } N \in \mathcal{A} :\ N \supset \{f > g\} \\ &\overset{(*)}{\Longleftrightarrow} \quad \mu(f > g) = 0, \end{aligned}$$

denn $(*)$ $f, g \in \mathcal{M}$.

(ii) Ist $f = I_{\mathbb{Q}}$, dann ist $f = 0[\mu]$.

**Bemerkung 2.44:**
Es gilt für alle $f \in \mathcal{M}$

(i) Ist $f = 0[\mu]$, so ist $\int f\, d\mu = 0$.

(ii) Ist $f \ge 0$ und $\int f\, d\mu = 0$, so ist $f = 0[\mu]$.

Im folgenden sei stets der Maßraum $(\Omega, \mathcal{A}, \mu)$ gegeben.

**Satz 2.45 (Satz von der monotonen Konvergenz):**
Sei $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$. Für alle $n \in \mathbb{N}$ sei $f_n \le f_{n+1}[\mu]$. Es existiere ein $h \in \mathcal{M}$ $\mu$-integrierbar mit $h \le f_1[\mu]$. Dann gilt

$$\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu \qquad \text{mit } f(\omega) := \begin{cases} \lim\limits_{n\to\infty} f_n(\omega) & \text{falls existend} \\ 0 & \text{sonst} \end{cases}$$

**Satz 2.46 (Lemma von Fatou):**
Für $h \in \mathcal{M}$ $\mu$-integrierbar und $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ gilt

(i) Ist $f_n \ge h[\mu]$ für alle $n \in \mathbb{N}$, so gilt:

$$\int \liminf f_n\, d\mu \le \liminf \int f_n\, d\mu.$$

(ii) Ist $f_n \le h[\mu]$ für alle $n \in \mathbb{N}$, so gilt:

$$\limsup \int f_n\, d\mu \ge \int \limsup f_n\, d\mu.$$

**Satz 2.47 (Satz von der beschränkten Konvergenz, Lebesgue):**
Sei $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ $\mu$-fast sicher konvergent und $\mu$-fast sicher $\mu$-integrierbar beschränkt (d.h. $\exists\, h \in \mathcal{M}$ : $\forall\, n \in \mathbb{N}$ : $|f_n| \leq h[\mu]$ und $h$ $\mu$-integrierbar). Dann gilt

(i) Es gilt

$$\lim \int f_n \, d\mu = \int f \, d\mu \qquad \text{mit } f(\omega) = \begin{cases} \lim\limits_{n\to\infty} f_n(\omega) & \text{falls existend} \\ 0 & \text{sonst} \end{cases}$$

(ii) $f$ ist $\mu$-integrierbar.

(iii) Es ist $\lim \int |f_n - f| \, d\mu = 0$.

**Satz 2.48:**
Sei $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}_{\mathbb{R}}$ $\mu$-integrierbar, $f_n \to f$ gleichmäßig $\mu$-endlich. Dann ist $f \in \mathcal{M}_{\mathbb{R}}$ $\mu$-integrierbar und

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

**Satz 2.49:**
Sei $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ mit $\sum\limits_{n\in\mathbb{N}} \int f_n^+ \, d\mu < \infty$ oder $\sum\limits_{n\in\mathbb{N}} \int f_n^- \, d\mu < \infty$. Dann gilt

(i) $\lim\limits_{n\to\infty} \sum\limits_{j=1}^{n} \int f_j \, d\mu = \sum\limits_{n\in\mathbb{N}} \int f_n^+ \, d\mu - \sum\limits_{n\in\mathbb{N}} \int f_n^- \, d\mu$

(ii) $\sum\limits_{n\in\mathbb{N}} f_n$ konvergiert $\mu$-fast sicher gegen eine $\mu$-quasiintegrierbare Funktion $f$.

(iii) $\int \sum\limits_{n\in\mathbb{N}} f_n \, d\mu = \sum\limits_{n\in\mathbb{N}} \int f_n \, d\mu$

Beweise: siehe zum Beispiel Bauer oder Hinderer.

**Bemerkung:**
Ist $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}^n, \mathbb{B}^n, \lambda^n)$. Sei $f \in \mathcal{M}(\mathbb{R}^n, \mathbb{B}^n) = \{f |\ f : (\mathbb{R}^n, \mathbb{B}^n) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})\}$. $f$ sei $\lambda^n$-quasiintegrierbar. Dann heißt

$$\int f \, d\lambda^n = \int f(x) \, d\lambda^n(x)$$

Lebesgue-Integral von $f$. Wir wollen uns jetzt den Zusammenhang Lebesgue-/Riemann-Integral für $n = 1$ etwas genauer anschauen:

**Satz 2.50:**
Für $f : [a, b] \to \mathbb{R}$ beschränkt bei $a, b \in \mathbb{R}$ mit $a < b$ und $f \cdot I_{[a,b]} : \mathbb{R} \to \mathbb{R}$ gilt:

(i) $f$ Riemann-integrierbar $\iff$ $f \cdot I_{[a,b]}$ stetig$[\mu]$.

(ii) Ist $f$ Riemann-integrierbar, so gilt:

(1) $\exists\, \tilde{f} \in \mathcal{M}_{\mathbb{R}}(\mathbb{R}, \mathbb{B})$ : $f \cdot I_{[a,b]} = \tilde{f}[\lambda]$

(2) $f \cdot I_{[a,b]}$ ist $\lambda$-integrierbar.

(3) Es ist

$$\int_a^b f(x) \, dx = \int f \cdot I_{(a,b]} \, d\lambda = \int_{(a,b]} f(x) \, d\lambda(x).$$

Für eigentliche Riemann-Integrale beschränkter Funktionen ist also das Lebesgue-Integral ($\lambda$-Integral) eine echte Verallgemeinerung (z.B. $f = I_{\mathbb{Q}\cap[0,1]}$). Für uneigentliche Riemann-Integrale gilt Satz 2.50 im allgemeinen nicht mehr. Zum Beispiel für $f(x) := \frac{\sin(x)}{x}$ existiert das uneigentliche Riemann-Integral $\int_{-\infty}^{\infty} f(x)\,dx$, aber $f$ ist nicht $\lambda$-quasiintegrierbar, da $\int f^+\,d\lambda = \int f^-\,d\lambda = \infty$.

**Satz 2.51:**
Für $f : \mathbb{R} \to \mathbb{R}$ mit $f \geq 0$ existiere das Riemann-Integral $\int_a^b f(x)\,dx$ für alle $a, b \in \mathbb{R}$ mit $a < b$. Dann gilt:

(i) $\int_{-\infty}^{\infty} f(x)\,dx := \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b f(x)\,dx = \int f\,d\lambda$

(ii) $\int_{-\infty}^{\infty} f(x)\,dx < \infty \iff \int f\,d\lambda < \infty$

## 2.3 Maße und Dichten

Problem: Gegeben sei ein Wahrscheinlichkeitsraum $(\mathbb{R}^n, \mathbb{B}^n, P)$. Wann gibt es eine Funktion $f \in \mathcal{M}_{\mathbb{R}}$, so daß für alle $A \in \mathbb{B}^n$ gilt: $P(A) = \int_A f\,d\lambda$ ?

Generelle Voraussetzungen: Maßraum $(\Omega, \mathcal{A}, \mu)$, $\mathcal{M}^+ := \{f \in \mathcal{M} : f \geq 0\}$.

**Satz 2.52:**
Für $f \in \mathcal{M}^+$ ist $\nu$, definiert durch

$$\nu(A) := \int_A f\,d\mu$$

mit $A \in \mathcal{A}$, ein Maß auf $\mathcal{A}$. $\nu$ heißt Maß mit der $\mu$-Dichte $f$, kurz $\nu = f\mu$ oder $d\nu = f\,d\mu$. Für diese $\nu$ gilt:

(i) Für alle $\varphi \in \mathcal{M}^+$ gilt $\int \varphi\,d\nu = \int \varphi f\,d\mu$.

(ii) Für alle $\varphi \in \mathcal{M}$ gilt

$$\varphi\ \nu\text{-integrierbar} \iff \varphi f\ \mu\text{-integrierbar} \Rightarrow \int \varphi\,d\nu = \int \varphi f\,d\mu$$

(iii) Für alle $g \in \mathcal{M}^+$ gilt: Ist $f = g[\mu]$, so ist $f\mu = g\mu$. Ist $f$ oder $g$ $\mu$-integrierbar, so gilt auch die Umkehrung „$\Leftarrow$“.

**Definition 2.53:**
Seien $\nu, \mu$ Maße auf $(\Omega, \mathcal{A})$.

(i)

$$\begin{aligned} \nu \text{ heißt } \mu\text{-stetig (oder totalstetig bzgl. } \mu) \quad &:\iff \quad \forall\, N \in \mathcal{A} \text{ gilt: } \big(\mu(N) = 0 \Rightarrow \nu(N) = 0\big) \\ &\iff : \quad \mu \text{ dominiert } \nu \\ &\iff : \quad \mu \gg \nu. \end{aligned}$$

(ii) $\mu$ und $\nu$ heißen äquivalent (kurz $\mu \sim \nu$) genau dann, wenn $\mu \gg \nu$ und $\nu \gg \mu$.

**Satz 2.54 (Radon, Nikodym):**
Für Maße $\mu, \nu$ auf $(\Omega, \mathcal{A})$ gilt bei $\sigma$-endlichem $\mu$:

$$\begin{aligned} \mu \gg \nu \quad &\iff \quad \exists\, f \in \mathcal{M}^+ : \nu = f\mu \\ &\iff \quad \nu \text{ besitzt eine } \mu\text{-Dichte} \end{aligned}$$

**Satz 2.55:**
Für Maße $\mu, \nu$ auf $(\Omega, \mathcal{A})$ mit $\mu$ $\sigma$-endlich, $\nu = f\mu$ und $f \in \mathcal{M}^+$ gilt:

(i) $f$ ist $\mu$-fast sicher eindeutig bestimmt,

(ii) $\nu$ $\sigma$-endlich $\iff$ $f$ $\mu$-fast sicher endlich.

**Definition 2.56:**
Sein $\mu, \nu$ Maße auf $(\Omega, \mathcal{A})$ mit $\mu$ $\sigma$-endlich und $\mu \gg \nu$, so heißt jedes $f \in \mathcal{M}^+$ mit $\nu = f\mu$ Radon-Nikodym-Dichte von $\nu$ bezüglich $\mu$.

**Corollar 2.57 (Kettenregel):**
Seien $f, g \in \mathcal{M}^+$, $\nu, \mu$ Maße auf $(\Omega, \mathcal{A})$ mit $\nu = f\mu$ und $\rho = g\nu$ (d.h. $d\nu = f\, d\mu$, $d\rho = g\, d\nu$). Dann gilt:

$$f \cdot g \in \mathcal{M}^+, \quad \rho = (gf)\mu \text{ bzw. } d\rho = (gf)\, d\mu.$$

Anmerkung: Schreibt man formal $\frac{d\rho}{d\nu} = g$ für $\rho = g\nu$ bzw. $d\rho = g\, d\nu$, so lautet Corollar 2.57:

$$\frac{d\rho}{d\mu} = \frac{d\rho}{d\nu} \cdot \frac{d\nu}{d\mu}.$$

**Beispiel 2.58:**
Sei $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathbb{B}, \lambda)$.

(a) Sei $P$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathbb{B})$ mit $\lambda \gg P$

$$\begin{aligned} &\overset{\text{R.N.}}{\Rightarrow} \quad \exists\, f \in \mathcal{M}^+ : P = f\lambda, \text{ da } \lambda \text{ } \sigma\text{-endlich} \\ &\overset{2.55\,(ii)}{\Rightarrow} \quad \text{o.B.d.A. } f \in \mathcal{M}^+_{\mathbb{R}}, \text{ da } P \text{ endlich.} \end{aligned}$$

Insbesondere ist $\int f\, d\lambda = 1$.

(b) Sei $f \in \mathcal{M}^+$ mit $\int f\, d\lambda = 1$. Dann ist $P := f\lambda$ ein Wahrscheinlichkeitsmaß.

(c) Sei $f \in \mathcal{M}^+$ $\lambda$-integrierbar und $\int f\, d\lambda > 0$, dann ist $P := \frac{1}{\int f\, d\lambda} f\lambda$ ein Wahrscheinlichkeitsmaß. Dabei ist $\frac{1}{\int f\, d\lambda}$ ein Normierungsfaktor.

(d) Ist $f_0(x) = \left(\frac{1}{\sqrt{x}} - 1\right) \cdot I_{(0,1]}(x)$ für $x \in \mathbb{R}$, dann ist $f_0 \in \mathcal{M}^+_{\mathbb{R}}$ und $\int f_0\, d\lambda = 1$. $f_0$ ist endlich, aber nicht $\lambda$-fast sicher beschränkt.

**Beispiel 2.59 (Exponentialverteilung EXP$(\mu, \sigma)$):**
Für $\mu \in \mathbb{R}$ und $\sigma > 0$ fest sei $f : \mathbb{R} \to \mathbb{R}$, definiert durch

$$f(x) = \frac{1}{\sigma} \exp\left\{-\frac{x-\mu}{\sigma}\right\} \cdot I_{[\mu,\infty)}(x). \qquad (x \in \mathbb{R})$$

$f$ ist stetig außer in $x = \mu$. Dann ist $P := f\lambda$ Wahrscheinlichkeitsmaß mit $\lambda$-Dichte $f$ wegen

$$P(\mathbb{R}) \quad = \quad \int f \, d\lambda = \int\limits_{[\mu,\infty)} \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} \, d\lambda(x) \overset{(*)}{=} \int\limits_0^\infty e^{-y} \, dy = -e^{-y}\Big|_0^\infty = 1$$

mit $y = \frac{x-\mu}{\sigma}$ und $dy = \frac{1}{\sigma} \, dx$. $P$ heißt Exponentialverteilung mit Lageparameter (Lokationsparameter) $\mu$ und Skalenparameter $\sigma$. Die Verteilungsfunktion $F$ zu $P$ ist

$$F(x) = P\big((-\infty, x]\big) = \int\limits_{(-\infty,x]\cap[\mu,\infty)} \frac{1}{\sigma} e^{-\frac{t-\mu}{\sigma}} \, d\lambda(t) = \begin{cases} 0 & x \le \mu \\ \int\limits_\mu^x \frac{1}{\sigma} e^{-\frac{t-\mu}{\sigma}} \, dt = 1 - e^{-\frac{x-\mu}{\sigma}} & x > \mu \end{cases}$$

Beachte: $f = F'[\lambda]$. Weitere $\lambda$- bzw. $\lambda^n$-stetige Verteilungen siehe Anhang.

## 2.4 Produktmaße und der Satz von Fubini

Durch Satz 2.21 wurde das Produktmaß eingeführt. Ist zum Beispiel $\mu = \mu_1 \otimes \mu_2$, so erhält man für Rechteckmengen $A = A_1 \times A_2 \in \mathcal{A}_1 \otimes \mathcal{A}_2$

$$\mu(A) = \mu_1(A_1)\mu_2(A_2).$$

Frage: Wie ist $\mu(A)$ für beliebige $A \in \mathcal{A}$ zu berechnen? Oder ist $P$ ein Wahrscheinlichkeitsmaß mit $\lambda^2$-Dichte $f$, wie ist

$$P(A) = \int\limits_A f(x_1, x_2) \, d\lambda^2(x_1, x_2)$$

zu berechnen?

Generelle Voraussetzungen seien jetzt die $\sigma$-endlichen Maßräume $(\Omega_1, \mathcal{A}_i, \mu_i)$ für $i = 1, \ldots, n$. Im folgenden sei stets $n = 2$ (Verallgemeinerung klar, benutze Induktion).

**Vorbemerkung:**

(i) Für $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ heißt $A_{\omega_i} = \{\omega_j \in \Omega_i : \ (\omega_1, \omega_2) \in A\}$ der $\omega_i$-Schnitt von $A$ für $i \neq j$.

(ii) Es gilt:

$$\omega_1 \to \mu_2(A_{\omega_1}) \text{ ist } \mathcal{A}_1 - \overline{\mathbb{B}}\text{-meßbar.}$$
$$\omega_2 \to \mu_1(A_{\omega_2}) \text{ ist } \mathcal{A}_2 - \overline{\mathbb{B}}\text{-meßbar.}$$

**Satz 2.60 (Ergänzung zu Satz 2.21):**
Unter den Voraussetzungen von Satz 2.21 gilt für das Produktmaß $\mu = \mu_1 \otimes \mu_2$

$$\forall \, A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : \qquad \mu(A) = \int \mu_2(A_{\omega_1}) \, d\mu_1 = \int \mu_1(A_{\omega_2}) \, d\mu_2.$$

Wir wollen jetzt dies für Integration von Funktionen verallgemeinern.

**Vorbemerkung:**
Für $f : \Omega_1 \times \Omega_2 \to \mathcal{X}$ heißen $f_{\omega_i}$ $\omega_i$-Schnitte von $f$ und sind definiert durch

$$f_{\omega_1} : \Omega_2 \to \mathcal{X}, \quad \omega_2 \mapsto f(\omega_1, \omega_2) \qquad \text{bzw.} \qquad f_{\omega_2} : \Omega_1 \to \mathcal{X}, \quad \omega_1 \mapsto f(\omega_1, \omega_2).$$

**Satz 2.61 (Satz von Fubini für Maße):**
Gegeben seien die $\sigma$-endlichen Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i)$ für $i = 1, 2$ und $f \in \mathcal{M}(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$. Dann gilt

(a) Ist $f \geq 0$, $g_i$ definiert durch $g_i(\omega_i) := \int\limits_{\Omega_j} f_{\omega_i}\, d\mu_j$ $(i \neq j)$, so gilt: $g_i \in \mathcal{M}(\Omega_i, \mathcal{A}_i)$, $g_i \geq 0$ und

$$\int f\, d\mu_1 \otimes \mu_2 = \int g_1\, d\mu_1 = \int g_2\, d\mu_2 \qquad (*)$$

(b) Ist $f$ schon $\mu_1 \otimes \mu_2$-integrierbar, $N_i := \{\omega_i : f_{\omega_i} \text{ nicht } \mu_j\text{-integrierbar}\}$, $g_i$ definiert durch

$$g_i(\omega_i) := I_{N_i^c}(\omega_i) \int\limits_{\Omega_j} f_{\omega_i}\, d\mu_j \qquad (i \neq j),$$

so gilt: $\mu_i(N_i) = 0$, $g_i \in \mathcal{M}(\Omega_i, \mathcal{A}_i)$ ist $\mu_i$-integrierbar mit $(*)$.

**Beispiel 2.62:**
Sei $(\mathbb{R}^2, \mathbb{B}^2, \lambda \otimes \lambda)$, $K = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ und $f = I_K \geq 0$. Dann ist

$$\int f\, d\lambda = \int\int f_x(y)\, d\lambda(y)\, d\lambda(x), \qquad \text{wobei } f_x(y) = \begin{cases} I_{[-\sqrt{1-x^2}, \sqrt{1+x^2}]}(y) & |x| < 1 \\ 0 & \text{sonst} \end{cases}$$

Also ist

$$\int f\, d\lambda^2 = \int\limits_{-1}^{1} \int\limits_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy\, dx = \int\limits_{-1}^{1} 2\sqrt{1-x^2}\, dx = \int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 2\cos^2 z\, dz = \pi$$

mit Hilfe von

$$\cos^2 z = \frac{1}{2} + \frac{1}{2}\cos(2z) \quad \Rightarrow \quad \int \cos^2 z\, dz = \frac{z}{2} + \frac{1}{4} \cdot \sin(2z).$$

**Integration bezüglich des Bildmaßes**:

Generelle Voraussetzungen: $(\Omega, \mathcal{A}, \mu)$ Maßraum, $(\mathcal{X}, \mathcal{B})$ Meßraum und $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B})$. Das Bildmaß ist $\mu^X = \mu \circ X^{-1}$. Dann ist $(\mathcal{X}, \mathcal{B}, \mu^X)$ Maßraum mit $\mu^X(B) = \mu\big(X^{-1}(B)\big)$.

**Satz 2.63 (Transformationsformel):**
Für $f \in \mathcal{M}(\mathcal{X}, \mathcal{B})$, d.h. $f : (\mathcal{X}, \mathcal{B}) \to (\overline{\mathbb{R}}, \overline{\mathbb{B}})$ gilt

(a) Es ist $\int f\, d\mu^X = \int f \circ X\, d\mu$, falls eines der Integrale definiert ist.

(b) $f$ ist genau dann $\mu^X$-quasiintegrierbar, wenn $f \circ X$ $\mu$-quasiintegrierbar ist. Analog ist $f$ genau dann $\mu^X$-integrierbar, wenn $f \circ X$ $\mu$-integrierbar ist.

(c) Insbesondere gilt für $B \in \mathcal{B}$ (Definiertheit vorausgesetzt)

$$\int\limits_B f\, d\mu^X = \int\limits_{X^{-1}(B)} f \circ X\, d\mu.$$

Beweis:

(a) Mit Induktion über meßbare Funktionen (vgl. Konstruktion des $\mu$-Integrals).

**1. Schritt:** Sei $f = I_B$ mit $B \in \mathcal{B}$, dann ist $f \circ X = I_B \circ X = I_{X^{-1}(B)}$ (denn $I_B\big(X(\omega)\big) = 1 \iff X(\omega) \in B \iff \omega \in X^{-1}(B) \iff I_{X^{-1}(B)}(\omega) = 1$). Es folgt

$$\int f \, d\mu^X = \int I_B \, d\mu^X = \mu^X(B) = \mu\big(X^{-1}(B)\big) = \int I_{X^{-1}(B)} \, d\mu = \int f \circ X \, d\mu.$$

**2. Schritt:** Sei $f = \sum \alpha_i I_{B_i}$ mit $\alpha_i \geq 0$ und $B_i \in \mathcal{B}$. Beweis leicht, selbst.

**3. Schritt:** Sei $f_n \uparrow f$ mit $f_n \geq 0$ primitiv. Dann ist $f \circ X = (\lim f_n) \circ X = \lim(f_n \circ X)$ und $f_n \circ X \uparrow f \circ X$. Mit dem Satz über die monotone Konvergenz folgt:

$$\int f \, d\mu^X = \lim \int f_n \, d\mu^X \overset{2.}{=} \lim \int f_n \circ X \, d\mu = \int f \circ X \, d\mu.$$

**4. Schritt:** Sei $f = f^+ - f^-$. Dann ist $f \circ X = (f \circ X)^+ - (f \circ X)^- = f^+ \circ X - f^- \circ X$ und es folgt

$$\int f \, d\mu^X \overset{(*)}{=} \int f^+ \, d\mu^X - \int f^- \, d\mu^X \overset{3.}{=} \int f^+ \circ X \, d\mu - \int f^- \circ \, d\mu \overset{(*)}{=} \int f \circ X \, d\mu,$$

$(*)$ falls definiert.

(b) Ist mit (a) und dem 4. Schritt klar.

(c) Es ist

$$\begin{aligned} \int\limits_B f \, d\mu^X &= \int f \cdot I_B \, d\mu^X \overset{(a)}{=} \int (f \cdot I_B) \circ X \, d\mu = \\ &= \int (f \circ X)(I_B \circ X) \, d\mu = \int (f \circ X) \cdot I_{X^{-1}(B)} \, d\mu = \int\limits_{X^{-1}(B)} f \circ X \, d\mu. \end{aligned}$$

**Definition 2.64:**
Seien $(\Omega, \mathcal{A})$, $(\mathcal{T}, \mathcal{D})$ Meßräume. $T : \Omega \to \mathcal{T}$ heißt bijektiv und bimeßbar (oder ein Meßisomorphismus) genau dann, wenn $T : \Omega \to \mathcal{T}$ bijektiv und $T : (\Omega, \mathcal{A}) \to (\mathcal{T}, \mathcal{D})$ bzw. $T^{-1} : (\mathcal{T}, \mathcal{D}) \to (\Omega, \mathcal{A})$.

**Satz 2.65:**
Seien $\mu, \nu$ Maße auf $(\Omega, \mathcal{A})$ mit $\nu = f\mu$, $f \in \mathcal{M}^+(\Omega, \mathcal{A})$ und $(\mathcal{T}, \mathcal{D})$ Meßraum. Sei weiter $T : \Omega \to \mathcal{T}$ bijektiv bimeßbar. Dann ist

$$\nu^T = (f \circ T^{-1})\mu^T.$$

Beweis: Sei $D \in \mathcal{D}$ beliebig.

$$\begin{aligned} \nu^T(D) &= \int\limits_D d\nu^T \overset{\text{Trafo}}{=} \int\limits_{T^{-1}(D)} d\nu \overset{\text{Vor.}}{=} \int\limits_{T^{-1}(D)} f \, d\mu = \\ &= \int\limits_{T^{-1}(D)} (f \circ T^{-1}) \circ T \, d\mu \overset{\text{Trafo}}{=} \int\limits_D (f \circ T^{-1}) \, d\mu^T \end{aligned}$$

**Beispiel 2.66:**
Voraussetzungen aus Satz 2.65 und $\varphi \in \mathcal{M}^+(\mathcal{T}, \mathcal{D})$ und $D \in \mathcal{D}$. Dann ist

$$\int\limits_D \varphi(t) f\big(T^{-1}(t)\big) \, d\mu^T(t) \overset{2.65}{=} \int\limits_D \varphi(t) \, d\nu^T(t) \overset{\text{Trafo}}{=} \int\limits_{T^{-1}(D)} \varphi\big(T(\omega)\big) \, d\nu(\omega) \overset{2.52}{=} \int\limits_{T^{-1}(D)} \varphi\big(T(\omega)\big) f(\omega) \, d\mu(\omega)$$

Die Anwendung von Satz 2.65 ist insbesondere für $(\mathcal{T},\mathcal{D}) = (\Omega,\mathcal{A})$ und $\mu^T = g\mu$ interessant. Mit der Kettenregel erhält man dann

$$\nu^T = (f\circ T^{-1})\,g\mu \qquad \text{bzw.} \qquad \nu^T(D) = \int f\big(T^{-1}(t)\big)g(t)\;d\mu(t).$$

Für das $n$-dimensionale Lebesgue-Maß $\lambda^n$ auf $(\mathbb{R}^n,\mathbb{B}^n)$ und eine affine Abbildung $T:\mathbb{R}^n\to\mathbb{R}^n$ mit $T(x)=Ax+b$, $A\in\mathbb{R}^{n\times n}$ mit $\det(A)=|A|\neq 0$, $b\in\mathbb{R}^n$, erhält man zum Beispiel

$$\lambda^{nT} = \frac{1}{|A|}\lambda^n \qquad \text{bzw. für alle } B\in\mathbb{B}^n\colon\quad \lambda^n\big(T^{-1}(B)\big) = \frac{1}{|A|}\lambda^n(B).$$

und $\lambda^n\big(T(B)\big) = |A|\cdot\lambda^n(B)$ (siehe Bauer (1990), Satz 8.4, Korollar 8.5).

**Satz 2.67 (Transformationsformel für $\lambda^n$-Integral):**
Für offene Mengen $G,\tilde{G}\subseteq\mathbb{R}^n$, $\varphi: G\to\tilde{G}$ ein Diffeomorphismus (d.h. $\varphi$ bijektiv und $\varphi,\varphi^{-1}$ stetig differenzierbar $\Rightarrow \tilde{G}=\varphi(G)$) und $f\in\mathcal{M}(\tilde{G},\tilde{G}\mathbb{B}^n)$ gilt

$$\begin{aligned} f \text{ ist } \lambda^n\text{-integrierbar über } \tilde{G}=\varphi(G) \quad &\Longleftrightarrow \quad f\circ\varphi|\det D\varphi| \text{ ist } \lambda^n\text{-integrierbar über } G \\ &\Rightarrow \quad \int\limits_{\varphi(G)} f\;d\lambda^n = \int\limits_G f\circ\varphi|\det D\varphi|\;d\lambda^n \end{aligned}$$

Hierbei ist

$$D\varphi = \begin{pmatrix} \frac{\partial\varphi_1(x)}{\partial x_1} & \cdots & \frac{\partial\varphi_1(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial\varphi_n(x)}{\partial x_1} & \cdots & \frac{\partial\varphi_n(x)}{\partial x_n} \end{pmatrix},$$

d.h. $|\det D\varphi|$ ist der Betrag der Determinante der Jakobi-Matrix. Für den Spezialfall $n=1$ sei $G=[a,b]$, $\lambda\to$ Riemann-Integral. Dann ist

$$\int\limits_a^b f\big(\varphi(x)\big)\varphi'(x)\;dx = \int\limits_{\varphi(a)}^{\varphi(b)} f(y)\;dy$$

die bekannte Substitutionsregel.

**Corollar 2.68 (Transformationsformel für $\lambda^n$-Dichten):**
Sei $P$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n,\mathbb{B}^n)$ mit $\lambda^n$-Dichte $f$ und $f=fI_G$ mit $G$ offen. Sei $T:\tilde{G}\to G$ ein Diffeomorphismus. Dann gilt: $P^T$ hat $\lambda^n$-Dichte $f^T$ mit

$$f^T(y) = f\big(T^{-1}(y)\big)\cdot|\det DT^{-1}(y)|\cdot I_{\mathrm{G}}(y).$$

**Beispiel 2.69:**
$X$ sei eine $(\mathbb{R}^n,\mathbb{B}^n)$-wertige Zufallsvariable mit Verteilung $P^X$ und $\lambda^n$-Dichte $f^X=f^XI_U$ mit $U$ offen. Sei $T:\mathbb{R}^n\to\mathbb{R}^n$ mit $T(x)=Ax+b$, $A\in\mathbb{R}^{n\times n}$, $b\in\mathbb{R}^n$ mit $|A|\neq 0$. Sei

$$\varphi := T^{-1}:\mathbb{R}^n\to\mathbb{R}^n, \qquad y\mapsto A^{-1}(y-b) =: \begin{pmatrix}\varphi_1(x)\\ \vdots\\ \varphi_n(x)\end{pmatrix} =: \varphi(y).$$

Dann ist

$$|\det DT^{-1}(y)| = |\det D\varphi(y)| = |\det A^{-1}| = \frac{1}{|\det A|}.$$

Wir erhalten

$$f^{T\circ X}(y) = f^{Ax+b}(y) = f^X\big(A^{-1}(y-b)\big)|\det A^{-1}|I_{AU+b}(y).$$

Einige Spezialfälle:

(a) Sei $n = 1$, $A = \sigma$, $b = \mu$. Dann ist $f^{\sigma T+b}(y) = \frac{1}{\sigma} f^X\left(\frac{y-b}{\sigma}\right)$.

(b) Sei $b = 0$ und $A$ orthonormal (d.h. $A' = A^{-1}$). Dann ist $|A| = 1$ und $f^{AX}(y) = f^X(A^{-1}y) I_{AU}(y)$.

(c) Sei $b = 0$ und $f^X = I_{(0,1)^n}$. Also ist $X$ eine Gleichverteilung auf $(0,1)^n$, d.h. $X \sim G\big((0,1)^n\big)$

$$f^{AX}(y) = f^X(A^{-1}y)\,|\det A^{-1}| I_{A(0,1)^n}(y) = \frac{1}{|\det A|} I_{A(0,1)^n}(y) \underbrace{I_{(0,1)^n}(A^{-1}y)}_{\equiv 1}$$

Also ist $AX \sim G\big(A(0,1)^n\big)$.

## 2.5 $\lambda^n$-stetige Verteilungen: Zusammenhang Verteilungsfunktion / Dichte

**Satz 2.70:**
Ist $P$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathbb{B})$ mit der Verteilungsfunktion $F$ und $f : \mathbb{R}^n \to [0, \infty]$ meßbar, so gilt

$$P = f\lambda^n \quad \Longleftrightarrow \quad \forall\, z \in \mathbb{R}^n : \quad F(z) = \int\limits_{(-\infty 1_n, z]} f \, d\lambda^n.$$

Zur Erinnerung:

$$P = f\lambda^n \quad \Longleftrightarrow \quad \forall\, B \in \mathbb{B}^n : \quad P(B) = \int\limits_B f(z)\, d\lambda^n(z).$$

Der Name Dichte $f(x)$ drückt aus

$$f(x) \overset{\epsilon \text{ klein}}{\approx} \frac{P\big((x - \epsilon 1_n, x + \epsilon 1_n)\big)}{\lambda^n\big((x - \epsilon 1_n, x + \epsilon 1_n)\big)} \approx \frac{P\big((x - \epsilon 1_n, x + \epsilon 1_n)\big)}{(2\epsilon)^n}.$$

Für $n = 1$ folgt

$$F'(x) \approx \frac{F(x + \epsilon) - F(x - \epsilon)}{2\epsilon}.$$

**Bemerkung 2.71 ($n = 1$):**

(a) Sei $P = f\lambda$. Dann ist für alle $z \in \mathbb{R}$ :

$$F(z) = \int\limits_{(-\infty, z]} f(t)\, d\lambda(t) = \int\limits_{-\infty}^{z} f(t)\, dt.$$

Ist zusätzlich $f$ stetig reell, so ergibt der Hauptsatz der Differential- und Integralrechnung, daß $F$ differenzierbar und $f = F'$ ist.

(b) Jede isotone Funktion $F : \mathbb{R} \to \mathbb{R}$ ist $\lambda$-fast sicher differenzierbar mit endlicher Ableitung, d.h. es existiert eine $\lambda$-Nullmenge $N$ mit

$$\forall\, x \in \mathbb{N}^c : \quad \exists\, F'(x) \in [0, \infty).$$

(Hewitt, Stromberg, 1965, S. 264 ff.)

(c) Sei $F$ Verteilungsfunktion eines Wahrscheinlichkeitsmaßes $P|_{\mathbb{B}}$. Dann ist $F'$ $\lambda$-fast sicher definiert, aber $F'$ ist i.a. keine $\lambda$-Dichte von $P$. Beispiele dafür sind die Verteilungsfunktionen der $B\left(1, \frac{1}{2}\right)$- und der Cantor-Verteilung (siehe Beispiel 2.72).

(d) Ist $P = \lambda f$, dann ist $f = F'[\lambda]$. Weiter gilt

$$F' \text{ } \lambda\text{-Dichte von } P \quad \Longleftrightarrow \quad F \text{ absolut stetig}$$

(e) Existiert für alle $x \in \mathbb{R}$ $F'(x)$, dann ist $P = F'\lambda$ (Natanson, 1961, S. 301).

(f) Die Verteilungsfunktion $F$ eines Wahrscheinlichkeitsmaßes $P|_{\mathbb{B}}$ sei $\lambda$-fast überall definiert. Existiert dann die Ableitung $F'$, dann ist $F'$ $\lambda$-integrierbar und für alle $a < b$ gilt

$$\int\limits_{(a,b]} F' \, d\lambda \leq F(b) - F(a).$$

(g) Es sei $F$ die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes $P|_{\mathbb{B}}$ mit

(i) $\int F' \, d\lambda = 1$ oder

(ii) $F$ stetig und stückweise stetig differenzierbar,

dann ist $F'$ $\lambda$-Dichte von $P$.

Beispiel: Sei $F'(x) = x \cdot I_{(0,1)}(x) + (2 - x) \cdot I_{[1,2)}(x)$ ($\Delta$-Verteilung). Dann ist

$$F(x) = \frac{x^2}{2} \cdot I_{(0,1]}(x) + \left(1 - \frac{(x-2)^2}{2}\right) \cdot I_{[1,2)}(x) + I_{[2,\infty)}(x).$$

Sind $X, Y \sim \mathrm{G}\big((0,1)\big)$ stochastisch unabhängig, dann gilt $X + Y \sim F$.

**Beispiel 2.72 (Cantor-Verteilung):**

(a) Es existiert eine Teilmenge $D \subset [0,1]$ mit $D$ überabzählbar, (aber) mit $\lambda(D) = 0$. Die Konstruktion von $D$ geschieht auf folgende Art und Weise: Nehme aus der Menge $[0,1]$ heraus:

$$\begin{array}{cccc} \left(\frac{1}{3}, \frac{2}{3}\right) & & & \\ \left(\frac{1}{9}, \frac{2}{9}\right) & & \left(\frac{7}{9}, \frac{8}{9}\right) & \\ \left(\frac{1}{27}, \frac{2}{27}\right) & \left(\frac{7}{27}, \frac{8}{27}\right) & \left(\frac{19}{27}, \frac{20}{27}\right) & \left(\frac{25}{27}, \frac{26}{27}\right) \quad \text{etc.} \end{array}$$

Dann bleibt $D$ übrig (per Definition). $D$ heißt Cantorsches Diskontinuum. Es ist

$$\lambda(D^c \cap [0,1]) = \frac{1}{3} + \frac{2}{9} + \frac{2^2}{27} + \frac{2^3}{3^4} + \ldots = \frac{\frac{1}{3}}{1 - \frac{2}{3}} = 1$$

und somit ist $\lambda(D) = 0$ aber $|D| = |[0,1]|$.

Beweis mit Hilfe von Trialzahlen $x = \sum\limits_{i=1}^{\infty} \frac{a_i}{3^i}$ mit $a_i \in \{0,1,2\}$. Dann ist

$$\begin{aligned} \left(\frac{1}{3}, \frac{2}{3}\right) &= \{x \in [0,1] : \; a_1 = 1\} \quad \text{oder} \\ \left(\frac{1}{9}, \frac{2}{9}\right) + \left(\frac{7}{9}, \frac{8}{9}\right) &= \{x \in [0,1] : \; a_1 \neq 1, \; a_2 = 1\} \quad \text{usw.} \end{aligned}$$

Dann ist $D = \big\{x : \; \forall \, i : \; a_i \in \{0,2\}\big\}$ und man erhält eine Bijektion $D \leftrightarrow [0,1]$ (Gleichmächtigkeit von Mengen).

(b) (Cantor-Verteilungsfunktion) Sei $x = 0, \alpha_1\alpha_2 \ldots \in [0,1]$ in triadischer Darstellung. Ersetze (immer) $x = 0, \alpha_1\alpha_2 \ldots \alpha_n 1\overline{0}$ durch $x = 0, \alpha_1\alpha_2 \ldots \alpha_n 0\overline{2}$ (daraus folgt die Eindeutigkeit der Darstellung). Definiere

$$k(x) := \begin{cases} \min\{n \in \mathbb{N} : \ \alpha_n = 1\} & x \in [0,1] \cap D^c \\ \infty & x \in D \end{cases}$$

und

$$F_D(x) := \begin{cases} 0 & x < 0 \\ \sum\limits_{k=1}^{k(x)-1} \frac{\alpha_k}{2} \cdot \frac{1}{2^k} + \frac{1}{2^{k(x)}} & x \in [0,1] \cap D^c \\ \sum\limits_{k=1}^{\infty} \frac{\alpha_k}{2} \cdot \frac{1}{2^k} & x \in D \\ 1 & x > 1 \end{cases}$$

d.h. setze $F_D(x)$ gleich $\frac{1}{2}$ auf $\left(\frac{1}{3}, \frac{2}{3}\right)$, $\frac{1}{4}$ auf $\left(\frac{1}{9}, \frac{2}{9}\right)$ und $\frac{3}{4}$ auf $\left(\frac{7}{9}, \frac{8}{9}\right)$ usw. und ergänze auf $D$ zu einer stetigen Verteilungsfunktion. Zeige: $F_D$ ist isoton, stetig und differenzierbar auf $D^c$ mit $F_D'(x) = 0$. $F_D$ ist also eine Verteilungsfunktion. Die zugehörige Wahrscheinlichkeitsverteilung $P_D$ heißt Cantor-Verteilung. Es gilt

$$\begin{aligned} P_D([0,1] \cap D^c) &= P_D\left(\left(\frac{1}{3}, \frac{2}{3}\right)\right) + P_D\left(\left(\frac{1}{9}, \frac{2}{9}\right)\right) + P_D\left(\left(\frac{7}{9}, \frac{8}{9}\right)\right) + \ldots = \\ &= F_D\left(\frac{2}{3}^-\right) - F_D\left(\frac{1}{3}\right) + F_D\left(\frac{2}{9}^-\right) - F_D\left(\frac{1}{9}\right) + \ldots = \\ &= \frac{1}{2} - \frac{1}{2} + \frac{1}{4} - \frac{1}{4} + \ldots = 0 \end{aligned}$$

und damit $P_D(D) = 1$, aber $\lambda(D) = 0$.

## 2.6 Erwartungswert, Varianz, Kovarianz, Momente und Ungleichungen

**Definition 2.73:**
Sei $X$ eine $\overline{\mathbb{R}}$-wertige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$. Ist $X$ quasiintegrierbar, so heißt

$$EX := \int X \, dP$$

Erwartungswert von $X$ (bzgl. $P$). Ist $EX \in \mathbb{R}$, so spricht man von einem endlichen Erwartungswert ($\iff X$ $P$-integrierbar).

**Satz 2.74 (Grundlegende Eigenschaften von Erwartungswerten):**
Seien $X, Y$ numerische Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$, $\alpha \in \mathbb{R}$, $A \in \mathcal{A}$. Dann gelten folgende Aussagen:

(a) Ist $X \geq 0 \Rightarrow X$ quasiintegrierbar $\iff EX$ ist definiert,

(b) $X$ quasiintegrierbar $\iff EX^+ < \infty$ oder $EX^- < \infty \Rightarrow EX = EX^+ - EX^-$,

(c) $X$ quasiintegrierbar $\Rightarrow |EX| \leq E|X|$,

(d) $X, Y$ quasiintegrierbar und $X \leq Y \Rightarrow EX \leq EY$,

(e) $X, Y$ quasiintegrierbar $\Rightarrow E(\alpha X + Y) = \alpha EX + EY$, falls $\alpha X + Y$ und $\alpha EX + EY$ definiert sind,

(f) $X = Y[P]$ und $X$ quasiintegrierbar $\Rightarrow EX = EY$,

(g) $|X| \leq Y$, $EY \in \mathbb{R} \Rightarrow E|X| \in \mathbb{R} \iff EX \in \mathbb{R}$.

Beweis: Einfaches Umschreiben der Eigenschaften des $P$-Integrals.

Die Transformationsformel ergibt folgenden Satz:

**Satz 2.75:**
Gegeben sei eine $(\mathcal{X},\mathcal{B})$-wertige Zufallsvariable $X$ auf $(\Omega,\mathcal{A},P)$ und $g:(\mathcal{X},\mathcal{B})\to(\overline{\mathbb{R}},\overline{\mathbb{B}})$. Dann gelten

(a) $g\circ X$ $P$-quasiintegrierbar $\iff$ $g$ $P^X$-quasiintegrierbar, d.h.

$$E(g\circ X)=\int g\big(X(\omega)\big)\,dP(\omega)=\int g(x)\,dP^X(x).$$

Ist speziell $(\mathcal{X},\mathcal{B})=(\overline{\mathbb{R}},\overline{\mathbb{B}})$, d.h. $g=\mathrm{id}_{\overline{\mathbb{R}}}$ quasiintegrierbar, dann ist $\int X\,dP=\int x\,dP^X(x)$.

(b) Hat $X$ die $\mu$-Dichte $f$ bei $\sigma$-endlichem $\mu|_{\mathcal{B}}$, so gilt

$$\begin{aligned} g\circ X \text{ } P\text{-quasiintegrierbar} &\iff gf \text{ } \mu\text{-quasiintegrierbar}\\ &\Rightarrow \int g\circ X\,dP=\int gf\,d\mu \end{aligned}$$

Ist speziell $\mu$ Zählmaß einer abzählbaren Menge $J\subseteq\mathcal{X}$, so ist

$$E(g\circ X)=\sum_{i\in J}g(i)P(X=i).$$

**Beispiel 2.76:**

(a) (Normalverteilung) $\varphi$ mit $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ heißt Dichte der Standard-Normalverteilung, kurz $\mathrm{N}(0,1)$-Verteilung. Ihre Verteilungsfunktion wird mit $\Phi$ bezeichnet:

$$\Phi(x)=\int_{\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt.$$

Ist $X\sim\mathrm{N}(0,1)$, so gilt

$$EX=\int x\,dP^X(x)=\int_{-\infty}^{\infty}x\varphi(x)\,dx=0,$$

denn

$$-\int_{-\infty}^{0}x\varphi(x)\,dx=\int_{0}^{\infty}x\varphi(x)\,dx=\left[-e^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}\right]_0^{\infty}=\frac{1}{\sqrt{2\pi}}.$$

Ist $X\sim\mathrm{N}(0,1)$, $Y:=\sigma X+\mu$, $\sigma>0$, so heißt $Y$ $\mathrm{N}(\mu,\sigma^2)$-Verteilung. Die $\lambda^1$-Dichte ist

$$f^Y(y)=\frac{1}{\sqrt{2\pi}\sigma}\cdot\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Mit Satz 2.74 (e) folgt

$$EY=E(\sigma X+\mu)=\sigma\underbrace{EX}_{=0}+\mu=\mu.$$

(b) (Cauchy-Verteilung) Sei $X\sim G\left(\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\right)$. Dann ist

$$f^X(x)=\frac{1}{\pi}\cdot I_{\left(-\frac{\pi}{2},\frac{\pi}{2}\right)}(x)$$

$\lambda^1$-Dichte von $P^X$. Zum Beispiel haben wir Lichtquellen $Q$ in einer Wand. Der Abstrahlwinkel soll zufällig gewählt werden (ist also gleichverteilt). Wir erhalten zu $X(\omega) = x$ eine neue Zufallsvariable $Y := \tan X$. Somit gilt für $y \in \mathbb{R}$

$$\begin{aligned} F^Y(y) &= P(Y \le y) = P(\tan X \le y) = P(X \le \arctan y) = \\ &= F^X(\arctan y) = \frac{1}{2} + \frac{1}{\pi} \cdot \arctan y \end{aligned}$$

Die Lebesgue-Dichte von $P^Y$ ist $f^Y(y) = \dfrac{1}{\pi(1+y^2)}$. Dies ist die Dichte der Cauchy-Verteilung $C(0,1)$. Aber der Erwartungswert von $Y$ existiert nicht, denn für $N > 1$ ist

$$\begin{aligned} \int_0^N y f^Y(y)\, dy &= \frac{1}{\pi} \int_0^N \frac{1}{1+y^2}\, dy > \frac{1}{\pi} \int_1^N \frac{1}{1+y}\, dy = \\ &= \frac{1}{\pi} \left[\ln(1+y)\right]_1^N \overset{N\to\infty}{\longrightarrow} \infty \end{aligned}$$

(c) (Poisson-Verteilung) Sei $X \sim \mathrm{Po}(\lambda)$ mit $\lambda > 0$. $\kappa$ sei Zählmaß auf $\mathbb{N}_0$. Dann ist

$$EX = \int x e^{-\lambda} \frac{\lambda^x}{x!}\, d\kappa(x) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda \underbrace{\sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!}}_{=1} = \lambda.$$

**Definition 2.77:**
Ist $X$ numerische $P$-integrierbare Zufallsvariable (d.h. $EX \in \mathbb{R} \iff E|X| \in \mathbb{R}$), so heißt

$$\mathrm{Var}\, X := E(X - EX)^2$$

Varianz von $X$. $\sqrt{\mathrm{Var}\, X}$ heißt Standardabweichung.

**Satz 2.78:**
Für eine numerische $P$-integrierbare Zufallsvariable $X$ und $\alpha, \beta \in \mathbb{R}$ gelten:

(a) $\mathrm{Var}\, X = EX^2 - (EX)^2 = E(X-\alpha)^2 - (EX-\alpha)^2$,

(b) $\mathrm{Var}(\alpha X + \beta) = \alpha^2 \cdot \mathrm{Var}\, X$,

(c) $\mathrm{Var}\, X < \infty \iff EX^2 < \infty$ ( $\iff : X$ $P$-quadratintegrierbar),

(d) $\mathrm{Var}\, X = 0 \iff X = EX[P]$.

Beweis:

(a) $\mathrm{Var}\, X = E(X-EX)^2 = E(X^2 - 2X \cdot EX + (EX)^2) \overset{\text{Satz}}{=} EX^2 - 2EX \cdot EX + (EX)^2 = EX^2 - (EX)^2$. Den Rest selbst beweisen.

(b) Es ist

$$\begin{aligned} \mathrm{Var}(\alpha X + \beta) &\overset{(a)}{=} E(\alpha X + \beta - \beta)^2 - (E(\alpha X + \beta - \beta))^2 = \\ &= \alpha^2 EX^2 - \alpha^2 (EX)^2 = \alpha^2 \cdot \mathrm{Var}\, X \end{aligned}$$

(c) Mit $\mathrm{Var}\, X = EX^2 - (EX)^2$ und $EX < \infty$ nach Obervoraussetzung folgt die Behauptung.

(d) Es ist $0 = E(X - EX)^2 \iff (X - EX)^2 = 0[P] \iff X = EX[P]$.

**Beispiel 2.76 - Fortsetzung:**

(a) Sei $X \sim \mathrm{N}(\mu, \sigma^2)$, dann ist $EX = \mu$ und

$$\begin{aligned}\operatorname{Var} X &= E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2}\, dy = \frac{4\sigma^2}{\sqrt{\pi}} \int_0^{\infty} y^2 e^{-y^2}\, dy = \\ &= \frac{4\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \int_0^{\infty} e^{-t} t^{\frac{3}{2}-1}\, dt = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = \sigma^2\end{aligned}$$

mit $\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1}\, dt$ und den Substitutionen

$$y = \frac{x-\mu}{\sqrt{2}\sigma}, \quad dy = \frac{1}{\sqrt{2}\sigma}\, dx \qquad \text{und} \qquad t = y^2, \quad dt = 2y\, dy.$$

(c) Sei $X \sim \mathrm{Po}(\lambda)$. Dann ist $EX = \lambda$ und

$$\operatorname{Var} X = EX^2 - (EX)^2 = E(X(X-1)) + EX - (EX)^2$$

sowie

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) e^{-\lambda} \frac{\lambda^x}{x!} = \lambda^2 \underbrace{\sum_{x=2}^{\infty} e^{-\lambda} \frac{\lambda^{x-2}}{(x-2)!}}_{=1} = \lambda^2$$

Damit ergibt sich $\operatorname{Var} X = \lambda^2 + \lambda - \lambda^2 = \lambda$. Also ist bei der Poisson-Verteilung $EX = \operatorname{Var} X$.

**Definition 2.79:**
$X$ sei numerische Zufallsvariable und $n \in \mathbb{N}$. Dann heißt, falls die jeweiligen Integrale definiert sind:

(a) $EX^n$ $n$-tes Moment von $X$,

(b) $E|X|^n$ $n$-tes absolutes Moment von $X$ und

(c) $E(X - EX)^n$ $n$-tes zentrales Moment von $X$.

**Satz 2.80 (Multiplikationssatz für Erwartungswerte):**
Für stochastisch unabhängige Zufallsvariablen $X, Y$ gilt

$$EXY = EX \cdot EY,$$

falls eine der folgenden Bedingungen erfüllt ist:

(a) $X \geq 0$ und $Y \geq 0$,

(b) $E|X| < \infty$ und $E|Y| < \infty$ ($\iff$ $X, Y$ $P$-integrierbar).

Beweis: Mit dem Satz von Fubini erhält man wegen $P^{(X,Y)} = P^X \otimes P^Y$ (da $X, Y$ stochastisch unabhängig)

$$\begin{aligned}EXY &= \int XY\, dP = \int xy\, dP^X(x) \otimes P^Y(y) = \\ &= \int\int xy\, dP^X(x)\, dP^Y(y) = \int y \left(\int x\, dP^X(x)\right) dP^Y(y) = EX \cdot EY.\end{aligned}$$

**Definition 2.81:**
Seien $X, Y$ $P$-integrierbare numerische Zufallsvariablen mit definiertem $EXY$. Dann heißt

$$\mathrm{Cov}(X,Y) := E(X-EX)(Y-EY)$$

Kovarianz von $X$ und $Y$.

**Satz 2.82:**
Seien $X, Y$ $P$-integrierbare numerische Zufallsvariablen mit definiertem $EXY$. Dann gelten

(a) $\mathrm{Cov}(X,Y) = EXY - EX \cdot EY$,

(b) $\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2 \cdot \mathrm{Cov}(X,Y)$, falls $X+Y$ und die rechte Summe definiert sind,

(c) $X, Y$ stochastisch unabhängig $\Rightarrow \mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y$, falls $X+Y$ definiert ist,

(d) $\mathrm{Cov}(X,X) = \mathrm{Var}\, X$ und $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$,

(e) $\mathrm{Cov}(\alpha X + Y, Z) = \alpha \cdot \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$.

Beweis: Elementares Nachrechnen, zum Beispiel für

(c) Sind $X, Y$ stochastisch unabhängig, dann sind auch $(X-EX)$ und $(Y-EY)$ stochastisch unabhängig. Somit ist

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E(X-EX+Y-EY)^2 = \\
&= E(X-EX)^2 + E(Y-EY)^2 + 2E(X-EX)(Y-EY) \overset{2.80}{=} \\
&= \mathrm{Var}\, X + \mathrm{Var}\, Y + 2 \cdot \underbrace{E(X-EX)}_{=0} \cdot \underbrace{E(Y-EY)}_{=0}
\end{aligned}$$

**Beispiel 2.83 (2-dimensionale Normalverteilung):**
Seien $X, Y \sim \mathrm{N}(0,1)$ stochastisch unabhängig mit $Z := \binom{X}{Y} \sim P^X \otimes P^Y$. Dann ist $\binom{X}{Y}$ 2-dimensional standard-normalverteilt ($\sim \mathrm{N}(0, I_2)$). Sei $A \in \mathbb{R}^{2\times 2}$ mit $|A| \neq 0$ sowie $\mu \in \mathbb{R}^2$. Dann ist $AZ + \mu \sim \mathrm{N}(\mu, AA^T)$ mit $\lambda^2$-Dichte

$$f^{AZ+\mu}(\omega) = \frac{1}{2\pi |\Sigma|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\omega-\mu)^T \Sigma^{-1} (\omega-\mu) \right\}$$

(ergibt sich aus Beispiel 2.69) mit $\Sigma := AA^T > 0$, d.h. $\Sigma$ ist positiv definit. Sei nun $\Sigma$ wie folgt gegeben

$$\Sigma := \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} > 0$$

mit $\sigma_1, \sigma_2 > 0$ und $\rho \in (-1,1)$. Sei weiter $\mu \in \mathbb{R}^2$, $Z := \binom{Z_1}{Z_2} \sim \mathrm{N}(\mu, \Sigma)$. Dann ist

$$|\Sigma|^{\frac{1}{2}} = \sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}} \qquad \text{und} \qquad \Sigma^{-1} = \frac{1}{|\Sigma|} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

Dann ist

$$f^Z(z) = \frac{\exp\left\{ -\frac{1}{2} \cdot \frac{1}{1-\rho^2} \left[ \left( \frac{z_1-\mu_1}{\sigma_1} \right)^2 - 2\rho \frac{z_1-\mu_1}{\sigma_1} \frac{z_2-\mu_2}{\sigma_2} + \left( \frac{z_2-\mu_2}{\sigma_2} \right)^2 \right] \right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

Es ist $Z_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Z_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ durch Berechnung der Randdichten

$$f^{Z_i}(z_i) = \int f^Z(z_1, z_2)\, dz_j \qquad (j \neq i).$$

Dann ist

$$\begin{aligned}
\mathrm{Cov}(Z_1,Z_2) &= E(Z_1-\mu_1)(Z_2-\mu_2) = \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(z_1-\mu_1)(z_2-\mu_2)f^Z(z_1,z_2)\,dz_1\,dz_2 \overset{\text{Subst.}}{=} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z_1\cdot z_2\cdot\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot \\
&\quad \cdot\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{z_1}{\sigma_1}\right)^2-2\rho\frac{z_1z_2}{\sigma_1\sigma_2}+\left(\frac{z_2}{\sigma_2}\right)^2\right]\right\}\,dz_1\,dz_2 = \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{z_2}{\sqrt{2\pi}\sigma_2}\cdot\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{z_2}{\sigma_2}\right)^2-\left(\frac{\rho}{\sigma_2}\right)^2 z_2^2\right]\right\}\cdot \\
&\quad \cdot\underbrace{\left[\int_{-\infty}^{\infty}\frac{z_1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\cdot\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left(z_1-\rho\frac{\sigma_1}{\sigma_2}z_2\right)^2\right\}\,dz_1\right]}_{(*)}\,dz_2 = \\
&= \underbrace{\rho\frac{\sigma_1}{\sigma_2}\int_{-\infty}^{\infty} z_2^2\frac{1}{\sqrt{2\pi}\sigma_2}\cdot\exp\left\{-\frac{1}{2}\cdot\frac{z_2^2}{\sigma_2^2}\right\}\,dz_2}_{(**)} = \rho\sigma_1\sigma_2
\end{aligned}$$

mit

$(*)$ $= \rho\frac{\sigma_1}{\sigma_2}z_2$ Erwartungswert einer Zufallsvariable $W\sim N\left(\rho\frac{\sigma_1}{\sigma_2}z_2,\sigma_1^2(1-\rho^2)\right)$,

$(**)$ $=$ Var $V=\sigma_2^2$ einer Zufallsvariable $V\sim \mathrm{N}(0,\sigma_2^2)$.

Spezialfall: $\sigma_1^2=\sigma_2^2=1,\ \rho\in(-1,1)$ und $\mu_1=\mu_2=0$:

$$f(z)=\frac{1}{2\pi\sqrt{1-\rho^2}}\cdot\exp\left\{-\frac{1}{2}\cdot\frac{1}{1-\rho^2}\left[z_1^2-2\rho z_1z_2+z_2^2\right]\right\}.$$

Ist $f(z)\equiv$ const, dann entspricht dies den Höhenlinien.

Im folgenden werden die wichtigsten Ungleichungen für Erwartungswerte behandelt.

**Vorbemerkung:**
Zwei Zahlen $p,q\in[1,\infty]$ mit $\frac{1}{p}+\frac{1}{q}=1$ heißen konjugierte Exponenten, zum Beispiel für $p=q=2$ oder $p=1,\ q=\infty$.

**Satz 2.84:**
Seien $p,q\in(0,\infty)$ mit $\frac{1}{p}+\frac{1}{q}=1$. Dann gelten für numerische Zufallsvariablen $X,Y$ auf $(\Omega,\mathcal{A},P)$

(a) die Hölder-Ungleichung
$$E|XY|\le(E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}},$$

(b) die Cauchy-Ungleichung (Teil (a) für $p=q=2$),

(c) die Minkowsky-Ungleichung für $p\in[1,\infty)$, falls $X+Y$ definiert ist,
$$(E|X+Y|^p)^{\frac{1}{p}}\le(E|X|^p)^{\frac{1}{p}}+(E|Y|^p)^{\frac{1}{p}}.$$

Beweis:

(a) Sei $A = (E|X|^p)^{\frac{1}{p}}$ und $B = (E|Y|^q)^{\frac{1}{q}}$.

(i) Falls $A = \infty$ oder $B = \infty$, so ist auch die rechte Seite unendlich.

(ii) Falls $A = 0 \quad \Rightarrow \quad X = 0[P] \quad \Rightarrow \quad$ Beh. ($B = 0$ analog).

(iii) Sei $0 < A, B < \infty$. Definiere $V := \frac{|X|}{A}$ und $W := \frac{|Y|}{B}$. Dann gilt: $V, W \in \mathcal{M}$ und $V, W \geq 0$. Es ist $EV^p = EW^q = 1$.

($\alpha$) Sei $0 < V(\omega), W(\omega) < \infty$: Dann existieren $s, t \in \mathbb{R}$ mit $V(\omega) = e^{\frac{s}{p}}$ und $W(\omega) = e^{\frac{t}{q}}$. Es folgt

$$\begin{aligned} V(\omega)W(\omega) &= e^{\frac{s}{p}} e^{\frac{t}{q}} = \exp\left\{\frac{1}{p}s + \left(1 - \frac{1}{p}\right)t\right\} \overset{\text{e konvex}}{\leq} \\ &\leq \frac{1}{p}e^s + \left(1 - \frac{1}{p}\right)e^t = \frac{1}{p}V^p(\omega) + \frac{1}{q}W^q(\omega) \qquad (*) \end{aligned}$$

($\beta$) Für $V(\omega)$ oder $W(\omega) \in \{0, \infty\}$ gilt $(*)$ ebenfalls.

Zusammen ergibt die Ungleichung

$$\frac{\int |XY|\, dP}{(E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}}} = \int VW\, dP \leq \frac{1}{p}\underbrace{EV^p}_{=1} + \frac{1}{q}\underbrace{EW^q}_{=1} = 1$$

die Behauptung.

(c) Der Fall $p = 1$ ist klar:

$$\int |X + Y|\, dP \leq \int (|X| + |Y|)\, dP = \int |X|\, dP + \int |Y|\, dP.$$

Für $1 < p < \infty$ gilt

$$\begin{aligned} \int |X+Y|^p\, dP &= \int |X+Y|^{p-1}|X+Y|\, dP \leq \int |X+Y|^{p-1}|X|\, dP + \int |X+Y|^{p-1}|Y|\, dP \leq \\ &\overset{(*)}{\leq} \left(\int |X+Y|^p\, dP\right)^{\frac{p-1}{p}} \left(\int |X|^p\, dP\right)^{\frac{1}{p}} + \\ &\quad + \left(\int |X+Y|^p\, dP\right)^{\frac{p-1}{p}} \left(\int |Y|^p\, dP\right)^{\frac{1}{p}} = \\ &= \left(\int |X+Y|^p\, dP\right)^{\frac{p-1}{p}} \cdot \left\{\left(\int |X|^p\, dP\right)^{\frac{1}{p}} + \left(\int |Y|^p\, dP\right)^{\frac{1}{p}}\right\}. \end{aligned}$$

$(*)$ Hölder-Ungleichung angewandt mit $\frac{1}{q} = 1 - \frac{1}{p} = \frac{p-1}{p}$.

Es folgt die Behauptung.

**Bemerkung 2.85:**

(a) In der Hölder-Ungleichung gilt Gleichheit genau dann, wenn

$$\exists\, \alpha, \beta \geq 0 : \qquad \alpha + \beta > 0 \text{ und } \alpha|X|^p + \beta|Y|^q[P].$$

(b) Definiert man das wesentliche Supremum von $|X|$ vermöge

$$P\text{-sup}\,|X| := \|X\|_\infty := \inf\left\{r \in \overline{\mathbb{R}} :\ P(|X| > r) = 0\right\},$$

so gilt die Hölder-Ungleichung auch für $p, q \in \{1, \infty\}$:

$$E|XY| \leq \|X\|_\infty \cdot E|Y|.$$

Die Minkowsky-Ungleichung gilt ebenfalls für $p = \infty$:

$$\|X + Y\|_\infty \leq \|X\|_\infty + \|Y\|_\infty.$$

(c) Die Ungleichungen gelten sogar für beliebige Maße $\mu$.

(d) Für alle $1 \leq p < q < \infty$ gilt

$$(E|X|^p)^{\frac{1}{p}} \leq (E|X|^q)^{\frac{1}{q}},$$

denn mit $q < \infty$, $r = \frac{p}{q} > 1$ folgt mit Hölder

$$E|X|^p = E\left(|X|^p \cdot 1\right) \leq (E|X|^{rp})^{\frac{1}{r}} \cdot 1 = (E|X|^q)^{\frac{p}{q}}.$$

Der Fall $q = \infty$ ist ebenfalls klar.

(e) Aus (d) folgt zum Beispiel

$$E|X|^2 < \infty \quad \Rightarrow \quad E|X| < \infty$$

und auch $(E|X|)^2 \leq E|X|^2$ (vgl. mit Var $X \geq 0 \iff EX^2 \geq (EX)^2$).

**Satz 2.86 (Jensen-Ungleichung):**
$X$ sei eine integrierbare reelle Zufallsvariable mit Werten in einem offenen Intervall $I \subseteq \mathbb{R}$ und $h : (I, I\mathbb{B}) \to (\mathbb{R}, \mathbb{B})$ konvex. Dann folgt

(a) $Eh(X) \in I$ und

(b) $h(EX) \leq Eh(X)$.

Beweis:

(a) Ist $\alpha = \sup\{x \mid x \in I\} < \infty$, so ist zu zeigen, daß $EX < \alpha$. Aus $X(\omega) < \alpha$ für alle $\omega \in \Omega$ folgt $EX \leq \alpha$. Wäre $EX = \alpha$, so ist $E(\alpha - X) = 0$. Also, da $\alpha - X(\omega) > 0$ für alle $\omega \in \Omega$, müßte $\alpha - X = 0[P]$ sein. Widerspruch zu $\alpha - X(\omega) > 0$ für alle $\omega \in \Omega$. Also gilt $EX < \alpha$.

(b) (Beweisskizze)

(i) Es existiert eine Familie $(g_j)$, $j \in J$, so daß für alle $y \in I$ gilt

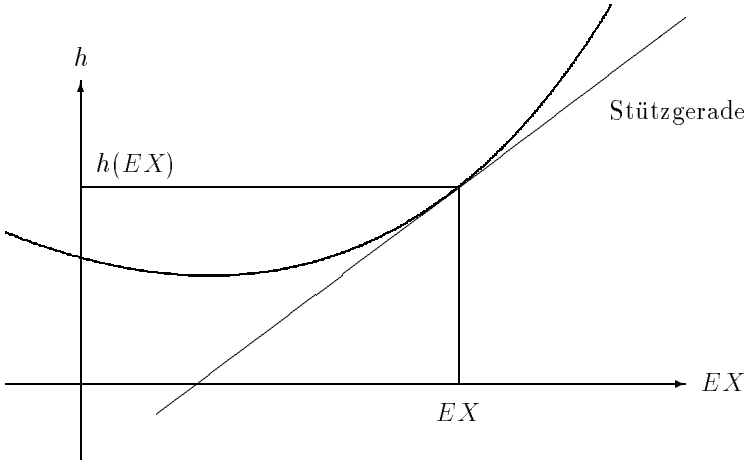
$$h(y) = \sup\{g_j(y) \mid j \in J\}.$$

(ii) Für alle $y \in I$ existiert ein $j \in J$, so daß

$$\exists\, a, b \in \mathbb{R} : \quad h(y) = g_j(y) = ay + b.$$

Es folgt

$$g_j(EX) = aEX + b = h(EX) \qquad \text{und} \qquad g_j(X(\omega)) \leq h(X(\omega)) \qquad \forall\, \omega \in \Omega.$$




Mit $h(EX) = Eg_j(X) \leq Eh(X)$ folgt die Behauptung.

**Bemerkung 2.87:**

(a) Für $h(x) = x^2$ gilt
$$h(EX) = (EX)^2 \leq EX^2 = Eh(X).$$

(b) Für $p > 1$, $h(x) = |x|^p$ gilt
$$h(E|X|) = (E|X|)^p \leq E|X|^p = Eh(X),$$
d.h. aus $E|X|^p < \infty$ folgt $E|X| < \infty$.

(c) Für $h(x) = \frac{1}{x}$ und $x > 0$ folgt $E\left(\frac{1}{X}\right) \leq \frac{1}{EX}$.

(d) Für $h(x) = e^x$ gilt $\exp(EX) \leq E\exp(X)$. Sei $X$ diskret verteilt mit $P(X_i = x_i) = \alpha_i$ für $i = 1, \ldots, n$ mit $\sum_{i=1}^{n} \alpha_i = 1$ und existierendem Erwartungswert. Dann gilt
$$\exp(EX) = \exp\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i \exp(x_i).$$
Mit $y_i = \exp(x_i)$ folgt die Ungleichung zwischen geometrischem und arithmetischem Mittel für $\alpha_i = \frac{1}{n}$
$$\prod_{i=1}^{n} y_i^{\alpha_i} \leq \sum_{i=1}^{n} \alpha_i y_i.$$

**Satz 2.88:**
Sei $X$ eine numerische Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $p, \epsilon > 0$. Dann gilt die Tschebyschev-Markov-Ungleichung
$$P\big(|X| \geq \epsilon\big) \leq \frac{1}{\epsilon^p} E|X|^p.$$
Für $p = 2$ ist dies die sogenannte Tschebyschev-Ungleichung. Für eine $P$-integrierbare Zufallsvariable $X$ erhält man
$$P\big(|X - EX| \geq \epsilon\big) \leq \frac{1}{\epsilon^2} \mathrm{Var}\, X,$$
bzw. für $0 < \mathrm{Var}\, X < \infty$ gilt
$$P\left(\frac{|X - EX|}{\sqrt{\mathrm{Var}\, X}} \geq \epsilon\right) \leq \frac{1}{\epsilon^2}.$$

Beweis: Es ist $B_\epsilon := \{|X| \geq \epsilon\} \in \mathcal{A}$ und

$$\int |X|^p \, dP \geq \int\limits_{B_\epsilon} |X|^p \, dP \geq \epsilon^p \int\limits_{B_\epsilon} dP = \epsilon^p \cdot P(|X| \geq \epsilon).$$

**Bemerkung 2.89:**

(a) Ist $X$ $P$-integrierbar mit $EX = 0$ und Var $X = 1$, so erhält man

$$P(|X| \geq 2) \leq \frac{1}{4}, \qquad P(|X| \geq 3) \leq \frac{1}{9}, \qquad P(|X| \leq 3) \geq 0.\overline{8}.$$

Für $EX = \mu$ und Var $X = \sigma^2$ mit $\sigma > 0$ erhält man

$$P(|X - EX| \geq k\sigma) \leq \frac{1}{k^2}, \text{ z.B. für } k = 3\text{: } P(|X - EX| \leq 3\sigma) \geq 0.\overline{8}.$$

Man sagt hierzu $3\sigma$-Bereich um $EX$. Ist $X \sim \mathrm{N}(0,1)$, dann ist

$$P(|X| \leq 1) \approx 0.68269, \quad P(|X| \leq 2) \approx 0.9545, \quad P(|X| \leq 3) \approx 0.9973$$

(b) (Verallgemeinerung von Satz 2.88) Sei $g$ eine isotone Funktion, $g : [0, \infty) \to [0, \infty)$, so folgt

$$P(|X| \geq \epsilon) \leq \frac{1}{g(\epsilon)} Eg(|X|).$$

(c) Sind $X_1, \ldots, X_n$ unkorreliert, d.h. $\mathrm{Cov}(X_i, X_j) = 0$ für alle $i \neq j$, dann gilt für alle $\epsilon > 0$

$$P\left(\left|\sum_{i=1}^n (X_i - EX_i)\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{i=1}^n \mathrm{Var}\, X_i.$$

Gilt zusätzlich $EX_i = \mu$ und Var $X_i = \sigma^2$ für alle $i$, dann gilt für alle $\epsilon > 0$:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right) \leq \frac{1}{(n\epsilon)^2} n\sigma^2 = \frac{\sigma^2}{n\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0,$$

d.h. $\frac{1}{n}\sum\limits_{i=1}^n X_i =: \overline{X}.$ „strebt“ mit wachsendem $n$ gegen $\mu = EX_i$ (siehe Kapitel 3: Schwaches Gesetzt der großen Zahlen). Wegen

$$E\overline{X}. = E\frac{1}{n}\sum X_i = \frac{1}{n}\sum EX_i = \frac{1}{n}\sum \mu = \mu$$

ist $\hat{\mu} := \overline{X}.$ ein „erwartungstreuer“ Schätzer für $n$ (siehe Kapitel 4).

## 2.7 Faltungsformel und spezielle Verteilungen

Es ist bereits bekannt:

$$X_1, \ldots, X_n \sim \mathrm{B}(1,p) \text{ stochastisch unabhängig } \Rightarrow \sum_{i=1}^n X_i \sim \mathrm{B}(n,p).$$

Allgemeine Frage: Sind $X_1, \ldots, X_n \sim P^{X_1}$ stochastisch unabhängig, wie ist dann $\sum\limits_{i=1}^n X_i$ verteilt?

**Definition 2.90:**
Sind $X_1, X_2$ $\mathbb{R}^n$-wertige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$, so heißt die Verteilung von $X_1 + X_2 =: Y$ Faltung von $P^{X_1}$ und $P^{X_2}$ genannt. Sie wird mit $P^{X_1} * P^{X_2}$ bezeichnet, d.h.

$$P^{X_1} * P^{X_2}(B) = P^{(X_1,X_2)}\big(\{(x_1,x_2) \in \mathbb{R}^{2n} : \; x_1 + x_2 \in B\}\big) \qquad \forall\, B \in \mathbb{B}^n.$$

**Satz 2.91:**
Unter den Voraussetzungen von Definition 2.90 gilt:

(a) Für diskrete Zufallsvariablen $X_1, X_2$ gilt

$$\forall\, z \in \mathbb{R}^n : \quad P^{X_1+X_2}(\{z\}) = P^{X_1} * P^{X_2}(\{z\}) = \sum_{x \in \mathbb{R}^n} P^{X_1}(\{x\}) P^{X_2}(\{z-x\}).$$

(b) Für $\lambda^n$-stetige stochastisch unabhängige Zufallsvariablen $X_1, X_2$ mit $\lambda^n$-Dichten $f, g$ gilt: $P^{X_1} * P^{X_2}$ hat die $\lambda^n$-Dichte $h$ definiert durch

$$\forall\, z \in \mathbb{R}^n : \quad h(z) = \int\limits_{\mathbb{R}^n} f(x) g(z-x)\, d\lambda^n(x) =: f * g(z).$$

Beweis: Es ist

$$k(x_1,x_2) = x_1 + x_2 \quad \Longleftrightarrow \quad k^{-1}(\{z\}) = \{(x_1,x_2) : \; x_1 + x_2 = z\}.$$

Also gilt

$$\begin{aligned} P^{X_1+X_2}(B) &= P^{k(X_1,X_2)}(B) = P^{(X_1,X_2)}\big(k^{-1}(B)\big) = \int I_{k^{-1}(B)}(x_1,x_2)\, dP^{(X_1,X_2)}(x_1,x_2) \overset{\text{Fubini, } X_i \text{ stu.}}{=} \\ &= \int\limits_{\mathbb{R}^n} \underbrace{\left[ \int\limits_{\mathbb{R}^n} I_{k^{-1}(B)}(x_1,x_2)\, dP^{X_2}(x_2) \right]}_{= P^{X_2}(B - x_1)} dP^{X_1}(x_1) = \int\limits_{\mathbb{R}^n} P^{X_2}(B - x_1)\, dP^{X_1}(x_1) \qquad (*) \end{aligned}$$

(a) Sei $B = \{z\}$. Dann ist

$$(*) = \sum_{x \in \mathbb{R}^n} g(z - x_1) f(x_1) = \sum_{x \in \mathbb{R}^n} P^{X_2}(\{z-x\}) P^{X_1}(\{x\}).$$

(b) In diesem Fall ist

$$(*) \quad = \quad \int\limits_{\mathbb{R}^n} \left[ \int\limits_{B-x} g(z) \cdot f(x)\, d\lambda^n(z) \right] d\lambda^n(x) \overset{\text{Fub., Trafo}}{=} = \int\limits_B \underbrace{\left[ \int\limits_{\mathbb{R}^n} g(z-x) f(x)\, d\lambda^n(x) \right]}_{=: h(z)} d\lambda^n(z).$$

**Bemerkung 2.92:**
Seien $X_i$ reellwertige Zufallsvariablen mit den Verteilungsfunktionen $F_i$ $(i = 1,2)$, $\mu_i$-Dichte $f_i$. Weiter seien $X_1, X_2$ stochastisch unabhängig. Dann ist die zugehörige Verteilungsfunktion von $X_1 + X_2$ gegeben durch

$$G(z) = \int F_1(z-x) f_2(x)\, d\mu_2(x) = \int F_2(z-x) f_1(x)\, d\mu_1(x).$$

**Beispiel 2.93 (Faltung von Poisson-/Normalverteilung):**

(a) Seien $X_i \sim \mathrm{Po}(\lambda_i)$, $i = 1, 2$, stochastisch unabhängig. $f_i$ sei die Dichte von $P^{X_i}$. Dann gilt für alle $z \in \mathbb{N}_0$:

$$\begin{aligned} P(X_1 + X_2 = z) &= \sum_{x \in \mathbb{N}_0} f_1(z-x) f_2(x) = \\ &= \sum_{x=0}^{z} e^{-\lambda_1} \frac{\lambda_1^{z-x}}{(z-x)!} \cdot e^{-\lambda_2} \frac{\lambda_2^x}{x!} = \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^z}{z!} \underbrace{\sum_{x=0}^{z} \binom{z}{x} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{z-x} \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{x}}_{\sim B\left(z, \frac{\lambda_2}{\lambda_1+\lambda_2}\right), \Rightarrow = 1} \end{aligned}$$

Also gilt $P^{X_1+X_2} = \mathrm{Po}(\lambda_1 + \lambda_2)$.

(b) Sei $X \sim \mathrm{N}(\mu, \sigma^2)$ und $Y \sim \mathrm{N}(\nu, \tau^2)$ stochastisch unabhängig. Dann ist $X + Y \sim \mathrm{N}(\mu + \nu, \sigma^2 + \tau^2)$. Sei dazu o.E. $\mu = \nu = 0$. Dann gilt

$$\begin{aligned} h(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi}\tau} \exp\left\{-\frac{1}{2\tau^2}(z-x)^2\right\} dx = \\ &= \frac{1}{\sqrt{2\pi} \cdot \sqrt{\sigma^2+\tau^2}} \exp\left\{-\frac{z^2}{2(\sigma^2+\tau^2)}\right\} \cdot \\ &\quad \cdot \int_{-\infty}^{\infty} \underbrace{\sqrt{\frac{\sigma^2+\tau^2}{2\pi\sigma^2\tau^2}} \exp\left\{\frac{z^2}{2(\sigma^2+\tau^2)} - \frac{z^2}{2\tau^2} - \frac{x^2}{2\tau^2} + \frac{xz}{\tau^2} - \frac{x^2}{2\sigma^2}\right\}}_{=(*)} dx \end{aligned}$$

Wie sehen

$$(*) = \sqrt{\frac{\sigma^2+\tau^2}{2\pi\sigma^2\tau^2}} \exp\left\{-\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2}\left(x - \frac{\sigma^2}{\sigma^2+\tau^2} z\right)^2\right\} =: w(x)$$

$w(x)$ entspricht der Dichte einer $\mathrm{N}\left(\frac{\sigma^2}{\sigma^2+\tau^2}z, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$- Verteilung. Also ist $\int_{-\infty}^{\infty} w(x)\, dx = 1$. Damit hat $h$ die $\lambda^1$-Dichte einer $\mathrm{N}(0, \sigma^2 + \tau^2)$-Verteilung.

**Beispiel 1.94 ($\chi_1^2$- und $\Gamma$-Verteilungen, Faltungen):**

Sind $X_1, \ldots, X_n \sim \mathrm{N}(0, \sigma^2)$ stochastisch unabhängig, so eignet sich $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ zum Schätzen von $\sigma^2$, denn

$$E\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\sum_{i=1}^{n} EX_i^2 = \frac{1}{n} \cdot n\sigma^2 = \sigma^2$$

und

$$\begin{aligned} \mathrm{Var}\, \frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\, X_i^2 = \frac{1}{n}\mathrm{Var}\, X_1^2 = \frac{1}{n} E(X_1^2 - \underbrace{EX_1^2}_{=\sigma^2})^2 = \\ &= \frac{1}{n}\sigma^4 E(Y^4 - 2EY^2 + 1) = \frac{\sigma^4}{n}(3 - 2 + 1) = \frac{2\sigma^4}{n} \overset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

wobei $Y \sim \mathrm{N}(0, 1)$. Frage jetzt: Wie ist $\sum X_i^2$ verteilt, wenn $X_1, \ldots, X_n \sim \mathrm{N}(0, 1)$ stochastisch unabhängig sind? Beispielsweise für $n = 1$ ($x > 0$):

$$P(X_1^2 \le x) = P(-\sqrt{x} \le X \le \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}).$$

Dies heißt für die $\lambda^1$-Dichte von $P^{X_1}$:

$$f(x) = \frac{d}{dx}\Phi\big(\sqrt{x} - \Phi(-\sqrt{x})\big) = \frac{1}{2}x^{-\frac{1}{2}}\varphi(\sqrt{x}) + \frac{1}{2}x^{-\frac{1}{2}}\varphi(-\sqrt{x}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x}x^{-\frac{1}{2}}.$$

$P^{X_1^2}$ heißt $\chi_1^2$-Verteilung. Für $n \geq 1$ erhält man $\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$ (zentrale $\chi^2$-Verteilung) mit $n$ „Freiheitsgraden“. Für $n \in (0,\infty)$ ist die $\lambda^1$-Dichte der $\chi_n^2$-Verteilung gegeben durch

$$f(x|n) = \frac{1}{\sqrt{2^n}\,\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}I_{(0,\infty)}(x).$$

Die $\chi_n^2$-Verteilung ist ein Spezialfall der $\Gamma$-Verteilung $\Gamma(\kappa,\sigma,\mu)$ mit $\kappa,\sigma > 0,\ \mu \in \mathbb{R}$, mit der $\lambda^1$-Dichte

$$g(x|\kappa,\sigma,\mu) = \frac{1}{\sigma^\kappa\Gamma(\kappa)}(x-\mu)^{\kappa-1}\exp\left\{-\frac{x-\mu}{\sigma}\right\}I_{(0,\infty)}(x).$$

Es gilt

$$\chi_n^2 = \Gamma\left(\frac{n}{2}, 2, 0\right), \quad \chi_2^2 = \Gamma(1,2,0) = \mathrm{EXP}\left(\frac{1}{2}\right).$$

Für $\Gamma$-Verteilungen gilt zum Beispiel

- hier fehlt wieder etwas -

d.h. es ist $\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$ (s.o.)

Beweis:

$$\begin{aligned}
&\int g(x|\kappa_1,\sigma,0)g(z-x|\kappa_2,\sigma,0)\,d\lambda^1(x) = \\
&= \int \frac{1}{\sigma^{\kappa_1+\kappa_2}\Gamma(\kappa_1)\Gamma(\kappa_2)}x^{\kappa_1-1}(z-x)^{\kappa_2-1}\exp\left(-\frac{x}{2}-\frac{x-z}{2}\right)\cdot I_{(0,\infty)}(x)\cdot I_{(0,\infty)}(z-x)\,d\lambda^1(x) = \\
&= \frac{\Gamma(\kappa_1+\kappa_2)}{\Gamma(\kappa_1)\Gamma(\kappa_2)}\frac{1}{\sigma^{\kappa_1+\kappa_2}\Gamma(\kappa_1+\kappa_2)}\exp\left(-\frac{z}{2}\right)z^{\kappa_1+\kappa_2-2}\cdot \underbrace{\int_0^z \left(\frac{x}{z}\right)^{\kappa_1-1}\left(1-\frac{x}{z}\right)^{\kappa_2-1}dx}_{\overset{y=\frac{x}{z}}{=} z\cdot \underbrace{\int_0^1 y^{\kappa_1-1}(1-y)^{\kappa_2-1}\,dy}_{(*)}} = \\
&= \frac{1}{\sigma^{\kappa_1+\kappa_2}\Gamma(\kappa_1+\kappa_2)}\exp\left(-\frac{z}{2}\right)z^{\kappa_1+\kappa_2-1} \qquad (z>0) \\
&\hat{=}\ \Gamma(\kappa_1+\kappa_2,\sigma,0)\text{-Dichte}
\end{aligned}$$

Dabei entspricht $(*)$ dem vollständigen Beta-Integral, es ist $\dfrac{\Gamma(\kappa_1)\Gamma(\kappa_2)}{\Gamma(\kappa_1+\kappa_2)}$.

**Satz 2.95 (Verteilung des Produktes von stochastisch unabhängigen Zufallsvariablen):**
Seien $X, Y$ reellwertige stochastisch unabhängige Zufallsvariablen über $(\Omega,\mathcal{A},P)$.

(a) Falls $X, Y$ diskret verteilt sind mit Träger $\mathcal{X}_0$ bzw. $\mathcal{Y}_0 \subset \mathbb{R}$, so gilt

$$P(XY=z) = \begin{cases} \sum\limits_{x\neq 0} P(X=x)P\left(Y=\frac{z}{x}\right) & z \in \mathbb{R}\setminus\{0\} \\ P(X=0) + P(X\neq 0)P(Y=0) & z=0 \end{cases}$$

(b) Falls $X, Y$ $\lambda^1$-stetige Zufallsvariablen mit Dichten $f, g$ sind, so ist $XY$ $\lambda^1$-stetig mit der $\lambda^1$-Dichte

$$h(z) = \int f(x) g\left(\frac{z}{x}\right) \frac{1}{|x|}\, d\lambda^1(x).$$

Beweis:

(a) Sei $z \neq 0$ und $B = \{(x,y) \in \mathcal{X}_0 \times \mathcal{Y}_0 : z = xy\}$. Dann folgt

$$\begin{aligned} P(XY = z) &= P^{(X,Y)}(B) \overset{X,Y \text{ stu.}}{=} \int\limits_B dP^X \otimes P^Y \overset{\text{Fubini}}{=} \\ &= \int P^Y(B_x)\, dP^X(x) \overset{(*)}{=} \sum_{x \in \mathcal{X}_0 \setminus \{0\}} P\left(Y = \frac{z}{x}\right) P(X = x) \end{aligned}$$

$(*)$ mit

$$P^Y(B_x) = \begin{cases} P\left(Y \in \left\{y \in \mathcal{Y}_0 : y = \frac{z}{x}\right\}\right) = P\left(Y = \frac{z}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Der Fall $z = 0$ ist klar.

(b) Sei $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x,y) = (x, xy)$. Dann ist $T^{-1}(u,v) = \left(u, \frac{v}{u}\right)$ für $u \neq 0$, d.h. $T^{-1}$ ist auf $(\mathbb{R}\setminus\{0\}) \times \mathbb{R}$ definiert (also bis auf eine $\lambda^2$-Nullmenge). Eine entsprechende Modifikation von Corollar 2.68 (Transformationsformel für $\lambda^n$-Dichten) liefert die $\lambda^2$-Dichte von $P^T$ gegeben durch

$$f(u) g\left(\frac{v}{u}\right) \cdot |\det DT^{-1}(u,v)| \cdot I_{(\mathbb{R}\setminus\{0\}) \times \mathbb{R}}(u,v) = f(u) g\left(\frac{v}{u}\right) \frac{1}{|u|} \quad (u \neq 0)$$

mit

$$|\det DT^{-1}(u,v)| = \left|\det \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{u} \end{pmatrix}\right| = \frac{1}{|u|}.$$

Also ist die $\lambda^1$-Dichte von $P^{XY}$ gegeben durch

$$\int f(u) g\left(\frac{v}{u}\right) \frac{1}{|u|}\, d\lambda^1(u).$$

Analog beweist man folgenden Satz:

**Satz 2.96:**
Sei $Y$ eine $\lambda^1$-stetige Zufallsvariable mit $\lambda^1$-Dichte $g$. Dann ist $Z = \frac{1}{Y}$ $\lambda^1$-stetige Zufallsvariable mit $\lambda^1$-Dichte

$$h(z) = g\left(\frac{1}{z}\right) \frac{1}{z^2}.$$

**Beispiel 2.97 (t-Verteilung mit $n$ Freiheitsgraden):**
Sind $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ stochastisch unabhängige Zufallsvariablen, so eignet sich

$$T(X_1, \ldots, X_n) := \sqrt{n} \frac{\overline{X}_\cdot - \mu_0}{S} \quad \text{mit } \overline{X}_\cdot = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_\cdot)^2,$$

zum Prüfen, ob $\mu \leq \mu_0$ oder $\mu > \mu_0$ ($\mu_0$ fest vorgegeben, $\mu, \sigma^2$ unbekannt). Wenn $T$ „groß“ ist (z.B. $\geq 3$), so spricht dies für $\mu > \mu_0$. Ist $T$ „klein“, so spricht dies für $\mu \leq \mu_0$. Man kann zeigen, daß $T$ die gleiche Verteilung hat wie

$$\frac{Y_0}{\sqrt{\frac{1}{\nu} \sum\limits_{i=1}^{\nu} Y_i^2}}$$

mit stochastisch unabhängigen $Y_0, Y_1, \ldots, Y_\nu \sim \mathrm{N}(0,1)$, $\nu = n - 1$ (falls $\mu = \mu_0$).

Ein Problem ist jetzt: Was ist

$$L\left(\frac{Y_0}{\sqrt{\frac{1}{\nu}\sum_{i=1}^{\nu} Y_i^2}}\right) = ? \qquad \text{(Law, Verteilungsgesetz)}$$

Wir definieren $W := \sum_{i=1}^{\nu} Y_i^2$. Dann gilt $W \sim \chi^2_\nu$ nach Beispiel 2.94 mit $\lambda^1$-Dichte

$$f(x|\nu) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} I_{(0,\infty)}(x)$$

Dann ist die $\lambda^1$-Dichte von $W^{\frac{1}{2}}$

$$u(x) := 2xf(x|\nu) \qquad \text{bzw.} \qquad u\left(\frac{1}{x}\right)\cdot\frac{1}{x^2} = \frac{2}{x^3} f\left(\frac{1}{x^2}\middle|\nu\right) =: \tilde{u}(x)$$

ist die $\lambda^1$-Dichte von $W^{-\frac{1}{2}}$. Wir können somit die $\lambda^1$-Dichte von $Y_0 W^{-\frac{1}{2}}$ errechnen:

$$\begin{aligned}
h(z) &= \int \tilde{u}(x)\varphi\left(\frac{z}{x}\right)\frac{1}{|x|}\, d\lambda^1(x) = \\
&= \int_0^\infty \frac{2}{x^4}\frac{\exp\left\{-\frac{1}{2x^2}\right\}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)x^{\nu-2}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{z}{x}\right)^2\right\} dx = \\
&= \int_0^\infty \frac{2\cdot 2^{\frac{\nu+2}{2}} w^{\frac{\nu+2}{2}}}{\sqrt{2\pi}\cdot 2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)(1+z^2)^{\frac{\nu+2}{2}}}\frac{(1+z^2)^{\frac{1}{2}}}{2^{\frac{3}{2}}w^{\frac{3}{2}}}\, dw = \\
&= \frac{(1+z^2)^{-\frac{\nu+1}{2}}}{\sqrt{\pi}\cdot\Gamma\left(\frac{\nu}{2}\right)} \underbrace{\int_0^\infty w^{\frac{1}{2}(\nu+1)-1} e^{-w}\, dw}_{=\Gamma\left(\frac{\nu+1}{2}\right)}
\end{aligned}$$

mit der Substitution

$$w = \frac{1}{2x^2}(1+z^2) \Rightarrow x = \frac{\sqrt{1+z^2}}{\sqrt{2w}}, \qquad dx = -\frac{x^3}{1+z^2}\, dw = \frac{\sqrt{1+z^2}}{(2w)^{\frac{3}{2}}}\, dw.$$

Somit ist die $\lambda^1$-Dichte von $\frac{Y_0}{\sqrt{\frac{1}{\nu}W}}$:

$$\frac{1}{\sqrt{\pi}} h\left(\frac{z}{\sqrt{\nu}}\right) =: t(z|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\cdot\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Die Verteilung von $\frac{Y_0}{\sqrt{\frac{1}{\nu}W}}$ heißt (zentrale) t-Verteilung mit Freiheitsgraden $\nu$ (und ist für alle $\nu \in (0,\infty)$ definiert). Eine andere Bezeichnung ist die Student'sche t-Verteilung (Gosset 1908). Bezeichnung: $\mathrm{t}_n$. Beachte: Für $\nu = 1$ ist

$$t(z|1) = \frac{1}{\pi}\frac{1}{1+\lambda^2} \mathrel{\hat{=}} \text{Cauchy}(0,1).$$

**Beispiel 2.98 (F-Verteilung $\mathrm{F}_{n,m}$):**
Seien $X_1,\ldots,X_n \sim \mathrm{N}(0,\sigma^2)$ und $Y_1,\ldots,Y_m \sim \mathrm{N}(0,\tau^2)$ mit unbekannten $\sigma,\tau > 0$. $X_1,\ldots,X_n,Y_1,\ldots,Y_m$ seien stochastisch unabhängig. Dann eignet sich

$$T(X_1,\ldots,Y_m) := \frac{1}{n}\sum_{i=1}^{n}\frac{X_i^2}{\frac{1}{m}\sum_{j=1}^{m} Y_j^2}$$

zum Vergleich der unbekannten Variablen $\sigma^2, \tau^2$: $T$ groß spricht für $\sigma^2 > \tau^2$. Umgekehrt bedeutet es $\tau^2 > \sigma^2$, wenn $T$ klein ist. Die Verteilung von $T$ heißt F-Verteilung mit Freiheitsgraden $n$ und $m$ (falls $\sigma^2 = \tau^2$).

Allgemeiner: $X \sim \chi^2_n$, $Y \sim \chi^2_m$ mit $n, m \in (0, \infty)$. Definiere $T := \frac{\frac{1}{n}X}{\frac{1}{m}Y}$. Dann ist $f(x|n)$ die $\lambda^1$-Dichte von $X$ sowie $f\left(\frac{1}{y}\middle|m\right)\frac{1}{y^2}$ die $\lambda^1$-Dichte von $Y^{-1}$. Für die $\lambda^1$-Dichte von $XY^{-1}$ berechnen wir

$$\begin{aligned}
g(z|n,m) &= \int f(x|n) f\left(\frac{x}{z}\middle|m\right)\frac{x^2}{z^2}\frac{1}{|x|}\,d\lambda^1(x) = \\
&= \int_0^\infty \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \frac{1}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} \left(\frac{x}{z}\right)^{\frac{m}{2}-1} e^{-\frac{x}{2z}} \frac{x}{z^2}\,dx = \\
&= \frac{1}{2^{\frac{m+n}{2}}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right) z^{\frac{m}{2}+1}} \int_0^\infty x^{\frac{n+m}{2}-1} \exp\left\{-\frac{x}{2}\left(1+\frac{1}{z}\right)\right\}\,dx = \\
&= \frac{z^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)(1+z)} \underbrace{\int_0^\infty w^{\frac{n+m}{2}-1} e^{-w}\,dw}_{=\Gamma\left(\frac{n+m}{2}\right)}
\end{aligned}$$

mit der Substitution $w = \frac{x}{2}\left(1+\frac{1}{z}\right)$, $x = \frac{2wz}{1+z}$ und $dx = \frac{2z}{1+z}\,dz$. Somit ist die $\lambda^1$-Dichte von $\frac{\frac{1}{n}X}{\frac{1}{m}Y}$:

$$\begin{aligned}
f(x|n,m) &= \frac{n}{m} g\left(\frac{n}{m}z\middle|n,m\right) = \\
&= \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \frac{z^{\frac{n}{2}-1}}{\left(1+\frac{n}{m}z\right)^{\frac{n+m}{2}}} \qquad (z>0).
\end{aligned}$$

Beachte: $Z \sim \mathrm{t}_m \Rightarrow Z^2 \sim \mathrm{F}_{1,m}$.

**Anmerkung:**
Die F-Verteilung taucht zum Beispiel in der Varianzanalyse auf: $X_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$. Prüfe, ob $\mu_1 = \ldots = \mu_k$. Denn wenn $\mu_1 = \ldots = \mu_k$, dann gilt

$$\frac{\frac{1}{k-1}\sum(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2}{S^2} \sim \mathrm{F}_{k-1, n_i-k}$$

mit $S^2 = \frac{1}{n_i - k}\sum\sum(X_{ij} - \overline{X}_{i\cdot})$.

**Beispiel 2.99 (Beta-Verteilung** $\mathrm{Beta}(\alpha,\beta)$**):**
Seien $X \sim \chi^2_n$ und $Y \sim \chi^2_m$ stochastisch unabhängig, $n, m \in (0, \infty)$. Setze $Z = XY^{-1}$ und $W = \frac{Z}{1+Z}$ (monotone Transformation). Dann gilt: $W \sim \mathrm{Beta}(\alpha, \beta)$ mit $\alpha = \frac{n}{2}$, $\beta = \frac{m}{2}$ und der $\lambda^1$-Dichte

$$g(w|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1}(1-w)^{\beta-1} I_{(0,1)}(w) \qquad (\alpha, \beta > 0)$$

Die Beta-Verteilung wird in der Bayes'schen Statistik als a-priori-Verteilung des Parameters $p \in (0,1)$ einer $\mathrm{B}(n,p)$-verteilten Zufallsvariable eine wichtige Rolle. Dort nimmt man an, daß der „Parameter" $p$ einer $\mathrm{Beta}(\alpha,\beta)$-Verteilung genügt und wählt $\alpha, \beta > 0$ auf der Basis gewisser Vorinformationen (oder des persönlichen Glaubens). Die Bayes'sche Formel für Dichten lautet in diesem Fall:

$$f(p|x) = \frac{f(x|p)g(p|\alpha,\beta)}{\int f(x|\tilde{p})g(\tilde{p}|\alpha,\beta)\,d\lambda^1(\tilde{p})} \qquad p \in (0,1)$$

mit $f(\cdot|p)$ Zähldichte der $\mathrm{B}(n,p)$-Verteilung und $g(\cdot|\alpha,\beta)$ $\lambda^1$-Dichte der $\mathrm{Beta}(\alpha,\beta)$-Verteilung, sowie $f(p|x)\hat{=}\mathrm{Beta}(\alpha+x, n+\beta-x)$-Dichte entspricht der posteriori-Dichte. Es gilt

$$\mathrm{Beta}(1,1)\hat{=}\mathrm{G}\big((0,1)\big).$$

# 3 Grenzwertsätze

## 3.1 Fragestellungen

(A) Ein Münzexperiment mit den Wahrscheinlichkeiten $p$ für Kopf und $1-p$ für Zahl ($p \in (0,1)$) wird unendlich oft wiederholt. Wie groß ist die Wahrscheinlichkeit, daß (i) unendlich oft Zahl und (ii) endlich oft Zahl fällt?

(B) Seien $X_1, \ldots, X_n$ stochastisch unabhängig identisch verteilt (kurz: stid., d.h. $P^{X_i} = P^{X_1}\ \forall\ i$) mit $EX_1 = \mu \in \mathbb{R}$, Var $X_1 = \sigma^2 \in (0, \infty)$. Dann gilt

$$E\overline{X}_\cdot^{(n)} = E\frac{1}{n}\sum_{i=1}^{n} X_i = \mu, \qquad \text{Var}\ \overline{X}_\cdot^{(n)} = \frac{1}{n}\sigma^2 \overset{n\to\infty}{\longrightarrow} 0.$$

1. Frage: Konvergiert die Zufallsvariable $\overline{X}_\cdot^{(n)}$ in irgendeinem Sinne gegen $\mu$? Gilt zum Beispiel

(i) $P(|\overline{X}_\cdot^{(n)} - \mu| \geq \epsilon) \overset{n\to\infty}{\longrightarrow} 0$ oder

(ii) $\lim\limits_{n\to\infty} \overline{X}_\cdot^{(n)} = \mu[P]$?

2. Frage: Zentriert und standardisiert man $\overline{X}_\cdot^{(n)}$ gemäß

$$W_n := \frac{\overline{X}_\cdot^{(n)} - \mu}{\sigma},$$

so gilt $EW_n = 0$ und

$$\text{Var}\ W_n = \frac{n}{\sigma^2} E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^2 \overset{(*)}{=} \frac{n}{\sigma^2}\frac{1}{n^2}\sum_{i=1}^{n}\underbrace{E(X_1-\mu)^2}_{=\sigma^2} = 1$$

Konvergiert die Verteilung von $W_n$, also $P^{W_n}$ für $n \to \infty$ gegen eine bekannte Verteilung?

**Bemerkung:**
Es führt

(A) zu den Null-Eins-Gesetzen,

(B) 1. (i) zu den schwachen Gesetzen der großen Zahlen (SGGZ) und zum Begriff der stochastischen Konvergenz,
(ii) zu den starken Gesetzen der großen Zahlen (StGGZ) und zum Begriff der fast-sicheren Konvergenz,
2. zu den zentralen Grenzwertsätzen (ZGWS) und zum Begriff der schwachen Konvergenz bzw. Verteilungskonvergenz von Zufallsvariablen.

Eine weitere Fragestellung: Gibt es eine Folge $(a_n)_{n\in\mathbb{N}}$, $a_n \neq 0$, so daß für eine stochastisch unabhängig identisch verteilte Folge $(X_n)_{n\in\mathbb{N}}$ mit $EX_n = 0$ und Var $X_n = \sigma^2$ gilt:

$$P(\limsup \frac{S_n}{a_n} = 1) = P(\liminf \frac{S_n}{a_n} = -1) = 1\ ?$$

Dabei sei $S_n := \sum\limits_{i=1}^{n} X_i$ ($n \in \mathbb{N}$). Äquivalent dazu

$$\limsup \frac{S_n}{a_n} = 1[P] \qquad \text{und} \qquad \liminf \frac{S_n}{a_n} = -1[P].$$

Dies führt zum Gesetz vom iterierten Logarithmus: Wähle

$$a_n = \sigma\sqrt{2n\log\log n} \qquad \text{(siehe Bauer, 1991)}$$

## 3.2 Null-Eins-Gesetz

Zunächst einige mengentheoretische Vorbereitungen:

**Definition 3.1:**
Sei $\Omega \neq \emptyset,\ \mathcal{E}, \mathcal{D} \subseteq 2^\Omega$.

(a) $\mathcal{D}$ heißt Dynkin-System, falls

(d1) $\Omega \in \mathcal{D}$,

(d2) $A, B \in \mathcal{D},\ A \subseteq B \Rightarrow B \setminus A \in \mathcal{D}$ und

(d3) $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkt $\Rightarrow \sum\limits_{n\in\mathbb{N}} A_n \in \mathcal{D}$.

(b) Es heißt

$$\mathcal{D}(\mathcal{E}) = \bigcap_{\substack{\mathcal{E}\subseteq\mathcal{D}'\subseteq\Omega \\ \mathcal{D}' \text{ Dynkin-Syst.}}} \mathcal{D}'$$

das von $\mathcal{E}$ erzeugte Dynkin-System.

**Lemman 3.2**

(a) Ist $\mathcal{D} \subseteq 2^\Omega$ ein Dynkin-System, so sind äquivalent

(i) $\mathcal{D}$ ist $\sigma$-Algebra,

(ii) $\mathcal{D}$ ist durchschnittstabil, d.h. $A, B \in \mathcal{D} \Rightarrow AB \in \mathcal{D}$.

(b) Ist $\mathcal{E} \subseteq 2^\Omega$ durchschnittstabil, so gilt $\mathcal{D}(\mathcal{E}) = \sigma(\mathcal{E})$.

Beweis: siehe Bauer, 1990.

**Bemerkung:**
Ist $\mathcal{D}$ eine $\sigma$-Algebra, so ist $\mathcal{D}$ auch ein Dynkin-System. Die Umkehrung ist im allgemeinen falsch.

**Definition 3.3 ($\sigma$-Algebra der terminalen Ereignisse):**
Für eine Folge von $\sigma$-Algebren $(\mathcal{A}_n)_{n\in\mathbb{N}}$ über $\Omega$ heißt

$$\mathcal{A}_\infty := \bigcap_{n\in\mathbb{N}} \sigma\left(\bigcup_{m\geq n} \mathcal{A}_m\right)$$

die $\sigma$-Algebra der terminalen Ereignisse (der Folge $(\mathcal{A}_n)_{n\in\mathbb{N}}$) oder abschnittsinvariante $\sigma$-Algebra. $\mathcal{A}_\infty$ ist $\sigma$-Algebra: Einfach $(\sigma 1)$ bis $(\sigma 3)$ nachrechnen.

**Beispiel 3.4:**

Seien $(A_n)_{n\in\mathbb{N}},\ A_n \subseteq \Omega,\ \mathcal{A}_n := \sigma(\{A_n\}) = \{\emptyset, A_n, A_n^c, \Omega\}$ und $\mathcal{B}_n := \sigma\left(\bigcup\limits_{m\geq n} \mathcal{A}_n\right)$. Dann ist

$$\mathcal{A}_\infty = \bigcap_{n\in\mathbb{N}} \mathcal{B}_n.$$

Setze $C_k := \bigcup\limits_{m\geq k} A_m$ für $k \geq n$. Dann ist $C_k \in \mathcal{B}_n$ für alle $k \geq n$ und wegen $C_k \downarrow$ gilt für alle $n \in \mathbb{N}$:

$$\limsup_{r\to\infty} A_r = \bigcap_{k=1}^{\infty} C_k = \bigcap_{k=n}^{\infty} C_k \in \mathcal{B}_n,$$

also $\limsup_{r\to\infty} A_r \in \mathcal{A}_\infty$. Analog kann man zeigen: $\liminf_{r\to\infty} A_r \in \mathcal{A}_\infty$.

Für unseren Münzwurf gilt: $\Omega = \{0,1\}^{\mathbb{N}}$, $A_i := \{\omega \in \Omega : \omega_i = 1\}$. $\omega_i = 1$ heißt: Kopf im $i$-ten Wurf. Dann entspricht

$$\begin{aligned} \limsup_{n\to\infty} A_n &\;\hat{=}\; \text{unendlich viele der } A_n \text{ treten ein,} \\ \liminf_{n\to\infty} A_n^c &\;\hat{=}\; \text{schließlich alle } A_n^c \text{ treten ein, d.h. höchstens endlich oft Kopf.} \end{aligned}$$

**Satz 3.5 (0-1-Gesetz von Kolmogoroff):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine stochastisch unabhängige Folge von $\sigma$-Algebren $(\mathcal{A}_n)_{n\in\mathbb{N}} \subset \mathcal{A}$. Dann gilt für alle $A \in \mathcal{A}_\infty$: $P(A) \in \{0,1\}$.

Beweis: Sei $A \in \mathcal{A}_\infty$ beliebig. Definiere $\mathcal{D} := \{D \in \mathcal{A} : P(AD) = P(A)P(D)\}$, d.h. $\mathcal{D}$ ist die Menge aller von $A$ stochastisch unabhängigen Ereignisse. Wie zeigen, daß $\mathcal{D}$ ein Dynkin-System ist:

(d1) $P(A\Omega) = P(A)P(\Omega)$, also gilt $\Omega \in \mathcal{D}$,

(d2) Seien $E, F \in \mathcal{D}$ mit $E \subseteq F$. Dann gilt

$$P\big(A(F \setminus E)\big) = P(AF) - P(AE) = P(A)P(F) - P(A)P(E) = P(A)P(F \setminus E),$$

also $F \setminus E \in \mathcal{D}$ nach Definition,

(d3) Sei $(D_n)_{n\in\mathbb{N}} \subset \mathcal{D}$ paarweise disjunkt. Dann gilt:

$$P\Big(A\sum_{n\in\mathbb{N}} D_n\Big) = \sum_{n\in\mathbb{N}} P(AD_n) = \sum_{n\in\mathbb{N}} P(A)P(D_n) = P(A)\sum_{n\in\mathbb{N}} P(D_n) = P(A)P\Big(\sum_{n\in\mathbb{N}} D_n\Big),$$

also $\sum_{n\in\mathbb{N}} \in \mathcal{D}$.

Sei $\mathcal{T}_n^* := \sigma(\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_n)$ und $\mathcal{T}_n := \sigma\big(\bigcup_{m\geq n} \mathcal{A}_m\big)$ für alle $n \in \mathbb{N}$. Dann sind, wegen der stochastischen Unabhängigkeit der $(\mathcal{A}_n)_{n\in\mathbb{N}}$, $\mathcal{T}_n^*$ und $\mathcal{T}_{n+1}$ stochastisch unabhängig (wegen Lemma 3.6 (c) s.u.). Wegen $A \in \mathcal{A}_\infty \subseteq \mathcal{T}_{n+1}$ sind daher $\{A\}$ und $\mathcal{T}_n^*$ stochastisch unabhängig für alle $n \in \mathbb{N}$. Nach Konstruktion der $\mathcal{D}$ gilt $\mathcal{T}_n^* \in \mathcal{D}$ für alle $n \in \mathbb{N}$ und somit

$$\mathcal{T}^* := \bigcup_{n=1}^{\infty} \mathcal{T}_n^* \subseteq \mathcal{D}.$$

$\mathcal{T}^*$ ist durchschnittstabil, denn $\mathcal{T}_n^* \uparrow$, d.h. für alle $E, F \in \mathcal{T}^*$ existiert ein $n \in \mathbb{N}$ mit $E, F \in \mathcal{T}_n^*$.

Nach Lemma 3.2 und mit $\mathcal{T}^* \subseteq \mathcal{D}$ folgt $\sigma(\mathcal{T}^*) = \mathcal{D}(\mathcal{T}^*) \subset \mathcal{D}$. Da $\mathcal{A}_n \subset \mathcal{T}^*$ für alle $n \in \mathbb{N}$, folgt

$$\forall\, n \in \mathbb{N}: \quad \mathcal{A}_n \subseteq \sigma(\mathcal{T}^*) \subseteq \mathcal{D}.$$

Somit gilt für alle $n \in \mathbb{N}$:

$$\sigma\Big(\bigcup_{m\geq n} \mathcal{A}_m\Big) \subseteq \sigma(\mathcal{T}^*) \subset \mathcal{D},$$

also $\mathcal{A}_\infty = \bigcap_{n=1}^{\infty} \sigma\big(\bigcup_{m\geq n} \mathcal{A}_m\big) \subseteq \sigma(\mathcal{T}^*) \subseteq \mathcal{D}$. Für unser vorgegebenes $A$ folgt damit $A \in \mathcal{D}$, d.h. $A$ ist von sich selbst unabhängig, also muß gelten:

$$P(A) = P(AA) = P(A)P(A) = \big(P(A)\big)^2 \in \{0,1\}.$$

**Lemma 3.6 (Nachtrag zur stochastischen Unabhängigkeit):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit $(\mathcal{E}_i)_{i\in I}$, $\mathcal{E}_i \subseteq \mathcal{A}$. Dann gelten:

(i) $(\mathcal{E}_i)_{i\in I}$ stochastisch unabhängig $\Rightarrow \big(\mathcal{D}(\mathcal{E}_i)\big)_{i\in I}$ stochastisch unabhängig,

(ii) Sind $(\mathcal{E}_i)_{i\in I}$ stochastisch unabhängig, $\forall\, i \in I:\ \mathcal{E}_i$ durchschnittstabil. Dann gilt: $\big(\sigma(\mathcal{E}_i)\big)_{i\in I}$ stochastisch unabhängig,

(iii) Sind $(\mathcal{E}_i)_{i\in I}$ stochastisch unabhängig, $\forall\, i \in I:\ \mathcal{E}_i$ durchschnittstabil und $I = \sum\limits_{j\in J} I_j$ sowie

$$\mathcal{A}_j := \sigma\Big(\bigcup_{i\in I_j} \mathcal{E}_i\Big).$$

Dann gilt

$$(\mathcal{A}_j)_{j\in J} \text{ stochastisch unabhängig} \qquad \text{(Zusammenfassen unabhängiger } \sigma\text{-Algebren).}$$

**Satz 3.7 (Lemma von Borel-Cantelli):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$, und

$$A_\infty := \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m\geq n} A_m \qquad (\text{„}A_n \text{ } \infty\text{ oft“}).$$

Dann gilt

(a) Ist $\sum\limits_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(A_\infty) = 0$,

(b) Sind $(A_n)_{n\in\mathbb{N}}$ stochastisch unabhängig und $\sum\limits_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(A_\infty) = 1$.

Beweis:

(a) Wegen $\sum P(A_n) < \infty$ gilt

$$P(A_\infty) = P\Big(\bigcap_{n=1}^{\infty} \bigcup_{m\geq n} A_m\Big) \leq P\Big(\bigcup_{m\geq n} A_m\Big) \leq \sum_{m\geq n} P(A_m) \overset{n\to\infty}{\longrightarrow} 0.$$

(b) Setze $\mathcal{A}_n := \sigma\big(\{A_n\}\big) = \{\emptyset, A_n, A_n^c, \Omega\}$. Dann sind $(\mathcal{A}_n)_{n\in\mathbb{N}}$ stochastisch unabhängig, d.h. wir sind in der Situation von Satz 3.5 und nach Beispiel 3.4 ist

$$A_\infty \in \mathcal{A}_\infty = \bigcap_{n=1}^{\infty} \sigma\Big(\bigcup_{m\geq n} \mathcal{A}_m\Big).$$

Also gilt $P(A_\infty) \in \{0, 1\}$. Weiter gilt

$$\begin{aligned} P(A_\infty^c) &= P(\liminf A_n^c) = P\Big(\bigcup_{n=1}^{\infty} \bigcap m \geq n A_m^c\Big) \overset{\text{Stetigkeitssatz}}{=} \lim_{n\to\infty} P\Big(\bigcap_{m\geq n} A_m^c\Big) = \\ &= \lim_{n\to\infty} \lim_{k\to\infty} P\Big(\bigcap_{m=n}^{k} A_m^c\Big) \overset{\text{stu}}{=} \lim_{n\to\infty} \lim_{k\to\infty} \prod_{m=n}^{k} P(A_m^c) = \lim_{n\to\infty} \prod_{m=n}^{\infty} \big(1 - P(A_m)\big) = 0, \end{aligned}$$

denn für $\alpha_n \in [0, 1]$, $n \in \mathbb{N}$, gilt:

$$\sum_{n\in\mathbb{N}} \alpha_n < \infty \iff \prod_{n\in\mathbb{N}} (1 - \alpha_n) > 0.$$

Da $\sum_{n\in\mathbb{N}} P(A_n) = \infty$, folgt $\sum_{m\geq n} P(A_m) = \infty$ für alle $n \in \mathbb{N}$, also $\prod_{m\geq n} \big(1-P(A_m)\big) = 0$ für alle $n \in \mathbb{N}$. Somit ist $P(A_\infty) = 1$.

**Beispiel 3.8 (Münzwurf):**
Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ mit $P(X_i = 1) = p,\ P(X_i = 0) = 1-p$. Sei definiert $A_n := \{\omega :\ X_n(\omega) = 1\}$ für alle $n \in \mathbb{N}$. Dann gilt $(A_n)_{n\in\mathbb{N}}$ stochastisch unabhängig. Dann gilt

$$A_\infty = \{\omega \in \Omega :\ X_n(\omega) = 1 \text{ für unendlich viele } n\} = \operatorname{limsup} A_n.$$

Wegen $\sum_{n=1}^{\infty} P(A_n) = \infty$ ist $P(A_\infty) = 1$. Weiter gilt für

$$B_\infty := \operatorname{liminf} A_n^c = \{\omega \in \Omega :\ X_n(\omega) = 1 \text{ für höchstens endlich viele } n\},$$

daß

$$P(B_\infty) = 1 - P(B_\infty^c) = 1 - P(A_\infty) = 0.$$

**Bemerkung:**
Es läßt sich zeigen, daß jede endliche Sequenz von Nullen und Einsen, zum Beispiel 010 mit Wahrscheinlichkeit 1 unendlich oft vorkommt.

**Beispiel 3.9 (Irrfahrt eines Teilchens):**
$(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ mit $P(X_i = -1) = p,\ P(X_i = 1) = 1-p$. Sei $S_n := \sum_{i=1}^{n} X_i$ die Summe der ersten $n$ Zufallsvariablen $X_i$. Sei $A_n := \{\omega \in \Omega :\ S_n(\omega) = 0\}$. Dann gilt

(a) Falls $p \neq \frac{1}{2}$, gilt $P(A_\infty) = 0$,

(b) falls $p = \frac{1}{2}$, so gilt $P(A_\infty) = 1$, d.h. „egal welchen Wert $S_n$ hat, man kommt mit Wahrscheinlichkeit 1 wieder zu $S_m = 0$ für ein $m > n$“.

Beweis:

(a) Wegen $A_{2n+1} = \infty \Rightarrow A_\infty = \operatorname{limsup} A_{2n}$. Es ist

$$P(A_{2n}) = \binom{2n}{n} p^n (1-p)^n.$$

Mit Hilfe der Stirling-Formel

$$g(n) \exp\left\{\frac{1}{12n+1}\right\} < n! < g(n) \exp\left\{\frac{1}{12n}\right\}$$

mit $g(n) = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$ erhält man

$$\binom{2n}{n} = 2^{2n} \frac{1}{\sqrt{\pi n}} (1+\delta_n) \qquad \text{mit } \delta_n \overset{n\to\infty}{\longrightarrow} 0.$$

Es folgt

$$P(A_{2n}) = 2^{2n} \frac{1}{\sqrt{\pi n}} (1+\delta_n) \big(p(1-p)\big)^n = \frac{1}{\sqrt{\pi n}} (1+\delta_n) \big(\underbrace{4p(1-p)}_{=:q<1}\big)^n,$$

also schließlich $P(A_{2n}) < q^n$:

$$\sum_{n\in\mathbb{N}} P(A_{2n}) < \sum_{n\in\mathbb{N}} q^n < \infty \quad \Rightarrow \quad P(A_\infty) = 0.$$

(b) Problem: $(A_n)_{n\in\mathbb{N}}$ sind nicht stochastisch unabhängig. Wähle deshalb eine Teilfolge $(n_k)_{k\in\mathbb{N}}$. Dann gilt für

$$C_k \in \sigma\Big(\bigcup_{n_k<j\le n_{k+1}} \sigma(X_j)\Big),$$

daß $(C_k)_{k\in\mathbb{N}}$ stochastisch unabhängig sind. Im folgenden werden $n_k$ und $C_k$ so gewählt, daß

(i) $\sum\limits_{k\in\mathbb{N}} P(C_k) = \infty$ und

(ii) $\operatorname{limsup} C_k \subseteq \operatorname{limsup} A_k =: A_\infty$.

Mit Borel-Cantelli folgt dann $P(A_\infty) \ge P(\operatorname{limsup} C_k) = 1$, also $P(A_\infty) = 1$. Angenommen, $(n_k)$ sei schon gewählt und $(m_k)$ mit $n_k \le m_k \le n_{k+1}$. Definiere

$$C_k := \{X_{n_k+1} + \ldots + X_{m_k} \le -n_k\} \cap \{X_{m_k+1} + \ldots + X_{n_{k+1}} \ge m_k\}.$$

Dann gilt für alle $\omega \in C_k$:

$$S_{m_k}(\omega) \le 0 \text{ und } S_{n_{k+1}}(\omega) \ge 0, \text{ d.h. } C_k \subseteq \bigcup_{j=n_k+1}^{n_{k+1}} A_j.$$

Somit gilt $\operatorname{limsup} C_k \subseteq A_\infty$, also ist (ii) erfüllt.

Wähle nun $(n_k), (m_k)$ so, daß $\sum\limits_{k=1}^{\infty} P(C_k) = \infty$. Dazu eine Zwischenbehauptung:

Für alle $\alpha \in (0,1)$ und $k \in \mathbb{N}$ existiert ein Index $r(k) \in \mathbb{N}$ mit $P(|S_{r(k)}| \le k) \le \alpha$.

Beweis der Zwischenbehauptung: Für $j = -n, -n+2, \ldots, n-2, n$ gilt

$$P(S_n = j) = \frac{1}{2^n}\binom{1}{\frac{1}{2}(n+j)}.$$

Es gilt $S_n = j$ genau dann, wenn $\frac{n-j}{2}$ der $X_1, \ldots, X_n = -1$ und $n - \frac{n-j}{2}$ der $X_1, \ldots, X_n = 1$. Also folgt

$$\begin{aligned}\lim_{n\to\infty} \sum_{|j|\le k} P(S_n = j) &= \lim_{\substack{n\to\infty \\ j\in\{-n,-n+2,\ldots,n\}}} \sum_{|j|\le k} \frac{1}{2^n}\binom{1}{\frac{1}{2}(n+j)} \le \lim_{\substack{n\to\infty \\ j\in\{-n,-n+2,\ldots,n\}}} \sum_{|j|\le k} \frac{1}{2^n}\binom{n}{\frac{n}{2}} \le \\ &\le (k+1)\lim_{n\to\infty}\frac{1}{2^n}\binom{n}{\frac{n}{2}} = (k+1)\lim_{n\to\infty}\frac{1}{2^n}\cdot 2^n\cdot\frac{1}{\sqrt{\pi n}}(1+\delta_n) = 0,\end{aligned}$$

also folgt die Zwischenbehauptung. Somit existiert für alle $\alpha \in (0,1)$ und alle $k \in \mathbb{N}$ ein $n = r(k)$ mit

$$\sum_{|j|\le k} P(S_n = j) \le \alpha.$$

Definiere $(n_k), (m_k)$ nun wie folgt:

$$m_k := n_k + r(n_k), \qquad n_{k+1} := m_k + r(m_k).$$

Es folgt:

$$\begin{aligned}P(C_k) &\overset{X_i \text{ stu}}{=} P(X_{n_k+1} + \ldots + X_{m_k} \le -n_k)\cdot P(X_{m_k+1} + \ldots + X_{n_{k+1}} \ge m_k) = \\ &\overset{\text{sym.}}{=} \frac{1}{4}P\big(|X_{n_k+1} + \ldots + X_{m_k}| \ge n_k\big)\cdot P\big(|X_{m_k+1} + \ldots + X_{n_{k+1}}| \ge m_k\big) = \\ &= \frac{1}{4}P\big(|S_{r(n_k)}| \ge n_k\big)\cdot P\big(|S_{r(m_k)}| \ge m_k\big) \ge \frac{1}{4}(1-\alpha)^2 > 0.\end{aligned}$$

Also folgt $\sum\limits_{k\in\mathbb{N}} P(C_k) = \infty$ und damit die Behauptung.

**Bemerkung 3.10 (Ergänzung zu Lemma 3.6):**
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und stochastisch unabhängige Zufallsvariablen

$$X_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$$

für $i \in I$. Sei weiter $I := \sum_{j \in J} I_j$ und $T_j : (\times_{i \in I_j} \Omega_i, \bigotimes_{i \in I_j} \mathcal{A}_i) \to (\Omega'_j, \mathcal{A}'_j)$. Dann sind

$$\big(T_j \circ (X_i : \ i \in I_j)\big)_{j \in J} \text{ stochastisch unabhängig.}$$

## 3.3 Gesetze der großen Zahlen

**Definition 3.11:**
Gegeben $(X_n)_{n \in \mathbb{N}}$, $X$ reellwertige Zufallsvariable auf $(\Omega, \mathcal{A}, P)$.

(a) $(X_n)_{n \in \mathbb{N}}$ konvergiert $P$-stochastisch gegen $X$ $(X_n \xrightarrow{P} X) :\Longleftrightarrow$

$$\forall\, \epsilon > 0 : \quad \lim_{n \to \infty} P\big(|X_n - X| \geq \epsilon\big) = 0.$$

(b) $(X_n)_{n \in \mathbb{N}}$ konvergiert $P$-fast-sicher gegen $X$ $(X_n \to X[P]) :\Longleftrightarrow$

$$\lim_{n \to \infty} X_n = X[P].$$

Dies ist äquivalent zu $P\big(\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}\big) = 1$.

**Definition 3.12:**
Für eine Folge $(X_n)_{n \in \mathbb{N}}$ von reellwertigen integrierbaren Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ gilt

(a) das schwache Gesetz der großen Zahlen (SGGZ) $:\Longleftrightarrow$

$$\forall\, \epsilon > 0 : \quad \lim_{n \to \infty} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i)\Big| \geq \epsilon\Big) = 0.$$

(b) das starke Gesetz der großen Zahlen (StGGZ) $:\Longleftrightarrow$

$$\lim_{n \to \infty} \frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) = 0[P].$$

**Bemerkung 3.13:**

(a) Es gilt: $X_n \to X[P] \Rightarrow X_n \xrightarrow{P} X$. Die Umkehrung gilt i.a. nicht.

(b) Wegen Teil (a) gilt:

$$(X_n)_{n \in \mathbb{N}} \text{ genügt dem StGGZ } \Rightarrow (X_n) \text{ genügt dem SGGZ.}$$

Die Umkehrung gilt i.a. nicht.

(c) Folgende Aussagen sind äquivalent:

(i) $X_n \to X[P]$ für eine reellwertige Zufallsvariable $X$,

(ii) $\forall\, \epsilon > 0 : \quad \lim\limits_{m\to\infty} P\big(\bigcup\limits_{n\geq m} \{|X_n - X| \geq \epsilon\}\big) = 0,$

(iii) $\forall\, \epsilon > 0 : \quad \lim\limits_{m\to\infty} P\big(\sup\limits_{n\geq m} |X_n - X| \geq \epsilon\big) = 0,$

(iv) $\forall\, \epsilon > 0 : \quad \lim\limits_{m\to\infty} P\big(\sup\limits_{n\geq m} |X_n - X_m| \geq \epsilon\big) = 0.$

Falls $(\alpha_n)_{n\in\mathbb{N}},\ \alpha_n > 0,\ \lim\limits_{n\to\infty} \alpha_n = 0$ und $\sum\limits_{i=1}^{\infty} P\big(|X_n - X| \geq \alpha_n\big) < \infty$, dann folgt (i).

(d) Ebenfalls äquivalent sind

(i) $X_n \xrightarrow{P} X$,

(ii) Zu jeder Teilfolge von $(X_n)_{n\in\mathbb{N}}$ gibt es eine fast überall gegen $X$ konvergente Teilfolge.

Im folgenden sei als generelle Voraussetzung eine Folge $(X_n)_{n\in\mathbb{N}}$ von reellwertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ gegeben.

**Satz 3.14 (SGGZ von Markov, publ. 1906):**
Sind die $(X_n)_{n\in\mathbb{N}}$ quadrat-integrierbar und ist die Markov-Bedingung

$$\lim_{n\to\infty} \frac{1}{n^2} \mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = 0$$

erfüllt, so gilt das schwache Gesetz der großen Zahlen (SGGZ).

Beweis: Die Tschebyschev-Ungleichung ergibt für alle $\epsilon > 0$:

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i)\Big| \geq \epsilon\Big) \quad \leq \quad \frac{1}{\epsilon^2 n^2} E\Big(\sum_{i=1}^{n}(X_i - EX_i)\Big)^2 = \frac{1}{\epsilon^2 n^2}\Big(\mathrm{Var}\sum_{i=1}^{n} X_i\Big) \to 0.$$

**Corollar 3.15:**

(a) Seien $X_n \sim \mathrm{B}(n, p)$ für $n \in \mathbb{N}$ stochastisch unabhängig, dann gilt das schwache Gesetz der großen Zahlen (SGGZ von Bernoulli, publ. 1718).

(b) Seien $(X_n)_{n\in\mathbb{N}}$ quadrat-integrierbar, paarweise unkorrelliert und identisch verteilt, dann gilt das SGGZ.

(c) Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig und $(\mathrm{Var}\, X_n)_{n\in\mathbb{N}}$ beschränk, dann gilt das SGGZ.

(d) Seien $(X_n)_{n\in\mathbb{N}}$ quadrat-integrierbar, paarweise unkorrelliert und

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\, X_i = 0,$$

dann gilt das SGGZ.

**Beispiel 3.16 (Irrfahrt eines Teilchens, Random Walk):**
Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig mit $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$, dann gilt das SGGZ. Für $S_n := \sum\limits_{i=1}^{n} X_i$ gilt also

$$\forall\, \epsilon > 0 : \quad P\big(|S_n| \geq n\epsilon\big) \overset{n\to\infty}{\longrightarrow} 0.$$

Im folgenden beschäftigen wir uns mit dem starken Gesetz der großen Zahlen (StGGZ). Der folgende Satz fast drei Varianten des StGGZ zusammen.

**Satz 3.17:**
Eine Folge $(X_n)_{n\in\mathbb{N}}$ reeller integrierbarer Zufallsvariablen genügt dem StGGZ, falls eine der drei folgenden Bedingungen erfüllt ist:

(a) $(X_n)_{n\in\mathbb{N}}$ identisch verteilt und paarweise stochastisch unabhängig (StGGZ von Etemadi, 1981),

(b) $(X_n)_{n\in\mathbb{N}}$ quadrat-integrierbar, stochastisch unabhängig und erfüllt die Kolmogorov-Bedingung

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathrm{Var}\, X_n < \infty$$

(StGGZ für quadrat-integrierbare, stochastisch unabhängige Zufallsvariablen von Kolmogorov von 1928),

(c) $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig identisch verteilt (StGGZ für integrierbare stochastisch unabhängige identisch verteilte Zufallsvariablen von Kolmogoroff, 1930).

Wir beweisen hier nur (b) und (c). Einen Beweis von (a) findet man z.B. in Bauer (1991), S. 86 ff. Es folgt (c) natürlich aus (a). Wir werden (c) jedoch auf (b) zurückführen und dabei die Technik des „Stutzens" von Zufallsvariablen kennenlernen. Zum Beweis von (b) jedoch noch drei vorbemerkende Resultate:

**Satz 3.18 (Kolmogorov-Ungleichung, 1930):**
Seien $X_1, \ldots, X_n$ stochastisch unabhängig und quadrat-integrierbar mit $EX_i = 0$. Sei weiter

$$\tau_n^2 := \sum_{i=1}^{n} \mathrm{Var}\, X_i; \quad S_j := \sum_{i=1}^{j} X_i.$$

Dann gilt für alle $\epsilon > 0$ :

$$P\big(\max_{1\le j\le n} |S_j| \ge \epsilon\big) \le \frac{\tau_n^2}{\epsilon^2}.$$

Beweis: Setze $A_j := \{|S_j| \ge \epsilon\}$ und $A := \big\{\max_{1\le j\le n} |S_j| \ge \epsilon\big\}$ sowie

$$A := \bigcup_{j=1}^{n} A_j =: \sum_{j=1}^{n} B_j\,, \quad B_1 := A_1,\ B_j := A_j A_1^c \cdots A_{j-1}^c.$$

$B_j \to$ erstmals bei $j$ : $|S_j| \ge \epsilon$.

Also ist zu zeigen: $P(A) \le \frac{1}{\epsilon^2}\mathrm{Var}\, S_n$ (da $\tau_n^2 = \sum\limits_{i=1}^{n} \mathrm{Var}\, X_i = \mathrm{Var}\, S_n$). Es gilt

$$\begin{aligned}
\mathrm{Var}\, S_n &= \int S_n^2\, dP \ge \int_A S_n^2\, dP = \sum_{j=1}^{n} \int_{B_j} (S_j - S_n + S_j)^2\, dP = \\
&= \sum_{j=1}^{n} \int_{B_j} (S_j^2 + 2S_j(S_j - S_n) + \underbrace{(S_n - S_j)^2}_{\ge 0})\, dP \ge \\
&\ge \sum_{j=1}^{n} \epsilon^2 P(B_j) + 2 \cdot \sum_{j=1}^{n} \int \underbrace{I_{B_j} S_j (S_j - S_n)}_{\substack{\text{stu., da} \\ (X_1,\ldots,X_j),(X_{j+1},\ldots,X_n) \text{ stu.}}}\, dP = \\
&= \epsilon^2 P(A) + 2 \cdot \sum_{j=1}^{n} \Big(\int_{B_j} S_j\, dP\Big) \underbrace{\big(E(S_j - S_n)\big)}_{=0} = \epsilon^2 P(A).
\end{aligned}$$

**Satz 3.19 (Zweireihensatz):**
Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig und quadrat-integrierbar mit

(i) $\sum_{n=1}^{\infty} EX_n$ konvergiert (in $\mathbb{R}$),

(ii) $\sum_{n=1}^{\infty} \operatorname{Var} X_n < \infty$,

dann konvergiert $\sum_{i=1}^{\infty} X_n$ fast sicher (in $\mathbb{R}$).

Beweis: Setze wieder $S_n := \sum_{i=1}^{n} X_i$. Wegen (i) kann ohne Einschränkung $EX_n = 0$ für alle $n \in \mathbb{N}$ angenommen werden (ansonsten rechne mit $Y_n := X_n - EX_n$ weiter). Für $m \in \mathbb{N}$ sei $Y_i := X_{i+m}$. Dann gilt mit

$$T_j := \sum_{i=1}^{j} Y_i = S_{j+m} - S_j$$

$$\begin{aligned} P\big(\max_{1\le j\le k} |S_{j+m} - S_j| \ge \epsilon\big) &= P\big(\max_{1\le j\le k} |T_j| \ge \epsilon\big) \overset{\text{Kolm.-Ungl.}}{\le} \\ &\le \frac{1}{\epsilon^2} \sum_{i=1}^{k} \operatorname{Var} X_i = \frac{1}{\epsilon^2} \sum_{i=m+1}^{m+k} \operatorname{Var} X_i \end{aligned}$$

Es folgt

$$P\big(\sup_j |S_{j+m} - S_m| \ge \epsilon\big) = \lim_{j\to\infty} P\big(|S_{j+m} - S_m| \ge \epsilon\big) \le \frac{1}{\epsilon^2} \sum_{i=m+1}^{\infty} \operatorname{Var} X_i$$

Wegen Voraussetzung (ii) gilt für alle $\epsilon > 0$:

$$\lim_{m\to\infty} P\big(\sup_j |S_{j+m} - S_m| \ge \epsilon\big) = 0.$$

Also folgt $S_n \to S[P]$ für eine reellwertige Zufallsvariable $S$ (nach Bemerkung 3.13 (c)) und damit die Behauptung.

**Lemma 3.20 (Kronecker-Lemma):**
Sei $0 < \tau_n < \infty$ mit $\tau_n$ monoton gegen $\infty$ konvergent, sowie $b_n \in \mathbb{R}$ mit $\sum_{n=1}^{\infty} \frac{b_n}{\tau_n}$ konvergent in $\mathbb{R}$. Dann gilt

$$\lim_{n\to\infty} \frac{1}{\tau_n} \sum_{j=1}^{n} b_j = 0.$$

Beweis: Setze $s_n := \sum_{i=1}^{n} \frac{b_i}{\tau_i}$, $b_0 := 0$, $\tau_0 := 0$, $s_0 := 0$. Nach Voraussetzung gilt $s_n \to s \in \mathbb{R}$. Wegen $\tau_n(s_n - s_{n-1}) = b$ für $n \ge 1$ folgt

$$\begin{aligned} \frac{1}{\tau_n} \sum_{i=1}^{n} b_i &= \frac{1}{\tau_n} \sum_{i=1}^{n} \tau_i(s_i - s_{i-1}) = \frac{1}{\tau_n}\Big(\sum_{i=1}^{n} \tau_i s_i - \underbrace{\sum_{i=2}^{n} \tau_i s_{i-1}}_{=\sum_{i=1}^{n-1} \tau_{i+1} s_i}\Big) = \\ &= \frac{1}{\tau_n}\Big(\Big(\sum_{i=1}^{n-1} (\tau_i - \tau_{i+1}) s_i\Big) + \tau_n s_n\Big) = s_n - \frac{1}{\tau_n} \sum_{i=1}^{n-1} (\tau_i - \tau_{i-1}) s_{i-1}. \end{aligned}$$

Aus der Analysis ist der folgende Sachverhalt bekannt:

$$s_n \to s,\ \sigma_n \geq 0 \text{ mit } \sum_{i=1}^{\infty} \sigma_n = \infty \quad \Rightarrow \quad \frac{\sigma_1 s_1 + \ldots + \sigma_n s_n}{\sigma_1 + \ldots + \sigma_n} \to s.$$

Wähle hierfür $\sigma_i := \tau_i - \tau_{i-1} \Rightarrow \sum_{i=1}^{n} \sigma_i = \tau_n$.

$$\Rightarrow \quad \frac{1}{\tau_n} \sum_{i=1}^{n-1} (\tau_i - \tau_{i-1}) s_{i-1} \to s, \quad \Rightarrow \quad \frac{1}{\tau_n} \sum_{i=1}^{n} b_i \to 0.$$

**Beweis von Satz 3.17 (b) (mit dem Zweireihensatz und dem Kronecker-Lemma):**
Setze $Y_n := \frac{1}{n}(X_n - EX_n) \Rightarrow EY_n = 0$ und $\text{Var } Y_n = \frac{1}{n^2}\text{Var } X_n$. Nach Voraussetzung folgt

$$\sum_{n=1}^{\infty} \text{Var } Y_n = \sum_{i=1}^{\infty} \frac{1}{n^2} \text{Var } X_n < \infty.$$

Weiterhin wissen wir $\sum_{n=1}^{\infty} EY_n = 0$. Somit folgt, daß

$$\sum_{n=1}^{\infty} Y_n = \sum_{n=1}^{\infty} \frac{X_n - EX_n}{n}$$

nach Satz 3.19 (Zweireihensatz) fast sicher konvergiert. Mit dem Kronecker-Lemma folgt

$$\frac{1}{n} \sum_{j=1}^{n} (X_j - EX_j) \to 0 [P].$$

Zum Beweis von Satz 3.17 (c) noch folgendes Lemma.

**Lemma 3.21:**
Seien $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ und $Y$ reellwertige Zufallsvariablen sowie $a < a_n \overset{n\to\infty}{\longrightarrow} \infty$ mit

$$\frac{1}{a_n} \sum_{i=1}^{n} Y_i \to Y [P] \quad \text{und} \quad \sum_{n=1}^{\infty} P(Y_n \neq X_n) < \infty.$$

Dann folgt $\frac{1}{a_n} \sum_{i=1}^{n} X_i \overset{n\to\infty}{\longrightarrow} Y [P]$.

Beweis: Nach Satz 3.7 (Borel-Cantelli) folgt $P(\lim_{n\to\infty} \sup\{Y_n \neq X_n\}) = 0$ und damit

$$P\Big(\underbrace{\lim_{n\to\infty} \inf\{Y_n = X_n\}}_{=:A=\{X_n = Y_n \text{ für fast alle } n\}}\Big) = 1.$$

Betrachte $B := \left\{ \frac{1}{n} \sum_{i=1}^{n} Y_i \to Y \right\}$. Dann gilt $P(AB) = 1$. Für $\omega \in AB$ existiert ein $n(\omega) \in \mathbb{N}$ mit $Y_n(\omega) = X_n(\omega)\ \forall\ n > n(\omega)$. Dann folgt für alle $n \geq n(\omega)$

$$\begin{aligned} \frac{1}{a_n} \sum_{i=1}^{n} X_i(\omega) &= \frac{1}{a_n} \sum_{i=1}^{n(\omega)} X_i(\omega) + \frac{1}{a_n} \sum_{i=n(\omega)+1}^{n} X_i(\omega) = \\ &= \underbrace{\frac{1}{a_n} \sum_{i=1}^{n(\omega)} X_i(\omega)}_{\to 0} + \underbrace{\frac{1}{a_n} \sum_{i=1}^{n} Y_i(\omega)}_{\to Y(\omega)} - \underbrace{\frac{1}{a_n} \sum_{i=1}^{n(\omega)} Y_i(\omega)}_{\to 0} \longrightarrow Y_n \end{aligned}$$

**Beweis von Satz 3.17 (c) (durch „Stutzen" der $X_n$ und Zurückführung auf (b)):**
Wir stutzen die $X_n$, um quadrat-integrierbare Zufallsvariablen zu erhalten:

$$U_n := X_n \cdot I_{[-n,n]}.$$

Dann sind die $(U_n)_{n\in\mathbb{N}}$ stochastisch unabhängig und quadrat-integrierbar (da beschränkt). Beachte: Die $(U_n)_{n\in\mathbb{N}}$ sind nicht identisch verteilt. Der Beweis bedient sich folgender drei Behautungen:

(i) $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}\mathrm{Var}\, U_n < \infty$ (Kolmogorov-Bedingung),

(ii) $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=1}^{n} EU_i = EX_1$ und

(iii) $\sum\limits_{n=1}^{\infty} P(U_n \neq X_n) < \infty$.

Aus (i) bis (iii) folgt das StGGZ für die $(X_n)_{n\in\mathbb{N}}$, denn aus (i) folgt das StGGZ für $(U_n)_{n\in\mathbb{N}}$ nach Satz 3.17 (b), d.h.

$$\frac{1}{n}\sum_{i=1}^{n}(U_i - EU_i) \longrightarrow 0[P].$$

Zusammen mit (ii) ergibt dies $\frac{1}{n}\sum\limits_{i=1}^{n} U_i \to EX_1[P]$. Mit Lemma 3.21 (setze $a_n := n$) folgt aus (iii) schließlich

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to EX_1[P] \quad \text{bzw.} \quad \frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \to 0[P] \quad (\text{da } (X_i) \text{ stid.})$$

und damit die Behauptung. Zum Beweis von (i) bis (iii):

(i) Setze $A_j := \{j-1 < |X_1| \le j\}$ für $j \in \mathbb{N}$. Dann sind $A_j$ paarweise disjunkt. Somit gilt

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{n^2}\mathrm{Var}\, U_n &\le \sum_{n=1}^{\infty}\frac{1}{n^2}EU_n^2 = \sum_{n=1}^{\infty}\frac{1}{n^2}\int\limits_{\{|X_1|\le n\}} X_1^2\, dP = \sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{j=1}^{n}\int\limits_{A_j} X_1^2\, dP \overset{\text{Fub. für Zählmaße}}{=} \\ &= \sum_{j=1}^{\infty}\Big(\underbrace{\sum_{n=j}^{\infty}\frac{1}{n^2}}_{\le \frac{2}{j}}\ \underbrace{\int\limits_{A_j} X_1^2\, dP}_{\le j\cdot\int\limits_{A_j}|X_1|\, dP}\Big) \le 2\cdot\sum_{j=1}^{\infty}\int\limits_{A_j}|X_1|\, dP = 2\cdot E|X_1| < \infty\end{aligned}$$

(ii) Es ist

$$EU_n = \int x\cdot I_{[-n,n]}\, dP^{X_1}(x) \overset{\text{beschr. Konv.}}{\longrightarrow} \int x\, dP^{X_1}(x) = EX_1.$$

Damit folgt $\frac{1}{n}\sum\limits_{i=1}^{n} EU_i \to EX_1$ (allg.: $a_n \to a \Rightarrow \frac{1}{n}\sum\limits_{i=1}^{n} a_i \to a$).

(iii) Es ist

$$\begin{aligned}\sum_{n=1}^{\infty} P(U_n \neq X_n) &= \sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty}\sum_{i=n}^{\infty} P(i < |X_1| \le i+1) = \\ &= \sum_{i=1}^{\infty}\sum_{n=1}^{i} P(i < |X_1| \le i+1) = \sum_{i=1}^{\infty} i\cdot P(i < |X_1| \le i+1) = \\ &= \sum_{i=1}^{\infty} i\cdot P^{|X_1|}\big((i, i+1]\big) \le E|X_1| < \infty\end{aligned}$$

Damit ist Teil (c) des Satzes 3.17 bewiesen.

Eine Art Umkehrung des starken Gesetzes der großen Zahlen (StGGZ) ist der folgende Satz.

**Satz 3.22:**
Sei $(X_n)_{n\in\mathbb{N}}$ eine stochastisch unabhängige identisch verteilte Folge von reellwertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Falls

$$S_n^* := \frac{1}{n} S_n := \frac{1}{n} \sum_{i=1}^{n} X_i \to X[P]$$

für eine reellwertige Zufallsvariable $X$ auf $(\Omega, \mathcal{A}, P)$, dann sind $X_n$ $P$-integrierbar für alle $n \in \mathbb{N}$ und $X = EX_1[P]$.

Beweis: Da $S_n^* \to X[P]$ nach Voraussetzung, gilt

$$\frac{1}{n} X_n = S_n^* - \frac{n-1}{n} S_{n-1}^* \to 0[P].$$

Betrachte $C_n := \{|X_n| \geq n\}$. Dann kann $C_n$ nur mit Wahrscheinlichkeit 1 für unendlich viele $n$ auftreten, d.h. $P(\operatorname{limsup} C_n) = 0$. Mit Satz 3.7 (b) (Borel-Cantelli) folgt $\sum\limits_{n=1}^{\infty} P(C_n) < \infty$. Da $(X_n)$ eine stid-Folge ist, gilt

$$P(C_n) = P\big(|X_k| > n\big) \qquad \forall\, k, n.$$

Also

$$\sum_{n=1}^{\infty} P(C_n) = \sum_{n=1}^{\infty} P\big(|X_k| \geq n\big) < \infty.$$

Mit $A_j := (j+1, j]$ für $j \in \mathbb{N}$ ist aber

$$\begin{aligned} \infty \;&>\; \sum_{n=1}^{\infty} P\big(|X_k| \geq n\big) = \sum_{n=1}^{\infty} \sum_{j>n} P\big(|X_k| \in A_j\big) = \sum_{j=1}^{\infty} (j-1) \cdot P\big(|X_k| \in A_j\big) = \\ &=\; \sum_{j=1}^{\infty} j \cdot P\big(|X_k| \in A_j\big) - \underbrace{\sum_{j=1}^{\infty} P\big(|X_k| \in A_j\big)}_{=1} \geq \Big(\sum_{j=1}^{\infty} \int_{A_j} |x| \; dP^{X_k}(x)\Big) - 1 = E|X_k| - 1, \end{aligned}$$

also ist $X_k$ $P$-integrierbar. Aus Satz 3.17 folgt sofort $S_n^* = \frac{1}{n} \sum\limits_{i=1}^{n} X_i \to EX_1[P]$. Nach Voraussetzung aber auch $S_n^* \to X[P]$, also $X = EX_1[P]$.

**Bemerkung 3.23:**
Zusammen mit der Rechnung im Beweis von Satz 3.17 (c) Teil (iii) ist damit für jede reellwertige integrierbare Zufallsvariable gezeigt:

$$\sum_{n=1}^{\infty} P\big(|X| \geq n\big) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P\big(|X| \geq n\big) \quad \text{bzw.} \quad X \text{ integrierbar} \iff \sum_{n=1}^{\infty} P\big(|X| \geq n\big) < \infty$$

**Corollar 3.24:**
Seien $(T_n)_{n\in\mathbb{N}}$ eine stochastisch unabhängige identisch verteilte Folge von $(\mathcal{T}, \mathcal{D})$-wertigen Zufallsvariablen sowie $g : (\mathcal{T}, \mathcal{D}) \to (\mathbb{R}, \mathbb{B})$ mit $E(g \circ T_1) < \infty$. Dann gilt

$$\frac{1}{n} \sum_{i=1}^{n} g(T_i) \to E(g \circ T_1)[P].$$

**Beispiel:**
Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig und Cauchy-verteilt. Dann ist $EX_i$ nicht definiert und

$$\overline{X}_\cdot^{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ist wieder Cauchy-verteilt. Es gilt

$$\neg\big[(\overline{X}_\cdot^{(n)} - \alpha) \to 0[P]\big].$$

**Bemerkung 3.26 (Folgerung von Satz 3.17):**

(a) $(X_n)_{n\in\mathbb{N}}$ stid., $EX_1 = \mu \Rightarrow \frac{1}{n}\sum\limits_{i=1}^{n} X_i \to \mu[P]$,

(b) $(X_n)_{n\in\mathbb{N}}$ stu., $EX_n = \mu_n$, $\operatorname{Var} X_n = \sigma_n^2 < \infty$ und $\sum\limits_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$. Dann gilt

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mu_i \to 0[P].$$

Gilt zusätzlich $\frac{1}{n}\sum\limits_{i=1}^{n} \mu_i \to \mu \in \mathbb{R}$, so gilt $\frac{1}{n}\sum\limits_{i=1}^{n} X_i \to \mu[P]$.

Zum Abschluß dieses Paragraphen noch einige Anwendungen zum starken Gesetz der großen Zahlen (StGGZ).

**Beispiel:**
Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig identisch verteilt, $X_n : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B})$ und $B \in \mathcal{B}$ vorgegeben. Schätzer für $P(X_1 \in B)$:

$$\frac{1}{n}\sum_{i=1}^{n} I_B\big(X_i(\omega)\big) = \frac{1}{n}\big(\{i \le n :\ X_i(\omega) \in B\}\big) \mathrel{\hat{=}} \text{relative Häufigkeiten}$$

Es gilt

$$\frac{1}{n}\sum_{i=1}^{n} I_B(X_i) \to P(X_1 \in B)[P],$$

denn $Y_i := I_B(X_i) \sim B(1,p)$ stid. mit $p = P(X_1 \in B)$. Daraus folgt die Behauptung mit dem StGGZ.

**Beispiel 3.27:**
Seien $(X_n)_{n\in\mathbb{N}}$ stid. und reellwertig. Dann heißt

$$\hat{F}_n(x,\omega) := \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}\big(X_i(\omega)\big) \qquad (x \in \mathbb{R})$$

empirische Verteilungsfunktion. $\hat{F}_n(x) := \hat{F}(x,\cdot) : \Omega \to \mathbb{R}$ ist eine Zufallsvariable für $x \in \mathbb{R}$. $\hat{F}_n(\cdot,\omega)$ ist monoton wachsend, rechtsseitig stetig und es gilt

$$\hat{F}_n(x,\omega) \overset{x\to\infty}{\longrightarrow} 1, \qquad \hat{F}_n(x,\infty) \overset{x\to-\infty}{\longrightarrow} 0 \qquad \forall\, \omega \in \Omega,$$

d.h. $\hat{F}_n(\cdot,\omega)$ ist Verteilungsfunktion für alle $\omega \in \Omega$. Wegen Beispiel 3.26 gilt für alle $x \in \mathbb{R}$

$$\hat{F}_n(x) \to \underbrace{P(X_1 \le x)}_{=F(x)}[P]$$

mit $F$ Verteilungsfunktion von $P^{X_1}$. Es gilt sogar stärker der folgende Satz:

**Satz 3.28 (Satz von Glirenko-Cantelli, Zentralsatz der Statistik):**
Seien $(X_n)_{n\in\mathbb{N}}$ stid. und reellwertig, $F(x) = P(X_1 \le x)$, $x \in \mathbb{R}$. Dann gilt

$$\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| \to 0[P].$$

**Beispiel 3.29 (Erinnerungsprozesse):**
Seien $(X_n)_{n\in\mathbb{N}}$ stid. und $X_n \ge 0$, z.B. wenn $X_n$ der Lebensdauer der $n$-ten Maschine entspricht. Frage: Wieviel Erneuerungen hat es bis zu einem Zeitpunkt $t$ gegeben? Setze $N_t(\omega)$ gleich der Anzahl der Erneuerungen bis einschließlich $t$ und

$$S_j(\omega) := \sum_{i=1}^{j} X_i(\omega), \quad S_0 := 0.$$

Dann gilt $\{\omega : N_t(\omega) = n\} = \{\omega : S_n(\omega) \le t < S_{n+1}(\omega)\}$ und somit Gleichheit der Ereignisse $\{N_t < m\} = \{S_m > t\}$. Dabei heißt $(N_t)_{t\ge 0}$ Erneuerungsprozeß. Falls diese $X_i \sim \mathrm{EXP}(\lambda)$, so heißt $(N_t)_{t\ge 0}$ Poisson-Prozeß, dann sind $X_i \sim \mathrm{EXP}(\lambda)$, dann ist $\sum_{i=1}^{m} X_i \sim \Gamma\left(m, \frac{1}{\lambda}, 0\right)$ und somit

$$\begin{aligned} P(S_m > t) &= 1 - P\Big(\sum_{i=1}^{m} X_i \le t\Big) = 1 - \int_0^t \frac{\lambda^m}{\Gamma(m)} e^{m-1} e^{-\lambda x}\, dx \overset{y=\lambda x}{=} \\ &= 1 - \int_0^{\lambda t} \frac{\lambda^m}{\Gamma(m)} \left(\frac{y}{\lambda}\right)^{m-1} e^{-y} \frac{1}{\lambda}\, dy = 1 - \int_0^{\lambda t} \frac{1}{\Gamma(m)} y^{m-1} e^{-\lambda}\, dy \overset{\text{part. Int.}}{=} \\ &= \sum_{j=0}^{m-1} \underbrace{\frac{(\lambda t)^j}{j!} e^{-\lambda t}}_{=P(Y_j)} \end{aligned}$$

mit $Y \sim \mathrm{Po}(\lambda t)$. Allgemein gilt (vorausgesetzt $P(X_n = \infty) = 0$):

$$\lim_{t\to\infty} N_t = \infty[P].$$

Sei $0 < EX_1 < \infty$, $\mu = EX_1$. Dann gilt mit dem StGGZ: $\frac{1}{n} S_n \to \mu[P]$. Da auch $N_t \to \infty[P]$, gilt

$$\frac{1}{N_t} S_{N_t} \overset{t\to\infty}{\longrightarrow} \mu[P].$$

Betrachte jetzt

$$\underbrace{\frac{S_{N_t}}{N_t}}_{\to\mu[P]} \le \frac{t}{N_t} < \frac{S_{N_t+1}}{N_t} = \underbrace{\frac{S_{N_t+1}}{N_t+1}}_{\to\mu[P]} \underbrace{\frac{N_t+1}{N_t}}_{\to 1}.$$

Also gezeigt: $\lim_{t\to\infty} \frac{t}{N_t} = \mu[P]$.

## 3.4 Schwache Konvergenz und zentraler Grenzwertsatz (ZGWS)

**Problemstellungen:**

(1) Seien $P_n$, $n \in \mathbb{N}$, $P_0$ Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A}, P)$. Was ist eine sinnvolle Definition von $P_n \overset{n\to\infty}{\longrightarrow} P_0$, welche Charakterisierungen hat diese Konvergenz?

(2) Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängige Zufallsvariablen und $S_n := \sum_{i=1}^{n} X_i$.

(a) Zentrales Grenzwertproblem: Unter welchen Bedingungen existieren $a_n > 0$, $b_n \in \mathbb{R}$, Wahrscheinlichkeitsmaß $P_0$ über $(\mathbb{R}, \mathbb{B})$, wobei $P_0$ kein Dirac-Maß, so daß

$$P^{\frac{S_n - b_n}{a_n}} \to P_0 \ ?$$

(b) Wie sehen mögliche $P_0$'s in (a) aus?

(c) Wann gilt $P_0 = \mathrm{N}(0,1)$?

Die in (2) angesprochenen Probleme sind alle gelöst. Hier wird nur ein kleiner Ausschnitt behandelt, nämlich (2) (c).

Zunächst zur Frage nach einer sinnvollen Definition von $P_n \to P_0$.

**Beispiel 3.30:**
Gegeben seien Wahrscheinlichkeitsmaße $P_n = \delta_{\frac{1}{n}}$ auf $(\mathbb{R}, \mathbb{B})$, d.h.

$$\forall\, n \in \mathbb{N}: \quad \forall\, A \in \mathbb{B}: \quad P_n(A) = \begin{cases} 1 & \frac{1}{n} \in A \\ 0 & \frac{1}{n} \notin A \end{cases}$$

Wenn $P_n \to P_0$, so sollte $P_0 = \delta_0$ sein. Würde man jedoch definieren

$$P_n \to P_0 \quad :\Longleftrightarrow \quad \forall\, A \in \mathbb{B}: \quad P_n(A) \to P_0(A), \qquad (*)$$

so wäre dies im Beispiel nicht erfüllt:

$$A = (-\infty, 0] \quad \Rightarrow \quad \forall\, n \in \mathbb{N}: \quad P_n(A) = 0,$$

aber $P_0(A) = 1$. Die Bedingung $(*)$ und selbst die schwächere Forderung an die zugehörigen Verteilungsfunktionen

$$\forall\, x \in \mathbb{R}: \quad F_n(x) \to F_0(x)$$

scheinen also zu stark. Als sinnvoll erweist sich die Forderung

$$F_n(x) \to F_0(x) \quad \forall\, x \in \mathrm{CP}(F_0) := \{x \in \mathbb{R}:\ F_0 \text{ stetig in } x\}. \qquad (**)$$

Es ist $(**)$ im obigen Beispiel offensichtlich erfüllt. Man spricht dann von Verteilungskonvergenz bzw. schwacher Konvergenz.

Da eine Reihe von zu $(**)$ äquivalenten Forderungen existieren, gibt es natürlich auch mehrere Möglichkeiten, „schwache Konvergenz“ zu definieren. Der Einfachheit halber werden nur Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathbb{B})$ betrachtet.

**Bezeichnung:**
Wir bezeichnen indexDefinition!Cb(R)@$\mathrm{C}^{\mathrm{b}}(\mathbb{R})$

$$\mathrm{C}^{\mathrm{b}}(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{R}:\ f \text{ stetig und beschränkt} \right\}.$$

Dabei gilt

$$f \text{ beschränkt} \quad \Longleftrightarrow \quad \sup_{x\in\mathbb{R}} |f(x)| < M < \infty.$$

**Definition 3.31:**

(a) Gegeben Wahrscheinlichkeitsmaße $P_n$, $n \in \mathbb{N}$, $P_0$ auf $(\mathbb{R}, \mathbb{B})$. Dann heißt $(P_n)_{n\in\mathbb{N}}$ schwach konvergent gegen $P_0$ [kurz: $P_n \overset{\text{w}}{\to} P_0$, w $\sim$ weak] $:\Longleftrightarrow$

$$\forall\, f \in \mathrm{C}^{\mathrm{b}}(\mathbb{R}): \quad \lim_{n\to\infty} \int f\, dP_n = \int f\, dP_0.$$

(b) Gegeben Wahrscheinlichkeitsräume $(\Omega_n, \mathcal{A}_n, P_n)$ und Zufallsvariablen $X_n : (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}, \mathbb{B})$ für $n \in \mathbb{N}_0$. $(X_n)_{n\in\mathbb{N}}$ heißt konvergent in Verteilung gegen $X_0 :\Longleftrightarrow P^{X_n} \overset{\text{w}}{\to} P^{X_0}$, d.h.

$$\forall\, f \in \mathrm{C}^{\mathrm{b}}(\mathbb{R}): \quad \lim_{n\to\infty} \int f\, dP^{X_n} = \int f\, dP^{X_0},$$

kurz: $X_n \overset{\text{w}}{\to} X_0$.

Mit

$$\begin{aligned} \mathrm{C}^{(\mathrm{v})}(\mathbb{R}) &= \{f \in \mathrm{C}^{\mathrm{b}}(\mathbb{R}):\ f \text{ gleichmäßig stetig auf } \mathbb{R}\} \\ \text{und } \mathrm{C}^{(\mathrm{r})}(\mathbb{R}) &= \{f \in \mathrm{C}^{\mathrm{b}}(\mathbb{R}):\ f \text{ } r\text{-mal differenzierbar und } f^{(i)} \in \mathrm{C}^{(\mathrm{v})}(\mathbb{R}),\ i = 0, 1, \ldots, r\} \end{aligned}$$

zeigen wir folgenden Satz:

**Satz 3.32:**
Unter den Voraussetzungen von Definition 3.31 (b) sind für jedes feste $r \in \mathbb{N}_0$ folgende Aussagen äquivalent:

(a) $X_n \overset{\text{w}}{\to} X_0$,

(b) $\forall\, f \in \mathrm{C}^{(\mathrm{r})}(\mathbb{R}): \quad \lim_{n\to\infty} \int f\, dP_n^{X_n} = \int f\, dP_0^{X_0}$,

(c) Für die Verteilungsfunktion $F_n(x) = P_n(X \le x)$, $n \in \mathbb{N}_0$, gilt:

$$\forall\, x \in \mathrm{CP}(F_0): \quad \lim_{n\to\infty} F_n(x) = F_0(x).$$

Bemerkung: Die Bedingung (b) wird sich beim Beweisen des Zentralen Grenzwertsatzes (ZGWS) als nützlich erweisen.

Beweis:

(a) $\Rightarrow$ (b): klar, da $\mathrm{C}^{(\mathrm{r})}(\mathbb{R}) \subset \mathrm{C}^{\mathrm{b}}(\mathbb{R})$.

(b) $\Rightarrow$ (c): Sei $x \in \mathrm{CP}(F_0) \Rightarrow \forall\, \epsilon > 0:\ \exists\, \delta > 0:\ \forall\, y \in [x-\delta, x+\delta]$ gilt

$$\big|F_0(x) - F_0(y)\big| < \epsilon.$$

Wähle $\underline{f}, \overline{f} \in \mathrm{C}^{(\mathrm{r})}(\mathbb{R})$, so daß

$$\underline{f}(y) = \begin{cases} 1 & \text{falls } y \le x - \delta \\ 0 & \text{falls } y \ge x \end{cases} \qquad \text{und} \qquad \overline{f}(y) = \begin{cases} 1 & \text{falls } y \le x \\ 0 & \text{falls } y \ge x + \delta \end{cases}$$

Es folgt

$$\begin{aligned} \overline{\lim}\, F_n(x) &= \overline{\lim} \int I_{\{X_n \le x\}}(t)\, dP_n^{X_n}(t) \le \\ &\le \overline{\lim} \int \overline{f}(t)\, dP_n^{X_n}(t) \overset{\text{(b)}}{=} \int \overline{f}(t)\, dP_0^{X_0}(t) \le \\ &\le \int I_{\{X_0 \le x+\delta\}}\, dP_0^{X_0} = F_0(x+\delta) \le F_0(x) + \epsilon \end{aligned}$$

und

$$\begin{aligned}\underline{\lim}\, F_n(x) &= \underline{\lim} \int I_{\{X_n \le x\}}\, dP_n^{X_n} \ge \underline{\lim} \int \underline{f}\, dP_n^{X_n} \overset{\text{(b)}}{=} \int \underline{f}\, dP_0^{X_0} \ge \\ &\ge \int I_{\{X_0 \le x-\delta\}}\, dP_0^{X_0} = F_0(x-\delta) \ge F_0(x) - \epsilon\end{aligned}$$

Da $\epsilon > 0$ beliebig, folgt $\lim\limits_{n\to\infty} F_n(x) = F_0(x)$.

(c) $\Rightarrow$ (a): Für $f \in \mathrm{C}^{\mathrm{b}}(\mathbb{R})$ ist zu zeigen, daß $\int f\, dP_n^{X_n} \to \int f\, dP_0^{X_0}$. Sei also $f \in \mathrm{C}^{\mathrm{b}}(\mathbb{R})$. Wähle $M > 0$, so daß $M, -M \in \mathrm{CP}(F_0)$ und $F_0(-M) < \epsilon$ sowie $F_0(M) > 1-\epsilon$ ($\epsilon \in (0, \frac{1}{2})$). Dann ist $f$ gleichmäßig stetig auf $[-M, M]$. Ferner existiert eine Folge $t_0, t_1, \ldots, t_m$ mit $-M = t_0 < t_1 < \ldots < t_m < M$, so daß

$$\sup\Big\{\big|f(x) - f(y)\big| :\ x, y \in [t_{j-1}, t_j]\Big\} < \epsilon \qquad \forall\, j = 1, \ldots, m.$$

Dabei seien (o.E.) $t_0, \ldots, t_m$ Stetigkeitspunkte von $F_0$, d.h.

$$F_n(t_j) \to F_0(t_j) \qquad \forall\, j = 0, \ldots, m.$$

Es folgt

$$\begin{aligned}\left|\int f\, dP_n^{X_n} - \int f\, dP_0^{X_0}\right| &= \Big|\int\limits_{\{X_n \le -M\}\cup\{X_n > M\}} f\circ X_n\, dP_n - \int\limits_{\{X_0 \le -M\}\cup\{X_0 > M\}} f\circ X_0\, dP_0 + \\ &\quad + \int\limits_{\{-M < X_n \le M\}} f\circ X_n\, dP_n - \int\limits_{\{-M < X_0 \le M\}} f\circ X_0\, dP_0 \le \\ &\le \|f\|_\infty \Big(\underbrace{F_n(-M) + 1 - F_n(M) + F_0(-M) + 1 - F_0(M)}_{\to \underbrace{2[F_0(-M)+1-F_0(M)]}_{<4\epsilon}}\Big) + \\ &\quad + \sum_{j=1}^{m} \underbrace{\Big|\int\limits_{\{t_{j-1} < X_n \le t_j\}} f\circ X_n\, dP_n - \int\limits_{\{t_{j-1} < X_0 \le t_j\}} f\circ X_0\, dP_0\Big|}_{=:\oplus}\end{aligned}$$

Mit $c_j := f\left(\frac{t_{j-1}+t_j}{2}\right)$ und $B_{n,j} := \{t_{j-1} < X_n \le t_j\}$ für $n \in \mathbb{N}_0$ ist

$$\begin{aligned}\oplus &= \Big|\int\limits_{B_{n,j}} f\circ X_n\, dP_n - \int\limits_{B_{0,j}} f\circ X_0\, dP_0\Big| \le \\ &\le \Big|\int\limits_{B_{n,j}} (f\circ X_n - c_j I_{\Omega_n})\, dP_n - \int\limits_{B_{0,j}} (f\circ X_0 - c_j I_{\Omega_0})\, dP_0\Big| + \\ &\quad + |c_j(F_n(t_j) - F_n(t_{j-1})) - c_j(F_0(t_j) - F_0(t_{j-1}))| \le \\ &\le \epsilon \cdot \underbrace{P_n(B_{n,j})}_{F_n(t_j)-F_n(t_{j-1})} + \epsilon \cdot \underbrace{P_0(B_{0,j})}_{F_0(t_j)-F_0(t_{j-1})} + |c_j|\Big(|\underbrace{F_n(t_j) - F_n(t_{j-1})}_{\to 0}| + |\underbrace{F_0(t_j) - F_0(t_{j-1})}_{\to 0}|\Big) \to \\ &\to 2 \cdot \epsilon \cdot P_0(B_{0,j})\end{aligned}$$

Insgesamt also

$$\overline{\lim}\left|\int f\, dP_n^{X_n} - \int f\, dP_0^{X_0}\right| \le 4\epsilon\|f\|_\infty + 2\epsilon \underbrace{\sum_{j=1}^{m} P_0(B_{0,j})}_{\le 1} \le \epsilon(4\|f\|_\infty + 2)$$

Mit $\epsilon \to 0$ folgt die Behauptung.

**Bemerkung 3.33:**
Schwache Konvergenz von Wahrscheinlichkeitsmaßen kann man natürlich auf allgemeine Wahrscheinlichkeitsräume definieren: Sind $(\mathcal{X}, d)$ ein metrischer Raum und $P_n$, $n \in \mathbb{N}$, Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{B})$ mit $\mathcal{B}$ Borel-$\sigma$-Algebra über $\mathcal{X}$, so definiert man

$$(*) \qquad P_n \overset{\mathrm{w}}{\to} P_0 \quad :\Longleftrightarrow \quad \forall\, B \in \mathcal{B} \text{ mit } P(\partial B) = 0 : \quad P_n(B) \to P_0(B)$$

Äquivalent zu $(*)$ ist das sogenannte Portmanteau-Theorem:

(a) $\forall\, F \in \mathcal{B}$, $F$ abgeschlossen: $\overline{\lim}\, P_n(F) \leq P_0(F)$,

(b) $\forall\, G \in \mathcal{B}$, $G$ offen: $\underline{\lim}\, P_n(G) \geq P_0(G)$,

(c) $\forall\, f : \mathcal{X} \to \mathbb{R}$ stetig und beschränkt: $\overline{\lim} \int f\, dP_n = \int f\, dP_0$ und

(d) $\forall\, f : \mathcal{X} \to \mathbb{R}$ (glm.) stetig und beschränkt: $\lim \int f\, dP_n = \int f\, dP_0$.

(siehe Billingsley, 1968)

**Beispiel 3.34:**

(a) Seien $x_n \in \mathbb{R}$ für $n \in \mathbb{N}$. Dann gilt

$$x_n \to x_0 \quad \Longleftrightarrow \quad \delta_{x_n} \overset{\mathrm{w}}{\to} \delta_{x_0}.$$

(b) Siehe Satz 1.47 (Poisson'scher Grenzwertsatz): $P_n = \mathrm{B}(n, p_n)$, $p_n = \frac{\lambda}{n}$ mit $\lambda > 0$, $n \in \mathbb{N}$ und $P_0 = \mathrm{Po}(\lambda)$. Dann gilt

$$\forall\, k \in \mathbb{N}_0 : \quad P_n(\{k\}) \to P_0(\{k\}),$$

d.h.

$$\forall\, x \in \mathbb{N}_0 : \quad F_{\mathrm{B}(n,p_n)}(x) \to F_{\mathrm{Po}(\lambda)}(x).$$

Somit gilt sogar

$$\forall\, x \in \mathbb{N} : \quad F_{\mathrm{B}(n,p_n)}(x) \to F_{\mathrm{Po}(\lambda)}(x) \Rightarrow P_n \overset{\mathrm{w}}{\to} P_0.$$

Wir gehen noch kurz auf den Zusammenhang zwischen stochastischer und schwacher Konvergenz ein.

**Lemma 3.35:**
Gegeben Wahrscheinlichkeitsräume $(\Omega, \mathcal{A}, P)$, $X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$, $n \in \mathbb{N}_0$:

(a) Falls $X_n \overset{P}{\to} X_0$, folgt $X_n \overset{\mathrm{w}}{\to} X_0$. Die Umkehrung ist i.a. falsch, es gilt jedoch

(b) Falls $X_0 = \mathrm{const}[P]$ und $X_n \overset{\mathrm{w}}{\to} X_0$, so folgt $X_n \overset{P}{\to} X_0$.

Beweis:

(a) Es gilt: $\forall\, \epsilon \in (0,1) : \exists\, M > 0 : \; P\big(|X_0| \leq M\big) > 1 - \epsilon$. Da $f \in \mathrm{C}^{\mathrm{b}}(\mathbb{R})$, ist $f$ gleichmäßig stetig auf $[-2M, 2M]$, d.h. $\exists\, \delta > 0$, so daß gilt:

$$|X_n|, |X_0| < 2M, \quad |X_n - X_0| \leq \delta \quad \Rightarrow \quad |f(X_n) - f(X_0)| \leq \epsilon.$$

Für diese $\delta$ existiert ein $n_0 \in \mathbb{N}$, so daß

$$\forall\, n \geq n_0 : \quad P\big(|X_n - X_0| > \delta\big) \leq \epsilon.$$

Es folgt für alle $n \geq n_0$:

$$\begin{aligned}
&\Big| \int f \, dP^{X_n} - \int f \, dP^{X_0} \Big| \leq \\
&\leq \Big| \int\limits_{\{|X_0|>M\}} (f \circ X_n - f \circ X_0) \, dP \Big| + \Big| \int\limits_{\{|X_0|\leq M\}} (f \circ X_n - f \circ X_0) \, dP \Big| \leq \\
&\leq 2\epsilon \|f\|_\infty + \Big| \int\limits_{\{|X_0|\leq M, \, |X_n - X_0|>\delta\}} (f \circ X_n - f \circ X_0) \, dP \Big| + \Big| \int\limits_{\{|X_0|\leq M, \, |X_n - X_0|\leq\delta\}} (f \circ X_n - f \circ X_0) \, dP \Big| \leq \\
&\leq 2\epsilon \|f\|_\infty + 2\epsilon \|f\|_\infty + \epsilon = \epsilon(4\|f\|_\infty + 1) \overset{\epsilon \to 0}{\longrightarrow} 0
\end{aligned}$$

Somit folgt die Behauptung, d.h. $X_n \overset{\mathrm{w}}{\to} X_0$.

Im Lemma 3.35 (a) wurde gezeigt: $X_n \overset{P}{\to} X_0 \Rightarrow X_n \overset{\mathrm{w}}{\to} X_0$. Daß die Umkehrung falsch ist, zeigt folgendes Beispiel:

**Beispiel 3.36:**
Sei $X_0 \sim \mathrm{B}\left(1, \frac{1}{2}\right)$ und $X_n = 1 - X_0$ für $n \in \mathbb{N}$. Dann ist $X_n \sim \mathrm{B}\left(1, \frac{1}{2}\right)$ für alle $n \in \mathbb{N}$, d.h. $P^{X_n} = P^{X_0}$ für alle $n \in \mathbb{N}$ und damit $X_n \overset{\mathrm{w}}{\to} X_0$, aber für alle $\epsilon \in (0,1)$ gilt

$$P\Big(\big|X_n - X_0\big| > \epsilon\Big) = P\Big(\big|2X_0 - 1\big| > \epsilon\Big) = 1 \qquad \forall \, n \in \mathbb{N}.$$

D.h. es gilt nicht $X_n \overset{P}{\to} X_0$. Dies zeigt auch, daß für $X_n \overset{\mathrm{w}}{\to} X_0$ zwar $P^{X_0}$ eindeutig bestimmt ist während $X_0$ nicht eindeutig bestimmt ist. [ Es gilt: $P = Q \iff \forall \, f \in \mathrm{C^b}(\mathbb{R}) : \int f \, dP = \int f \, dQ$ ]

Im folgenden behandeln wir das angekündigte zentrale Grenzproblem. Generelle Voraussetzungen seien jetzt: Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$, $(X_n)$ stochastisch unabhängige reellwertige Zufallsvariablen, also

$$X_n : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$$

sowie $\sum\limits_{i=1}^{n} X_i =: S_n$.

**Definition 3.37:**
Sind die $X_n$ quadratintegrierbar mit $\mathrm{Var}\, X_n = \sigma_n^2 > 0$, so gilt für die Folge $(X_n)_{n \in \mathbb{N}}$ der zentrale Grenzwertsatz (ZGWS), falls die sogenannten standardisierten Summenvariablen

$$T_n := \frac{S_n - ES_n}{\sqrt{\mathrm{Var}\, S_n}} \overset{\mathrm{w}}{\to} X_0,$$

also schwach gegen $X_0$ konvergieren, wobei $X_0 \sim \mathrm{N}(0,1)$, bzw. $P^{T_n} \overset{\mathrm{w}}{\to} P^{X_0}$.

**Bemerkung 3.38:**

(a) Beachte: $ET_n = 0$ und $\mathrm{Var}\, T_n = 1$.

(b) Ist $(X_n)_{n \in \mathbb{N}}$ eine stid-Folge mit $EX_i = \mu$ und $\mathrm{Var}\, X_i = \sigma^2 > 0$, so ist

$$T_n = \frac{\Big(\sum\limits_{i=1}^{n} X_i\Big) - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}.$$

(c) Gilt in (b) zusätzlich $X_i \sim \mathrm{B}(1,p)$ für $p \in (0,1)$, so ist

$$T_n = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Für den Spezialfall $p = \frac{1}{2}$ also $T_n = \frac{2}{\sqrt{n}}\left(S_n - \frac{n}{2}\right)$. Hierfür wurde der zentrale Grenzwertsatz (ZGWS) zuerst gezeigt:

**Zentraler Grenzwertsatz (ZGWS) von de Moivre-Laplace:**
Es gilt

$$S_n \sim \mathrm{B}(n,p) \quad \Rightarrow \quad P^{T_n} \xrightarrow{\mathrm{w}} \mathrm{N}(0,1).$$

Der zentrale Grenzwertsatz von de Moivre-Laplace wird sich als Spezialfall des Zentralen Grenzwertsatzes von Lindeberg ergeben. Die entscheidende Bedingung für die Gültigkeit des zentralen Grenzwertsatzes ist die Lindeberg-Bedingung:

**Lindeberg-Bedingung (Lindeberg, 1876-1932):**
Es sei $\tau := \sqrt{\mathrm{Var}\ S_n}$ und für alle $\epsilon > 0$ gelte

$$\forall\, n \in \mathbb{N}: \qquad L_n(\epsilon) := \frac{1}{\tau_n^2} \sum_{j=1}^{n} \int\limits_{\{|X_j - EX_j| > \epsilon\tau_n\}} (X_j - EX_j)^2\, dP.$$

$(X_n)_{n\in\mathbb{N}}$ genügt der Lindeberg-Bedingung, falls

$$\text{(L)} \qquad \forall\, \epsilon > 0: \quad \lim_{n\to\infty} L_n(\epsilon) = 0.$$

Die Lindeberg-Bedingung (L) impliziert, daß die Varianzen der Summanden, also Var $X_j$, gleichmäßig klein gegenüber der Varianz der Summe wird (d.h. $X_j$ hat gegenüber den restlichen $X_i$ in $S_n$ keinen dominierenden Einfluß). Um dies einzusehen, führen wir die Feller-Bedingung ein:

**Feller-Bedingung:**
$(X_n)_{n\in\mathbb{N}}$ genügt der Feller-Bedingung, falls

$$\text{(F)} \qquad \max_{1\le j\le n} \frac{\mathrm{Var}\ X_j}{\tau_n^2} \xrightarrow{n\to\infty} 0.$$

**Lemma 3.39:**
Sind die $X_n$ quadratintegrierbar und stochastisch unabhängig, so gilt

$$\text{(L)} \quad \Rightarrow \quad \text{(F)} \quad \Rightarrow \quad \lim_{n\to\infty} \tau_n^2 = \infty.$$

Beweis: Sei $\epsilon > 0$, $Y_j := X_j - EX_j$ für $j \in \mathbb{N}$ und sei $n \in \mathbb{N}$. Dann gilt für alle $j = 1, \ldots, n$:

$$\begin{aligned} \mathrm{Var}\ X_j &= EY_j^2 = \int\limits_{\{|Y_j|\le\epsilon\tau_n\}} Y_j^2\, dP + \int\limits_{\{|Y_j|>\epsilon\tau_n\}} Y_j^2\, dP \le \\ &\le \epsilon^2\tau_n^2 + \sum_{j=1}^{n} \int\limits_{\{|Y_j|>\epsilon\tau_n\}} Y_j^2\, dP = \epsilon^2\tau_n^2 + \tau_n^2 L_n(\epsilon) \end{aligned}$$

Also gilt:

$$\max_{1\le j\le n} \frac{\mathrm{Var}\ X_j}{\tau_n^2} \le \epsilon^2 + L_n(\epsilon).$$

Damit folgt die Behauptung (L) $\Rightarrow$ (F). Also ist noch der zweite Teil der Behauptung zu zeigen. Es gilt: Da die $X_n$ stochastisch unabhängig sind, ist $\tau_n^2 = \sum\limits_{j=1}^{n} \mathrm{Var}\ X_j$ wachsend in $n$.

Nehmen wir an, daß $a := \sup\limits_{n\in\mathbb{N}} \tau_n^2 < \infty$, dann gilt für alle $n \in \mathbb{N}$: $\frac{1}{\tau_n^2} \ge \frac{1}{a}$. Somit gilt für alle $n \in \mathbb{N}$:

$$\underbrace{\max_{1\le j\le n} \frac{\mathrm{Var}\ X_j}{\tau_n^2}}_{\xrightarrow{n\to\infty} 0 \text{ wegen (F)}} \ge \frac{1}{a} \max_{1\le j\le n} \mathrm{Var}\ X_j.$$

Also folgt $\max\limits_{1\le j\le n} \operatorname{Var} X_j \overset{n\to\infty}{\longrightarrow} 0$. Da $\max\limits_{1\le j\le n} \operatorname{Var} X_j$ wachsend in $n$, muß $\operatorname{Var} X_j = 0$ für alle $j \in \mathbb{N}$ gelten - ein Widerspruch.

Beachte: Für den Beweis von (L) $\Rightarrow$ (F) wurde nicht ausgenutzt, daß $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig sind.

**Satz 3.40 (Zentraler Grenzwertsatz von Lindeberg):**
Seien $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig quadratintegrierbar mit $\operatorname{Var} X_n > 0$ für alle $n \in \mathbb{N}$. Dann gilt:

$$\text{Ist (L) erfüllt, so gilt der ZGWS für } (X_n)_{n\in\mathbb{N}}.$$

Wir beweisen diesen Satz mit Hilfe des folgenden Lemma.

**Lemma 3.41:**
Seien $X, Y$ reellwertige Zufallsvariablen mit $EX = EY$, $EX^2 = EY^2 < \infty$, $E|Y|^3 < \infty$ und sei $f \in \mathrm{C}^{(3)}(\mathbb{R}) = \{f \in \mathrm{C}^{\mathrm{b}} : f \text{ 3-mal differenzierbar, } f^{(i)} \in \mathrm{C}^{\mathrm{b}},\ i = 1,2,3\}$. Dann existiert ein $c = c(f)$, so daß für alle $\epsilon > 0$ gilt

$$\sup_{z\in\mathbb{R}} \big|Ef(X+z) - Ef(Y+z)\big| \le c \cdot \Big[\epsilon \cdot EX^2 + \int\limits_{\{|X|>\epsilon\}} X^2\, dP + E|Y|^3\Big].$$

Beweis: Taylor-Entwicklung von $f$ liefert für $u, z \in \mathbb{R}$:

$$\begin{aligned} f(z+u) &= f(z) + uf'(z) + \frac{u^2}{2}f''(z) + R_2(z,u) = \\ &= f(z) + uf'(z) + R_1(z,u) \end{aligned}$$

mit

$$\begin{aligned} R_2(z,u) &= \frac{u^3}{6}f'''(z+\delta_2 u) \quad \text{für ein } \delta_2 = \delta_2(z,u) \in (0,1) \\ R_1(z,u) &= \frac{u^2}{2}f''(z+\delta_1 u) \quad \text{für ein } \delta_1 = \delta_1(z,u) \in (0,1) \end{aligned}$$

Dann gelten die Abschätzungen

$$|R_2(z,u)| \le \frac{u^3}{6}\|f'''\|_\infty \tag{1}$$

$$\begin{aligned} |R_2(z,u)| &\le |R_1(z,u) - \frac{u^2}{2}f''(z)| = \\ &= \left|\frac{u^2}{2}f''(z+\delta_1 u) - \frac{u^2}{2}f''(z)\right| \le u^2\|f''\| - \infty \end{aligned} \tag{2}$$

für alle $u, z \in \mathbb{R}$. Für große $u$ ist (2) besser, für kleine $u$ (1); so werden (1) und (2) auch eingesetzt. Sei $c = \max\left\{\frac{1}{6}\|f'''\|_\infty, \|f''\|_\infty\right\}$. Dann folgt für alle $z \in \mathbb{R}$ und alle $\epsilon > 0$:

$$\begin{aligned} |Ef(X+z) - Ef(Y+z)| &= \left|\int f(x+z)\, dP^X(x) - \int f(y+z)\, dP^Y(y)\right| \overset{\text{da } EX=EY}{\le} \\ &\le \left|\int f(x+z)\, dP^X(x) - f(z) - f'(z)EX - \frac{1}{2}f''(z)EX^2\right| + \\ &\quad + \left|f(z) + f'(z)EY + \frac{1}{2}f''(z)EY^2 - \int f(y+z)\, dP^Y(y)\right| = \\ &= \left|\int \left[f(x+z) - f(z) - xf'(z) - \frac{1}{2}x^2f''(z)\right] dP^X(x)\right| + \\ &\quad + \left|\int \left[f(z) + yf'(z) + \frac{1}{2}y^2f''(z) - f(y+z)\right] dP^Y(y)\right| \le \\ &\le \int \big|R_2(z,x)\big|\, dP^X(x) + \int \big|R_2(z,y)\big|\, dP^Y(y) \end{aligned}$$

Es ist $\int |R_2(z,y)|\, dP^Y(y) = \int |R_2(z,Y)|\, dP \le c \cdot E|Y|^3$ wegen (1). Dann ist für alle $\epsilon > 0$:

$$\begin{aligned}\int \left|R_2(z,x)\right| dP^X(x) &= \int \left|R_2(z,X)\right| dP = \\ &= \underbrace{\int_{\{|X|\le\epsilon\}} \left|R_2(z,X)\right| dP}_{\overset{(1)}{\le} c\int_{\{|X|\le\epsilon\}} |X|^3\, dP} + \underbrace{\int_{\{|X|>\epsilon\}} \left|R_2(z,X)\right| dP}_{\overset{(2)}{\le} c\int_{\{|X|>\epsilon\}} |X|^3\, dP} \le \\ &\le c\cdot\epsilon\cdot EX^2 + c\int_{\{|X|>\epsilon\}} X^2\, dP\end{aligned}$$

Es folgt die Behauptung.

**Beweis von Satz 3.40 (mit der Ersetzungsmethode von Lindeberg/Levy):**
Wegen Satz 3.32 genügt es, für alle $f \in \mathrm{C}^{(3)}(\mathbb{R})$ zu zeigen:

$$\int f(T_n)\, dP \to \int f\, dP_0 \quad \text{mit } P_0 = N(0,1).$$

Sei $f \in \mathrm{C}^{(3)}(\mathbb{R})$ und $X_{nj} := \frac{1}{\tau_n}(X_j - EX_j),\ 1 \le j \le n,\ n \in \mathbb{N}$. Dann gilt

$$T_n = \frac{S_n - ES_n}{\tau_n} = \sum_{j=1}^n X_{nj}$$

und für alle $n \in \mathbb{N}$: $X_{n1}, \ldots, X_{nn}$ stu. mit $EX_{nj} = 0$, $\operatorname{Var} X_{nj} = \frac{\operatorname{Var} X_j}{\tau_n^2}$. Für jedes $n \in \mathbb{N}$ wähle ein reelle Zufallsvariablen (auf $(\Omega, \mathcal{A}, P)$, evtl. nach geeigneter Erweiterung) $Y_{n1}, \ldots, Y_{nn}$ mit $P^{Y_{nj}} \sim \mathrm{N}(0, \operatorname{Var} X_{nj})$ und $X_{n1}, \ldots, X_{nn}, Y_{n1}, \ldots, Y_{nn}$ stochastisch unabhängig. Dann ist

$$\tilde{S}_n := \sum_{j=1}^n Y_{nj} \sim N(0,1),$$

da $1 = \sum\limits_{j=1}^n \operatorname{Var} X_{nj}$. Definiere weiter einen neuen Satz von Zufallsvariablen:

$$Z_{nj} := X_{n1} + \ldots + X_{n,j-1} + Y_{n,j+1} + \ldots + Y_{nn} \qquad \text{für } 1 \le j \le n,$$

speziell also $Z_{n1} = Y_{n2} + \ldots + Y_{nn}$ und $Z_{nn} = X_{n1} + \ldots + X_{n,n-1}$. Es folgt

$$\begin{aligned} Z_{nj} + X_{nj} &= Z_{n,j+1} + Y_{n,j+1} \qquad (1 \le j \le n-1) \\ Z_{nn} + X_{nn} &= T_n \quad \text{und} \\ Z_{n1} + Y_{n1} &= \tilde{S}_n \end{aligned}$$

Betrachte nun

$$\begin{aligned}\left|\int f(T_n)\, dP - \int f\, dP_0\right| &= \left|Ef(T_n) - Ef(\tilde{S}_n)\right| = \\ &= \underbrace{\left|\sum_{j=1}^n \left[\underbrace{Ef(Z_{nj}+X_{nj})}_{=:a_j} - \underbrace{Ef(Z_{nj}+Y_{nj})}_{=:b_j}\right]\right|}_{a_1-b_1+a_2-b_2+\ldots+a_n-b_n = a_n-b_1} \le \\ &\le \sum_{j=1}^n \left|Ef(Z_{nj}+X_{nj}) - Ef(Z_{nj}+Y_{nj})\right| =: (*)\end{aligned}$$

Da $Z_{nj}, X_{nj}$ stochastisch unabhängig sind, ebenso wie $Z_{nj}, Y_{nj}$, gilt

$$\begin{aligned}
&\left|Ef(Z_{nj}+X_{nj}) - Ef(Z_{nj}+Y_{nj})\right| = \\
&= \left|\int \Big[ \underbrace{\int f(z+x)\, dP^{X_{nj}}(x)}_{=Ef(z+X_{nj})} - \underbrace{\int f(z+y)\, dP^{Y_{nj}}(y)}_{=Ef(z+Y_{nj})} \Big]\, dP^{Z_{nj}}(z)\right| \leq \\
&\leq \sup_{z\in\mathbb{R}} \left|Ef(z+X_{nj}) - Ef(z+Y_{nj})\right|
\end{aligned}$$

Es folgt mit Lemma 3.41

$$\begin{aligned}
(*) \quad &\leq c\cdot\epsilon\cdot \underbrace{\sum_{j=1}^n EX_{nj}^2}_{=\operatorname{Var} T_n = 1} + c\cdot \underbrace{\sum_{j=1}^n \int_{\{|X_{nj}|>\epsilon\}} X_{nj}^2\, dP}_{=L_n(\epsilon)} + c\cdot\sum_{j=1}^n E|Y_{nj}|^3 = \\
&= c\epsilon + cL_n(\epsilon) + cE|X_0|^3\cdot\sum_{j=1}^n \left(\operatorname{Var} X_{nj}\right)^{\frac{3}{2}}
\end{aligned}$$

mit $X_0 \sim \mathrm{N}(0,1)$, denn wegen $Y_{nj} \sim \mathrm{N}(0, \operatorname{Var} X_{nj}$ folgt $\dfrac{1}{\sqrt{\operatorname{Var} X_{nj}}} Y_{nj} \sim \mathrm{N}(0,1)$ und damit

$$E|Y_{nj}|^3 = \left(\operatorname{Var} X_{nj}\right)^{\frac{3}{2}} E|X_0|^3.$$

Nun gilt

(i) Für alle $\epsilon > 0$ gilt $L_n(\epsilon) \to 0$ wegen Voraussetzung (L),

(ii) $E|X_0|^3 < \infty$, da $EX_0^4 < \infty$,

(iii) Da mit (L) auch (F) erfüllt ist, gilt

$$\sum_{j=1}^n \left(\operatorname{Var} X_{nj}\right)^{\frac{3}{2}} \leq \Big(\max_{1\leq j\leq n} \left(\operatorname{Var} X_{nj}\right)^{\frac{1}{2}}\Big) \cdot \underbrace{\sum_{j=1}^n \operatorname{Var} X_{nj}}_{=1} = \sqrt{\max_{1\leq j\leq n} \frac{\operatorname{Var} X_{nj}}{\tau_n^2}} \overset{n\to\infty}{\longrightarrow} 0.$$

Also geht gesamter abzuschätzender Ausdruck gegen 0, es folgt die Behauptung.

**Corollar 3.42 (ZGWS von Lindeberg/Levy):**
Ist $(X_n)_{n\in\mathbb{N}}$ eine stochastisch unabhängige identisch verteilte Folge von quadrat-integrierbaren Zufallsvariablen mit $EX_n = \mu$ und $\operatorname{Var} X_n = \sigma^2$, so gilt der zentrale Grenzwertsatz (ZGWS), d.h.

$$\frac{1}{\sqrt{n\sigma^2}} \sum_{j=1}^n (X_j - \mu) \overset{\mathrm{w}}{\to} X_0 \sim \mathrm{N}(0,1).$$

Die folgende nach A. U. Lyapunov (1857 - 1918) benannte Bedingung ist stärker als (L):

**Lyapunov-Bedingung:**
Für $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig und quadrat-integrierbar gilt die Lyapunov-Bedingung genau dann, wenn

$$\text{(Ly)} \qquad \exists\, \delta > 0 : \quad \lim_{n\to\infty} \tau^{-2-\delta} \sum_{j=1}^n E|X_j - EX_j|^{2+\delta} = 0.$$

**Corollar 3.43 (ZGWS von Lyapunov):**
Ist $(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängig und quadrat-integrierbar und gilt (Ly), so genügt $(X_n)_{n\in\mathbb{N}}$ dem zentralen Grenzwertsatz (ZGWS).

Alle bisher behaupteten zentralen Grenzwertsätze und Bedingungen, also (L), (F) und (Ly), lassen sich nicht nur für Folgen, sondern sogar für $\Delta$-Schemata von Zufallsvariablen beweisen bzw. formulieren. $(X_n)_{n\in\mathbb{N}}$ wird ersetzt durch $(X_{nj})_{j=1,\dots,k_n}$, $n\in\mathbb{N}$, $k_1 \le k_2 \le \dots$, wobei $X_{n1},\dots,X_{nk_n}$ jeweils als stochastisch unabhängig angenommen werden. So eine Situation liegt zum Beispiel beim Poisson'schen Grenzwertsatz vor:

$$X_{ni} := X_i \sim \mathrm{B}(1,p_n),\ i=1,\dots,n,\ \text{stu. mit } p_n = \frac{\lambda}{n},\ \lambda>0.$$

In den in (L), (F) und (Ly) auftretenden Größen ist nur $n$ durch $k_n$ zu ersetzen, dann $n\to\infty$, also zum Beispiel

$$S_n = \sum_{j=1}^{k_n} X_j, \quad \tau_n = \sqrt{\operatorname{Var} S_n}, \quad L_n(\epsilon) = \frac{1}{\tau_n^2}\sum_{j=1}^{k_n}\int \dots \qquad \text{(etc.)}$$

Außer dieser Verallgemeinerungen liefert der folgende Satz eine wesentlich genauere Charakterisierung der Gültigkeit eines zentralen Grenzwertsatzes (ZGWS). Auf den Beweis verzichten wir jedoch hier.

**Satz 3.44 (ZGWS von Lindeberg/Feller):**
Sind $X_{n1},\dots,X_{nk_n}$ für $n\in\mathbb{N}$ stochastisch unabhängige reellwertige quadrat-integrierbare Zufallsvariablen (mit $k_n\to\infty$), so sind äquivalent

(a) $(X_{nj})_{\substack{j=1,\dots,k_n\\ n\in\mathbb{N}}}$ erfüllt die Feller-Bedingung

$$\text{(F)} \qquad \max_{1\le j\le k_n} \frac{\operatorname{Var} X_{nj}}{\tau_n^2} \overset{n\to\infty}{\longrightarrow} \infty$$

und es gilt der zentrale Grenzwertsatz für $(X_{nj})$, d.h.

$$\frac{1}{\tau_n}\sum_{j=1}^{k_n}(X_{nj}-EX_{nj}) \overset{\mathrm{w}}{\to} X_0 \sim \mathrm{N}(0,1).$$

(b) Es gilt die Lindeberg-Bedingung, d.h.

$$\text{(L)} \qquad \forall\,\epsilon>0: \quad \lim_{n\to\infty}\frac{1}{\tau_n^2}\int\limits_{\{|x-EX_{nj}|\ge\epsilon\tau_n\}} (x-EX_{nj})^2\,dP^{X_{nj}}(x) = 0$$

**Beispiel 3.45:**
$X_i\sim\mathrm{B}(1,p)$, $i=1,\dots,n$, stochastisch unabhängig ($n$-facher Münzwurf)

$$S_n = \sum_{i=1}^{n} X_i \sim \mathrm{B}(n,p),$$

zum Beispiel $n=1000$, $p=\frac{1}{2}$. Was ist $P(a\le S_n\le b)$ für $a=450$ und $b=550$. Dazu

$$\begin{aligned} P(a\le S_n\le b) &= P\left(\frac{a-np}{\sqrt{np(1-p)}} \le \frac{S_n-np}{\sqrt{np(1-p)}} \le \frac{b-np}{\sqrt{np(1-p)}}\right) \overset{\text{ZGWS}}{\approx} \\ &\approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right) \overset{\text{Bsp.}}{=} \end{aligned}$$

$$= \Phi\left(\frac{550-500}{\sqrt{\frac{1}{4}\cdot 1000}}\right) - \Phi\left(\frac{450-500}{\sqrt{\frac{1}{4}\cdot 1000}}\right) \approx \Phi\left(\frac{50}{15.81}\right) - \Phi\left(-\frac{50}{15.81}\right) =$$

$$= 1 - 2\cdot\Phi\left(-\frac{50}{15.81}\right) \approx 1 - 2\cdot\Phi(-3.16) \approx 1 - 2\cdot 0.001 = 0.998$$

**Bemerkung 3.46:**
Ein Problem bei der Anwendung des zentralen Grenzwertsatzes ist die Konvergenzordnung. Man kann zeigen (Satz von Berry-Esseen)

(a) $(X_n)$ stu. mit $EX_n = 0$, $0 < \mathrm{Var} X_n < \infty$. Dann gilt

$$\sup_{x\in\mathbb{R}} \left|F_n(x) - \Phi(x)\right| \le \frac{6}{\tau_n^3} \sum_{k=1}^{n} E|X_k|^3,$$

wobei $F_n(x) := P(T_n \le x)$, $\Phi(x)$ Verteilungsfunktion der $\mathrm{N}(0,1)$-Verteilung.

(b) $(X_n)$ stid. mit $\mathrm{Var}\, X_1 = \sigma^2 > 0$. Dann gilt

$$\sup_{x\in\mathbb{R}} \left|F_n(x) - \Phi(x)\right| \le 6\cdot E|X_1 - EX_1|^3 \cdot \frac{1}{\sigma^3\sqrt{n}}.$$

Im Beispiel 3.45 erhält man wegen

$$E|X_1 - EX_1|^3 = \sum_{x=0}^{1} |x-p|^3 p^x (1-p)^{1-x} = p^3(1-p) + (1-p)^3 p = \frac{1}{8}$$

$$\sigma^3 = (sigma^2)^{\frac{3}{2}} = \big(p(1-p)\big)^{\frac{3}{2}} = \left(\frac{1}{4}\right)^{\frac{3}{2}} = \frac{1}{8}$$

Somit folgt

$$\sup_{x\in\mathbb{R}} \left|F_n(x) - \Phi(x)\right| \le 6\cdot\frac{1}{8}\cdot\frac{1}{\frac{1}{8}\sqrt{1000}} \approx 0.1897$$

van Beek (1972) konnte zeigen, daß man die Konstante 6 auf der rechten Seite in (b) durch 0.7995 ersetzen kann. Im Beispiel erhält man dann für die rechte Seite 0.025.

**Bemerkung 3.47:**
Es gibt natürlich auch multivariate zentrale Grenzwertsätze, d.h. zentrale Grenzwertsätze für $k$-dimensionale Zufallsvariablen, zum Beispiel den multivariaten zentralen Grenzwertsatz von Lindeberg/Levy:

$(X_n)_{n\in\mathbb{N}}$ stochastisch unabhängige identisch verteilte Folge von $\mathbb{R}^k$-wertigen Zufallsvariablen mit $EX_1 = \mu \in \mathbb{R}$ und $\mathrm{Cov}\, X_1 = \Sigma \in \mathbb{R}_+^{k\times k}$. Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{\mathrm{w}} \tilde{X}_0 \sim \mathrm{N}_k(0, \Sigma)$$

oder auch formal anders notiert

$$\frac{1}{\sqrt{n}} \Sigma^{-\frac{1}{2}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{\mathrm{w}} \tilde{\tilde{X}}_0 \sim \mathrm{N}_k(0, I_k).$$

Zum Abschluß dieses Kapitels noch ein Resultat (ohne Beweis), welches sich u.a. in der Statistik als äußerst nützlich erweist (siehe zum Beispiel Roussas (1973), pp. 152).

**Satz 3.48:**
$(X_n)_{n\in\mathbb{N}}$, $X$, $(Y_n)_{n\in\mathbb{N}}$ reellwertige Zufallsvariablen und $c \in \mathbb{R}\setminus\{0\}$. Falls $X_n \xrightarrow{\text{w}} X$ und $Y_n \xrightarrow{P} c$, so folgt

(i) $X_n + Y_n \xrightarrow{\text{w}} X + c$,

(ii) $X_n Y_n \xrightarrow{\text{w}} cX$ und

(iii) $\frac{X_n}{Y_n} \xrightarrow{\text{w}} \frac{X}{c}$, falls $P(Y_n \neq 0) = 1$ für alle $n \in \mathbb{N}$.

# 4 Einführung in die Statistik

## 4.1 Problemstellungen

**Beispiel 4.1 (Meinungsumfrage):**
Es werden $n$ zufällig ausgewählte Personen nach der Partei ihrer Wahl befragt. Bei dieser Meinungsumfrage interessiert nur die Frage FDP JA/NEIN. Es antworten $h$ Personen mit JA. Somit wird die Prognose, daß $\frac{h}{n} \cdot 100$ % FDP wählen werden.

(a) Deskription des Experimentes: Codierung 1 = JA, 0 = Nein. $x_i$ entspricht der Antwort der $i$-ten befragten Person $(i = 1, \ldots, n)$, $h$ sei die Häufigkeit, $\frac{h}{n}$ die relative Häufigkeit.

(b) Statistisches Modell: $x_i$ Realisation einer Zufallsvariable $X_i$ mit $P(X_i = 1) =: p$ und $P(X_i = 0) = 1 - p$ sowie $p \in (0, 1)$ sei der „wahre" Anteil der FDP-Wähler in der wahlberechtigten Bevölkerung. Dann gilt $X_i \sim \mathrm{B}(1, p)$. Eine weitere Annahmen sei $X_1, \ldots, X_n$ stochastisch unabhängig.

Probleme dabei sind

(1) Schätzproblem: $p = ?$

(2) Testproblem: $H : p \geq 0.05$ vs. $K : p < 0.05$. Dabei entspricht $H$ der Nullhypothese und $K$ der Alternativhypothese.

Ergebnisse aus der Wahrscheinlichkeitstheorie sind

(i) $X. = \sum_{i=1}^{n} X_i \sim \mathrm{B}(n, p)$

(ii) Es gilt mit $\overline{X}. = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$E\overline{X}. = p \qquad \text{und} \qquad \mathrm{Var}\, \overline{X}. = \frac{1}{n} p(1-p)$$

(iii) $\mathrm{Var}\, \overline{X}. \overset{n \to \infty}{\longrightarrow} 0$

(iv) $\overline{X}. \overset{P}{\to} p$ wegen SGGZ

(v) $\overline{X}. \to p[P]$ wegen StGGZ

(vi) Es ist

$$\begin{aligned} P(\overline{X}. \leq c) &= P\Big(\sum_{i=1}^{n} X_i \leq nc\Big) = \sum_{x=0}^{[|nc|]} \binom{n}{x} p^x (1-p)^{n-x} = \\ &= P\Big(\frac{\sum X_i - np}{\sqrt{np(1-p)}} \leq \frac{nc - np}{\sqrt{np(1-p)}}\Big) \overset{\text{ZGWS}}{\approx} \Phi\Big(\frac{nc - np}{\sqrt{np(1-p)}}\Big) \end{aligned}$$

Man könnte jedoch auch ein asymptotisches Modell betrachten: $(X_i)_{i \in N}$ stochastisch unabhängig und $\mathrm{B}(1, p)$-verteilt. Ein Schätzer $d : \{0, 1\}^n \to [0, 1]$ mit $d(x) = \overline{x}.$ erscheint vernünftig.

Beim Test für $H$ gegen $K$ könnte man so vorgehen:

(i) Intuitiv: Entscheide für $K$, falls

$$P_p(\overline{X}. \leq \overline{x}.) \qquad \forall\, p \in \left[\frac{5}{100}, 1\right)$$

sehr klein, d.h. kleiner als 5 % oder auch kleiner als 1 %.

(ii) Test: $\varphi : \{0,1\}^n \to \{0,1\}$, im Beispiel von der Form

$$\varphi(x) = \begin{cases} 1 & \sum_{i=1}^{n} x_i < c \\ 0 & \sum_{i=1}^{n} x_i \geq c \end{cases}$$

mit geeignetem $c$.

(iii) Interpretation:

$$\varphi(x) = \begin{cases} 1 & \text{Hypothese } H \text{ ablehnen, Entscheidung für } K \\ 0 & \text{Hypothese } H \text{ kann nicht abgelehnt werden} \end{cases}$$

Hierbei sind jetzt verschiedene Fehlentscheidungen möglich:

| | Entscheidung für $K$ | $H$ nicht abgelehnt |
|---|---|---|
| wahres $p \in H$ | Fehler 1. Art | keine Fehlentscheidung |
| wahres $p \in K$ | keine Fehlentscheidung | Fehler 2. Art |

Der Fehler 1. Art ist die Entscheidung für $K$, obwohl $H$ richtig ist, der Fehler 2. Art ist die Nicht-Entscheidung für $K$, obwohl $K$ richtig ist.

**Mathematische Problemstellungen:**

(a) Gibt es in einem gewissen Sinne „optimale“ Schätzer? Zum Beispiel Schätzer für $p$ mit minimaler Varianz?

(b) Gibt es in einem gewissen Sinne „optimale“ Tests mit minimaler Fehlerwahrscheinlichkeiten?

Im Beispiel 4.1 wird sich $d^*(x) := \overline{x}_\cdot$ tatsächlich als gleichmäßig bester erwartungstreuer Schätzer für $p$ erweisen, d.h. für alle $d : \{0,1\}^n \to [0,1]$ mit $E_p d(X) = p \; \forall\, p \in (0,1)$ gilt:

$$\mathrm{Var}_p\; d(X) \geq \mathrm{Var}_p\; d^*(X) \qquad \forall\, p \in (0,1),$$

d.h. $d^*$ ist erwartungstreuer Schätzer mit minimaler Varianz.

**Beispiel 4.2 (Taxiproblem):**
In einer Stadt gibt es $t$ Taxis mit den Taxinummern $1, \ldots, t$. Ich habe 5 Taxis mit den Nummern 73, 179, 280, 405 und 440 gesehen. Nun möchte ich $t$ schätzen. In der Statistik ist die Frage jetzt: Welcher erwartungstreuer Schätzer hat die kleinste Varianz?

Nehme folgendes Modell an: $X_1, \ldots, X_n \sim \mathrm{G}(\{1, \ldots, t\})$ stochastisch unabhängig. $t \in \mathbb{N}$ sei unbekannt. Gesucht ist $d^* : \{1, \ldots, t\}^n \to \mathbb{N}$ mit

$$E_t d^*(X) = t \qquad \text{und} \qquad \mathrm{Var}_t\; d^*(X) \leq \mathrm{Var}_t\; d(X) \quad \forall\, d \text{ mit } E_t d(X) = t.$$

Dann sind mögliche Schätzer:

(a) $d(X) := 2\overline{X}_\cdot - 1$ ist erwartungstreu, denn

$$E_t X_1 = \frac{t+1}{2}, \qquad E_t d(X) = E_t(2\overline{X}_\cdot - 1) = 2\frac{1}{n} n \cdot EX_1 - 1 = 2\frac{t+1}{2} - 1 = t.$$

(b) (Median-Schätzer) $d(X) := 2\hat{M}(X) - 1$ mit

$$\hat{M}(X) := \begin{cases} X_{(k)} & n = 2k-1 \\ \frac{X_{(k)} + X_{(k+1)}}{2} & n = 2k \end{cases}$$

Nachteil bei (a) als auch (b) ist, daß manchmal $d(x) < x_{(n)}$ ist.

(c) (Lückenschätzer) Schätze zunächst die Lücke $L$ zwischen $X_{(n)}$ und $t$ durch das arithmetische Mittel der Stichprobenlücken:

$$\hat{L} := \frac{1}{n}\sum_{j=1}^{n}(X_{(j)} - X_{(j-1)} - 1) \qquad \text{mit } X_{(0)} := 1.$$

Dann ist $d(X) := X_{(n)} + \hat{L} + 1 = \ldots = \left(1 + \frac{1}{n}\right) X_{(n)}$

(d) Der gleichmäßig beste erwartungstreue Schätzer ist jedoch

$$d^*(X) := \frac{X_{(n)}^{n+1} - (X_{(n)} - 1)^{n+1}}{X_{(n)}^n - (X_{(n)} - 1)^n}.$$

Leichter Rechnen läßt es sich im (nicht diskreten) Modell: $X_1, \ldots, X_n \sim G([0,t])$ mit $t > 0$ unbekannt. Dann ist

$$P(X_{(n)} \le x) = P(X_1 \le x)^n = \left(\frac{x}{t}\right)^n \qquad x \in [0,t]$$

Dann ist die $\lambda^1$-Dichte $f_{(n)}(x) = \frac{n}{t^n} x^{n-1} I_{[0,t]}(x)$ und somit

$$E_t X_{(n)} = \int_0^t x \frac{n}{t^n} x^{n-1}\, dx = \frac{n}{n+1}\frac{1}{t^n} x^{n+1}\Big|_0^t = \frac{n}{n+1} t$$

Wir setzen $d(X) := \left(1 + \frac{1}{n}\right) X_{(n)}$ und erhalten $E_t d(X) = t$. Dann ist

$$\begin{aligned}
E_t X_{(n)}^2 &= \int_0^t \frac{n}{t^n} x^{n+1}\, dx = \ldots = \frac{n}{n+2} t^2 \\
\mathrm{Var}_t\, X_{(n)} &= \frac{n}{n+2} t^2 - \left(\frac{n}{n+1}\right)^2 t^2 = \\
&= t^2 n \frac{(n+1)^2 - n(n+2)}{(n+2)(n-1)^2} = t^2 \frac{n}{(n+2)(n-1)^2} = t^2 \frac{1}{n^2 + 4n + 5 + \frac{2}{n}} \\
\mathrm{Var}_t\, d(X) &= \left(\frac{n+1}{n}\right)^2 \frac{1}{n^2 + 4n + 5 + \frac{2}{n}} t^2
\end{aligned}$$

Wir vergleichen mit $\tilde{d}(X) := 2\overline{X}$.

$$\begin{aligned}
E_t \tilde{d}(X) &= t \qquad \forall\, t \\
\mathrm{Var}_t\, X_1 &= \int_0^t x^2 \frac{1}{t}\, dx - \left(\frac{t}{2}\right)^2 = \frac{x^3}{3t}\Big|_0^t - \frac{t^2}{4} = \frac{1}{12} t^2 \\
\mathrm{Var}_t\, \tilde{d}(X) &= 4\frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\, X_i = 4\frac{1}{n}\frac{1}{12} t^2 = \frac{t^2}{3n}
\end{aligned}$$

Wir erkennen $\mathrm{Var}_t\, d(X) \le \mathrm{Var}_t\, \tilde{d}(X)$ mit Gleichheit für $n = 1$, d.h. für große $n$ ist $d(X)$ besser. Es gilt sogar: $d(X)$ ist gleichmäßig bester erwartungstreuer Schätzer.

Beim Testen von Hypothesen ist es i.a. nicht möglich, beide Fehlerarten gleichzeitig zu minimieren. Man versucht deshalb, den Fehler 1. Art unter Kontrolle zu halten, indem man für ein festes $\alpha \in (0,1)$, zum Beispiel $\alpha = 0.05$ oder $\alpha = 0.01$, fordert, daß

$$\forall\, p \in H: \quad P_p(\varphi(X) = 1) \le \alpha$$

bzw.

$$\sup_{p\in H} P_p(\varphi(X)=1)\le\alpha$$

bzw.

$$\sup_{p\in H} E_p\varphi(X)\le\alpha.$$

$\varphi$ heißt Test zum Niveau $\alpha$ (Signifikanzniveau). Sei

$$\Phi_\alpha=\Big\{\varphi:\varphi \text{ Test zum Niveau } \alpha \text{ für } H \text{ vs. } K\Big\}.$$

Man sucht dann unter allen $\varphi\in\Phi_\alpha$ einen Test $\varphi^*$, der die Fehlerwahrscheinlichkeit 2. Art für alle $p\in K$ minimiert, d.h. gesucht ist $\varphi^*\in\Phi_\alpha$ mit

(1) $\varphi^*\in\Phi_\alpha$,

(2) $\forall\,\varphi\in\Phi_\alpha:\quad\forall\,p\in K:\quad E_p\varphi^*\ge E_p\varphi.$

Probleme sind

(i) Häufig ist $\Phi_\alpha=\{\varphi:\sup\limits_{\theta\in H}E_\theta\varphi=\alpha\}=\emptyset$ (bei diskreten Verteilungen). Im Beispiel ist

$$\sup_{p\in H}P_p\Big(\sum_{i=1}^n X_i<c\Big)=\sup_{p\in H}\sum_{\substack{k<c\\ k\in\mathbb{N}_0}}\binom{n}{k}p^k(1-p)^{n-k}\overset{!}{=}\alpha$$

nur für ausgewählte $\alpha\in(0,1)$ möglich.

(ii) Selbst wenn $\Phi_\alpha\neq\emptyset$ (v.a. bei stetigen Verteilungen), muß kein $\varphi^*$ mit (2) existieren.

Es läßt sich (i) leicht durch einen kleinen Kunstgriff lösen: Man erweitert $\Phi_\alpha$ einfach mit Werten, indem man alle $\varphi$ mit Werten in $[0,1]$ zuläßt.

Interpretation: Lehne $H$ mit Wahrscheinlichkeit $\varphi(X)$ ab! Im Beispiel wird sich

$$\varphi^*(x)=\begin{cases}1 & T(x)<c\\ \gamma & T(x)=c\\ 0 & T(x)>c\end{cases}$$

mit eindeutig bestimmten $\gamma\in[0,1)$, $c\in\{0,1,\dots,n\}$ aus

$$E_{p_0=0.05}\varphi(X)=\alpha\quad\Longleftrightarrow\quad\sum_{k=1}^{c-1}\binom{n}{k}p_0^k(1-p_0)^{n-k}+\gamma\binom{n}{c}p_0^c(1-p_0)^{n-c}\overset{!}{=}\alpha$$

als optimal erweisen. $\varphi^*$ heißt gleichmäßig bester Test zum Niveau $\alpha$ für $H$ gegen $K$.

Wird $T(x)=c$ beobachtet, so ist mit Wahrscheinlichkeit $\gamma$ für $K$ zu entscheiden (Dies kann durch ein Zufallsexperiment realisiert werden: $Y\sim\mathrm{G}([0,1])$, lehne $H$ ab, falls $Y\le\gamma$).

**Beispiel 4.3:**
Modell: $X\sim\mathrm{N}(\mu,1)$, $\mu\in\{0,2\}$. Hypothese: $H:\mu=0$ gegen $K:\mu=2$. Naheliegend ist: Lehne $H$ ab, falls $X(\omega)$ „groß“, d.h. ein Test von der Form

$$\varphi(x)=\begin{cases}1 & x\ge c\\ 0 & x<c\end{cases}\qquad(*)$$

Sei dazu $\alpha \in (0,1)$ vorgegeben. Gesucht ist ein $\varphi$ der Form $(*)$ mit $P_0(\varphi(X)=1)=\alpha$, also ein $c \in \mathbb{R}$ mit

$$\begin{aligned} P_0(X \geq c) = \alpha \quad &\Longleftrightarrow \quad 1-\Phi(c)=\alpha \iff \Phi(c)=1-\alpha \iff \\ &\Longleftrightarrow \quad c=\Phi^{-1}(1-\alpha) \overset{\alpha=0.05}{\Longleftrightarrow} c \approx 1{,}64 \end{aligned}$$

Es folgt

$$\begin{aligned} P_2(\varphi(X)=0) &= P_2(X<c) = P_0(X+2<c) = P_0(X<c-2) \approx \\ &\approx \Phi(-0{,}35) = 1-\Phi(0{,}35) \approx 1-0{,}64 = 0{,}36 \end{aligned}$$

d.h. ist $\mu = 2$ „wahl“, so ist die Wahrscheinlichkeit für den Fehler 2. Art ungefähr 0,36 (also die Wahrscheinlichkeit, $H$ irrtümlich anzunehmen). Nun ist

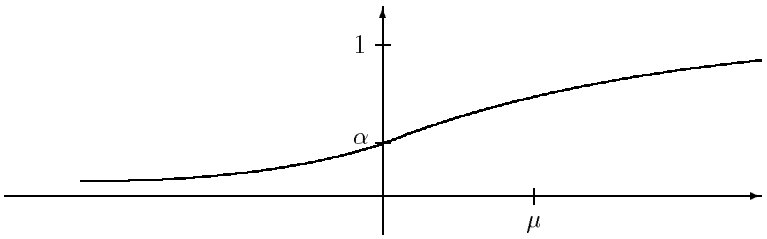
$$P_\mu(X \geq c) = P_0(X+\mu \geq c) = 1-\Phi(c-\mu)$$

wachsend in $\mu$, d.h. obiger Test $\varphi$ ist auch ein Test zum Niveau $\alpha$ für $\tilde{H}: \mu \leq 0$ gegen $\tilde{K}: \mu > 0$ (im erweiterten Modell mit $\mu \in \mathbb{R}$).

Es gilt: $\varphi$ ist gleichmäßig bester Test zum Niveau $\alpha$ $(= 0.05)$ für $\tilde{H}$ gegen $\tilde{K}$ und heißt einseitiger Gauß-Test. Es heißt

$$\beta_\varphi(\mu) := P_\mu(\varphi(X)=1) = 1-\Phi(c-\mu)$$

Gütefunktion von $\varphi$. Es gibt $\beta_\varphi(\mu)$ die Wahrscheinlichkeit bei Vorliegen von $\mu$ an, die Nullhypothese abzulehnen.




**Generelle Voraussetzungen und Bezeichnungen:**
Die Beobachtungen sind Realisierungen von Zufallsvariablen (ein- oder mehrdimensionale), deren Verteilungen nicht (vollständig) bekannt sind. Dies ist häufig als Grundannahme der Statistik bezeichnet.

$$X: (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B}) \text{ mit möglichen Verteilungen } P_\vartheta^X \text{ bzw. } P_\vartheta^Y,\ \vartheta \in \Theta.$$

**Verteilungsannahme:**
Die wahre, aber unbekannte Verteilung liegt in $\mathcal{P} := \{P_\vartheta : \vartheta \in \Theta\}$. Es heißt $\Theta$ Parameterraum. Ist $\Theta \in \mathbb{R}^k$, so spricht man von einer parametrischen Verteilungsannahme, zum Beispiel

$$\mathcal{P} = \left\{P_{(\mu,\sigma^2)} = \mathrm{N}(\mu,\sigma^2) :\ (\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_+\right\}.$$

Andernsfalls spricht man von einer nicht-parametrischen Verteilungsannahme, zum Beispiel

$$\mathcal{P} = \left\{P : P \text{ Wk-Maß über } (\mathbb{R}^1, \mathbb{B}^1) \text{ mit } \lambda^1\text{-Dichte}\right\} = \left\{P_\vartheta : \vartheta \in \Theta\right\} \quad \text{mit } \Theta = \left\{f : f\ \lambda^1\text{-Dichte}\right\}$$

Es heißt $(\mathcal{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$ statistischer Raum oder statistisches Experiment. Unser Ziel ist es jetzt, Aussagen über den unbekannten Parameter $\vartheta \in \Theta$ bzw. die unbekannte Verteilung $P \in \mathcal{P}$ zu machen.

## 4.2 Das Fundamentallemma von Neyman und Pearson

**Definition 4.4:**
Gegeben sei ein statistischer Raum $(\mathcal{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$, $H + K = \Theta$ mit $H, K \neq \emptyset$.

(a) $\varphi : (\mathcal{X}, \mathcal{B}) \to (\{0,1\}, 2^{\{0,1\}})$ heißt (nicht-randomisierter) Test.

(b) $\varphi : (\mathcal{X}, \mathcal{B}) \to ([0,1], [0,1]^{\mathbb{B}^1})$ heißt (randomisierter) Test, Bezeichnung: $\varphi \in \Phi$.

(c) Ein Test $\varphi$ heißt Test zum Niveau $\alpha$ ($\alpha \in [0,1]$) für $H$ gegen $K$ genau dann, wenn für alle $\vartheta \in H$ gilt: $E_\vartheta \varphi \leq \alpha$. Die Bezeichnung ist

$$\Phi_\alpha(H) := \left\{\varphi : \varphi \text{ Test mit } \sup_{\vartheta \in H} E_\vartheta \varphi \leq \alpha\right\}.$$

(d) $\varphi^* \in \Phi_\alpha(H)$ heißt gleichmäßig bester Test (UMP-Test, uniformly most powerful) für $H$ gegen $K$ $:\Longleftrightarrow$

$$\forall\, \varphi \in \Phi_\alpha(H): \quad \forall\, \vartheta \in K: \quad E_\vartheta \varphi^* \geq E_\vartheta \varphi.$$

(e) $\varphi \in \Phi_\alpha(H)$ heißt unverfälscht $:\Longleftrightarrow \forall\, \vartheta \in K:\ E_\vartheta \varphi \geq \alpha$.

(f) $\varphi \in \Phi_\alpha(H)$ schöpft das Niveau $\alpha$ aus $:\Longleftrightarrow \sup_{\vartheta \in H} E_\vartheta \varphi = \alpha$.

(g) Falls $|H| = |K| = 1$, d.h. $|\Theta| = 2$, so nennt man das Testproblem „einfach gegen einfach“. Ein gleichmäßig bester Test $H$ gegen $K$ $(H : K)$ heißt dann schlicht bester Test.

Wir behandeln und lösen zunächst das Testproblem „einfach gegen einfach“. Dazu folgende Vorbemerkung.

**Vorbemerkung:**
Seien $P_0, P_1$ zwei Wahrscheinlichkeitsmaße auf $\mathcal{X}_0 = \{x_1, x_2, \ldots\}$ mit Zähldichten $p_0$ und $p_1$. Sei $H := \{P_0\}$ und $K := \{P_1\}$.

Heuristik: Ist $\frac{p_1(x)}{p_0(x)}$ groß bzw. klein, so spricht dies für $K$ bzw. $H$.

Vorgehensweise: Ordne $r(x) := \frac{p_1(x)}{p_0(x)}$ der Größe nach, etwa

$$r(x_{i_1}) \geq r(x_{i_2}) \geq \ldots$$

Wähle nun $m$ mit $\sum_{j=1}^{m} p_0(x_{i_j}) \leq \alpha$ und $\sum_{j=1}^{m+1} p_0(x_{i_j}) > \alpha$ und bestimme $\gamma$ aus

$$\sum_{j=1}^{m} p_0(x_{i_j}) + \gamma \cdot p_0(x_{i_{m+1}}) \stackrel{!}{=} \alpha.$$

Setze

$$\varphi^*(x) := \begin{cases} 1 & x \in \{x_{i_1}, \ldots, x_{i_m}\} \\ \gamma & x = x_{i_{m+1}} \\ 0 & \text{sonst} \end{cases}$$

Somit ist $E_{P_0}\varphi^* = \alpha$. Ist $\varphi^*$ aber auch bester Test für $H = \{P_0\}$ gegen $K = \{P_1\}$ ?

**Bemerkung 4.5:**
Interpretiert man $p_0$ als Buchpreis und $p_1$ als Buchdicke und will man möglichst viel Buch für möglichst wenig Geld, so wird man zunächst die Bücher, die nichts kosten ($\frac{p_1}{p_0} = \infty$) und dann die Bücher geordnet nach Größe ihres Leistung-Preis-Verhältnisses $\frac{p_1}{p_0}$ kaufen, soweit der Etat reicht.

**Bemerkung 4.6:**
Sind $P_0, P_1$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{B})$, so gilt: $P_0 + P_1 \gg P_i$ für $i = 0, 1$, d.h. es existiert stets ein Maß $\mu$ auf $(\mathcal{X}, \mathcal{B})$, so daß $\mu$-Dichten von $P_0$ und $P_1$ existieren (Radon-Nikodym).

**Definition 4.7:**
Seien $P_i$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{B})$ mit $\mu$-Dichten $p_i$, $i = 0, 1$. Es heißt

$$g : (\mathcal{X}, \mathcal{B}) \to \left([0, \infty], [0, \infty]\mathbb{B}\right)$$

Dichtequotient (DQ) von $P_1$ bzgl. $P_0$ genau dann, wenn $g = \frac{p_1}{p_0}[P_0 + P_1]$, wobei für alle $x \in \{p_0 = 0\} \cap \{p_1 > 0\}$ gilt: $\frac{p_1(x)}{p_0(x)} := +\infty$.

**Bemerkung 4.8:**
Sind $g_1$ und $g_2$ Dichtequotienten von $P_1$ bzgl. $P_0$, so gilt: $g_1 = g_2[P_0 + P_1]$.

**Lemma 4.9:**
Sei $\frac{p_1}{p_0}$ Dichtequotient von $P_1$ bzgl. $P_0$. Dann gilt für $a : \overline{\mathbb{R}} \to [0, 1]$ mit $a(k) = P_0\left(\frac{p_1}{p_0} > k\right)$

(a) $a$ ist monoton nicht wachsend, d.h. aus $k_1 < k_2$ folgt $a(k_2) \leq a(k_1)$.

(b) $a$ ist rechtsseitig stetig, d.h. $a(k^+) = a(k)$.

(c) Es gilt $a(k^-) = P_0\left(\frac{p_1}{p_0} \geq k\right)$ und $a(k^-) - a(k) = P_0\left(\frac{p_1}{p_0} = k\right)$.

(d) Es ist $a(0^-) = 1$ und $a(0) = P_0\left(\frac{p_1}{p_2} > 0\right)$ sowie $a(\infty) = 0 = a(\infty^-)$.

(e) Für alle $\alpha \in [0, 1]$ existiert ein $k \in [0, \infty]$ mit $a(k^-) \geq \alpha \geq a(k)$.

(f) Gilt (e) für $0 \leq k_1 < k_2 \leq \infty$, so ist $\left\{k_1 < \frac{p_1}{p_0} < k_2\right\} = \emptyset[P_0 + P_1]$.

Beweis: (a) - (c) folgen aus

$$1 - a(k) = P_0\left(\frac{p_1}{p_0} \leq k\right) \text{ Verteilungsfunktion von } \frac{p_1}{p_0} \text{ unter } P_0$$

und $\frac{p_1}{p_0} < \infty[P_0]$.

(f) Aus $a(k_1^-) \geq \alpha \geq a(k_1) \geq a(k_2^-) \geq \alpha \geq a(k_2)$ folgt

$$0 = a(k_1) - a(k_2^-) = P_0\left(\frac{p_1}{p_0} > k_1\right) - P_0\left(\frac{p_1}{p_0} \geq k_2\right) = P_0\left(k_1 < \frac{p_1}{p_0} < k_2\right),$$

d.h. $N := \left\{k_1 < \frac{p_1}{p_0} < k_2\right\}$ ist $P_0$-Nullmenge. Jetzt ist also nur noch zu zeigen, daß $N$ auch $P_1$-Nullmenge ist:

$$P_1(N) = \int_N p_1 \, d\mu = \underbrace{\underbrace{\int_{N \cap \{p_0 > 0\}} \frac{p_1}{p_0} \underbrace{p_0 \, d\mu}_{= dP_0}}_{= 0, \text{ da } P_0(N) = 0} + \underbrace{\int_{N \cap \{p_0 = 0\}} p_1 \, d\mu}_{= 0, \text{ da } N \cap \{p_0 = 0\} = \emptyset}}_{} = 0$$

**Satz 4.10 (Fundamentallemma von Neyman und Pearson):**
Seien $P_i$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{B})$ mit $\mu$-Dichten $p_i,\ i = 0, 1$, und Dichtequotient $\frac{p_1}{p_0}$, $\alpha \in [0, 1]$ und $\Phi = \{\varphi : \varphi \text{ Test}\}$. Dann gilt:

(a) (Existenz) $\exists\, k^* \in [0, \infty] \exists\, \varphi^* \in \Phi$:

(FL1) $E_0\varphi^* = \alpha$, dabei gilt bekanntlich $E_0\varphi^* := \int \varphi^*\, dP_0$.

(FL2) Es ist

$$\varphi^*(x) = \begin{cases} 1 & \frac{p_1}{p_0} > k^*[P_0 + P_1] \\ 0 & \frac{p_1}{p_0} < k^*[P_0 + P_1] \end{cases}$$

(b) (Hinreichende Bedingung) Falls $\varphi^* \in \Phi$ mit

(i) (FL1), (FL2) mit $k^* \in [0, \infty)$

oder

(ii) (FL1), (FL2) mit $k^* = \infty$ und $\{\varphi^* = 1\} \supseteq \left\{\frac{p_1}{p_0} = \infty\right\}[P_1]$,

so gilt:

$\varphi^*$ (gleichmäßig) bester Test zum Niveau $\alpha$ für $H = \{P_0\}$ gegen $K = \{P_1\}$.

(c) (Notwendige Bedingung) Falls $\varphi^* \in \Phi_\alpha(H)$ bester Test für $H$ gegen $K$ ist, so folgt (FL2), und, falls

(i) (FL2) mit $k^* \in (0, \infty]$

oder

(ii) (FL2) mit $k^* = 0$ und $P_0(p_1 > 0) = \alpha$

auch (FL1).

Beweis: Die Grenzfälle $\alpha = 0$ und $\alpha \geq P_0(p_1 > 0$ sind gesondert zu behandeln (als Übung). Hinweis hierzu:

$$\alpha = 0 \quad \Rightarrow \quad k^* = \infty \quad \Rightarrow \quad \varphi^* = \begin{cases} 1 & \frac{p_1}{p_0} = \infty \\ 0 & \frac{p_1}{p_0} < \infty \end{cases}$$

Sei $\alpha = \alpha_1 := P_0(p_1 > 0) \Rightarrow k^* = 0 \Rightarrow$

$$\Rightarrow \quad \varphi^* = \begin{cases} 1 & \frac{p_1}{p_0} > 0 \\ 0 & \frac{p_1}{p_0} = 0 \end{cases} \qquad \Rightarrow \quad \varphi^* = \begin{cases} 1 & \frac{p_1}{p_0} > 0 \\ \frac{\alpha - \alpha_1}{1 - \alpha_1} & \frac{p_1}{p_0} = 0 \end{cases}$$

Sei also $0 < \alpha < P_0(p_1 > 0)$.

(a) Nach Lemma 4.9 (e) existiert ein $k^* \in [0, \infty]$ mit $a(k^{*-}) \geq \alpha \geq a(k^*)$. Angenommen, es sei

(i) $k^* = \infty$. Dann folgt nach Lemma 4.9 (d), daß $a(\infty^-) = 0 \geq \alpha$. Dies ist aber ein Widerspruch zur Voraussetzung $\alpha > 0$.

(ii) $k^* = 0$. Dann ist aber nach Lemma 4.9 (d) $a(0) = P_0(p_1 > 0) \leq \alpha$. Widerspruch zur Voraussetzung $\alpha < P_0(p_1 > 0)$.

Somit existiert also ein $k^* \in (0,\infty)$ mit $a(k^{*-}) \geq \alpha \geq a(k^*)$. Definiere damit

$$\varphi^*_{k^*} := \begin{cases} 1 & \frac{p_1}{p_0} > k^* \\ \gamma^* & \frac{p_1}{p_0} = k^* \\ 0 & \frac{p_1}{p_0} < k^* \end{cases} \qquad \gamma^* := \begin{cases} \frac{\alpha - a(k^*)}{a(k^{*-}) - a(k^*)} & a(k^{*-}) > a(k^*) \\ 1 & a(k^{*-}) = a(k^*) \end{cases}$$

Dann gilt (FL2) trivialerweise und (FL1) wegen

$$E_0\varphi^*_{k^*} = P_0\left(\frac{p_1}{p_0} > k^*\right) + \gamma^* \cdot P_0\left(\frac{p_1}{p_0} = k^*\right) \overset{4.9\,(c)}{=} a(k^*) + \gamma^*\Big(a(k^{*-}) - a(k^*)\Big) = \alpha.$$

(b) $\varphi^*$ erfülle (FL1) und (FL2) mit $k^* \in (0,\infty)$ und es sei $\varphi \in \Phi_\alpha(H)$. Zu zeigen ist: $E_1\varphi^* \geq E_1\varphi$. Wegen $E_0\varphi^* = \alpha \geq E_0\varphi$ ist zunächst folgendes richtig:

$$k^*\big(E_0\varphi^* - E_0\varphi\big) = k^*(\alpha - E_0\varphi) \geq 0,$$

d.h. die Behauptung folgt, wenn

$$E_1\varphi^* - E_1\varphi \overset{!}{=} k^*(\alpha - E_0\varphi)$$

Dies ist gezeigt, wenn gilt

$$\begin{aligned} 0 \overset{!}{\leq} E_1(\varphi^* - \varphi) - k^*E_0(\varphi^* - \varphi) &= \int (\varphi^* - \varphi)(p_1 - k^*p_0)\, d\mu = \\ &= \int\limits_{\{\varphi^* < \varphi\}} (\varphi^* - \varphi)(p_1 - k^*p_0)\, d\mu + \int\limits_{\{\varphi^* > \varphi\}} (\varphi^* - \varphi)(p_1 - k^*p_0)\, d\mu \end{aligned}$$

(i) Sei $\varphi^* < \varphi \leq 1$. Nach (FL2) gilt $\frac{p_1}{p_0} \leq k^*[P_0 + P_1]$, wegen $k^* < \infty$ gilt $p_1 \leq k^*p_0[P_0 + P_1]$ und damit

$$\underbrace{(\varphi^* - \varphi)}_{<0}\underbrace{(p_1 - k^*p_0)}_{\leq 0} \geq 0[P_0 + P_1]$$

(ii) Sei $\varphi^* > \varphi \geq 0$. Nach (FL2) gilt $\frac{p_1}{p_0} \geq k^*[P_0 + P_1]$ und dann $p_1 \geq k^*p_0[P_0 + P_1]$ sowie

$$\underbrace{(\varphi^* - \varphi)}_{>0}\underbrace{(p_1 - k^*p_0)}_{\geq 0} \geq 0[P_0 + P_1]$$

Es folgt die Behauptung.

(c) Sei $\varphi$ (gleichmäßig) bester Test zum Niveau $\alpha$ für $P_0 : P_1$. $\varphi^*$ erfülle (FL2) und (FL2) mit $k^* \in (0,\infty)$ ($\exists$ nach (a), optimal nach (b)). Also zu zeigen: (FL1) und (FL2) gelten für $\varphi$. Es ist zunächst $E_1\varphi \geq E_1\varphi^*$, da $\varphi$ als bester Test vorausgesetzt wurde, und $E_0\varphi \leq \alpha = E_0\varphi^*$. Dann folgt

$$0 \leq E_1\varphi - E_1\varphi^* - k^*\underbrace{(E_0\varphi - E_0\varphi^*)}_{\leq 0} \overset{(b)}{\leq} 0.$$

Somit gilt $\{\varphi \neq \varphi^*\} \cap \left\{\frac{p_1}{p_0} \neq k^*\right\} = \emptyset[P_0 + P_1]$, d.h. es gilt (FL2) für $\varphi$. Aus (b) folgt $E_1\varphi^* \geq E_1\varphi$ und dann, da $\varphi$ bester Test ist, $E_1\varphi^* = E_1\varphi$. Insgesamt also $E_0\varphi = E_0\varphi^* = \alpha$, d.h. (FL1) für $\varphi$.

**Bemerkung 4.11:**
Merke die konstruktiven Definitionen aus dem Beweis von Satz 4.10 der Tests $\varphi^*_k$ mit $k \in [0,\infty]$.

1. $k^* = \infty$ $(\alpha = 0)$:

$$\varphi^*_\infty = \left\{ \begin{array}{ll} 1 & \frac{p_1}{p_0} = \infty \\ 0 & \frac{p_1}{p_0} < \infty \end{array} \right\} = \left\{ \begin{array}{ll} 1 & p_1 > 0,\ p_0 = 0 \\ 0 & p_0 > 0 \end{array} \right.$$

   Somit ist $E_0\varphi^*_\infty = 0$ und $E_1\varphi^*_\infty = P_1(p_0 = 0)$.

2. $0 < k^* < \infty$ $(0 < \alpha < P_0(p_1 > 0))$:

$$\varphi^*_{k^*} = \left\{ \begin{array}{ll} 1 & \frac{p_1}{p_0} > k^* \\ \gamma^* & \frac{p_1}{p_0} = k^* \\ 0 & \frac{p_1}{p_0} < k^* \end{array} \right. \qquad \gamma^* = \left\{ \begin{array}{ll} 1 & P_0\left(\frac{p_1}{p_0} = k^*\right) = 0 \\ \frac{\alpha - P_0\left(\frac{p_1}{p_0} > k^*\right)}{P_0\left(\frac{p_1}{p_0} = k^*\right)} & \text{sonst} \end{array} \right.$$

   $\Rightarrow \gamma^* \in [0,1]$, $E_0\varphi^*_{k^*} = \alpha$ und $E_1\varphi^*_{k^*} < 1$, denn angenommen $E_1\varphi^*_{k^*} = 1 \Rightarrow$

$$\{\varphi^*_{k^*} = 1\} \supseteq \{p_1 > 0\}[P_1]$$

   $\Rightarrow P_0(p_1 > 0) \leq E_0\varphi^*_{k^*} = \alpha$. Widerspruch zur Wahl $\alpha < P_0(p_1 > 0)$.

3. $k^* = 0$ $(\alpha \geq P_0(p_1 > 0))$:

$$\varphi^*_0 = \left\{ \begin{array}{ll} 1 & \frac{p_1}{p_0} > 0 \\ 0 & \frac{p_1}{p_0} = 0 \end{array} \right\} = \left\{ \begin{array}{ll} 1 & p_1 > 0 \\ 0 & p_1 = 0,\ p_0 > 0 \end{array} \right.$$

   Somit ist $E_0\varphi^*_0 = P_0(p_1 > 0)$ und $E_1\varphi^*_0 = 1$.

**Bemerkung 4.12:**
Falls $\alpha_1 = P_0(p_1 > 0)$, so gilt für alle $\alpha \in [\alpha_1, 1]$: $\varphi^*_0$ ist (gleichmäßig) bester Test zum Niveau $\alpha$ für $P_0 : P_1$ wegen $E_1\varphi^*_0 = 1$ und $E_0\varphi^*_0 = \alpha_1 \leq \alpha$. Für alle $\alpha > \alpha_1$ gilt:

$$E_0\tilde{\varphi}_{0,\alpha} = \underbrace{P_0\left(\frac{p_1}{p_0} > 0\right)}_{= P_0(p_1 > 0)} + \underbrace{\frac{\alpha - \alpha_1}{1 - \alpha_1}}_{> 0} \cdot \underbrace{P_0\left(\frac{p_1}{p_0} = 0\right)}_{= P_0(p_1 = 0, p_0 > 0)} > E_0\varphi^*_0,$$

falls $P_0(p_1 = 0, p_0 > 0) > 0$.

**Bemerkung 4.13:**
Für alle $\varphi \in \Phi$, $\forall\ k^* \in [0, \infty]$ gilt

$$\left. \begin{array}{l} E_0\varphi \leq E_0\varphi^*_{k^*} \\ E_1\varphi \geq E_1\varphi^*_k \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} E_0\varphi = E_0\varphi^*_{k^*} \\ E_1\varphi = E_1\varphi^*_{k^*} \end{array} \right.$$

**Corollar 4.14:**
Unter den Voraussetzungen von Satz 4.10 sei $\varphi^*$ bester Test zum Niveau $\alpha$ für $P_0 : P_1$ mit

$$E_1\varphi^* < 1 \quad \text{oder} \quad E_1\varphi^* = 1 \text{ und } E_0\varphi^* = P_0(p_1 > 0).$$

Dann ist $1 - \varphi^*$ bester Test zum Niveau $1 - \alpha$ für $P_1 : P_0$.

Beweis: Aus Satz 4.10 (c) folgt (FL2), d.h.

$$\varphi^* = \left\{ \begin{array}{ll} 1 & \frac{p_1}{p_0} > k^*[P_0 + P_1] \\ 0 & \frac{p_1}{p_0} < k^*[P_0 + P_1] \end{array} \right.$$

($\alpha$) Sei $k^* \in (0,\infty]$. Dann ist

$$1-\varphi^* = \left\{ \begin{array}{ll} 0 & \frac{p_1}{p_0} > k^*[P_0+P_1] \\ 1 & \frac{p_1}{p_0} < k^*[P_0+P_1] \end{array} \right\} = \left\{ \begin{array}{ll} 1 & \frac{p_0}{p_1} > \frac{1}{k^*}[P_0+P_1] \\ 0 & \frac{p_0}{p_1} < \frac{1}{k^*}[P_0+P_1] \end{array} \right.$$

Dabei ist $\frac{p_1}{p_0} \left\{ \begin{array}{ll} = +\infty & p_0 = 0,\ p_1 > 0 \\ \in (0,\infty) & p_0 > 0,\ p_1 > 0 \\ = 0 & p_0 > 0,\ p_1 = 0 \end{array} \right.$ mit $\frac{1}{k^*} < \infty$ $\left(\frac{1}{\infty} := 0\right)$. Mit Satz 4.10 (b) (setze dort $k' := \frac{1}{k^*} \in [0,\infty)$). Es folgt die Behauptung.

($\beta$) Sei $k^* = 0$. Dann ist nach Voraussetzung

$$\begin{aligned} E_1\varphi^* &\geq P_1\left(\frac{p_1}{p_0} > k^* = 0\right) = P_1(p_1 > 0) = 1 \\ E_0\varphi^* &= P_0(p_1 > 0) \end{aligned}$$

Weiter ist

$$E_0\varphi^* = \int \varphi^* p_0 \, d\mu = \int\limits_{\{p_0>0,p_1=0\}} \varphi^* p_0 \, d\mu + \int\limits_{\{p_0>0,p_1>0\}} \varphi^* p_0 \, d\mu$$

Wegen $\{p_0 > 0, p_1 > 0\} = \left\{\frac{p_1}{p_0} > k^* = 0\right\}[P_0+P_1] \subseteq \{\varphi^* = 1\}[P_0+P_1]$ ist jedoch

$$\int\limits_{\{p_0>0,p_1>0\}} \varphi^* p_0 \, d\mu = \int\limits_{\{p_0>0,p_1>0\}} p_0 \, d\mu = P_0(p_1 > 0)$$

und damit $\int\limits_{\{p_0>0,p_1=0\}} \varphi^* p_0 \, d\mu = 0$. Insgesamt also $\{\varphi^* > 0\} \cap \{p_0 > 0, p_1 = 0\} = \emptyset[P_0]$ und damit

$$\{\varphi^* > 0\} \cap \{p_0 > 0, p_1 = 0\} = \emptyset[P_0+P_1].$$

Also ist

$$\begin{aligned} \varphi^* &= \left\{ \begin{array}{ll} 1 & \frac{p_1}{p_0} > 0[P_1+P_0] \\ 0 & \frac{p_1}{p_0} = 0[P_1+P_0] \end{array} \right. \\ 1-\varphi^* &= \left\{ \begin{array}{ll} 1 & \frac{p_0}{p_1} = \infty[P_0+P_1] \\ 0 & \frac{p_0}{p_1} < \infty[P_0+P_1] \end{array} \right. \end{aligned}$$

Mit Satz 4.10 (b) folgt die Behauptung.

Die bisherige Wahl einer konstanten Randomisierung muß nicht immer „ideal“ sein, d.h. $\gamma^*$ kann auch ersetzt werden durch $\gamma^*(x)$, siehe dazu folgendes Beispiel.

**Beispiel 4.15 (Rechteckverteilungen mit Skalenparameter):**
Seien $P_0 = \mathrm{G}([0,1])$, $P_1 = \mathrm{G}([0,2])$ und $0 < \alpha < \frac{1}{2}$. Dann ist $p_0(x) = I_{[0,1]}(x)$ und $p_1(x) = \frac{1}{2}I_{[0,2]}(x)$. Definiere dann

$$q(x) := \frac{p_1(x)}{p_0(x)} = \frac{1}{2}I_{[0,1]}(x) + \infty I_{(1,2]}(x).$$

Mit

$$\varphi_1^*(x) = \left\{ \begin{array}{ll} 1 & q(x) > 0,5 \\ \alpha & q(x) = 0,5 \\ 0 & q(x) < 0,5 \end{array} \right\} = \left\{ \begin{array}{ll} 1 & x > 1[P_0+P_1] \\ \alpha & x \leq 1[P_0+P_1] \end{array} \right.$$

ist dann $E_0\varphi_1^* = \alpha$ und $E_1\varphi_1^* = \frac{1+\alpha}{2}$. Nach dem Fundamentallemma ist $\varphi_1^*$ bester Test zum Niveau $\alpha$ für $P_0 : P_1$, andererseits auch

$$\varphi_2^*(x) = \begin{cases} 1 & x \geq 1-\alpha \\ 0 & x < 1-\alpha \end{cases}$$

Der Test $\varphi_2^*$ hat den Vorteil, daß er nicht randomisiert ist. Der N-P-Test ist also keineswegs eindeutig!

## 4.3 Monotone Dichtequotienten

**Beispiel 4.16 (Einseitiger Gauß-Test, vgl. Beispiel 4.3):**
Modell: $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma_0^2)$ stochastisch unabhängig, $\mu \in \mathbb{R}^1$ ($\sigma_0^2 \in \mathbb{R}_+^1$). Die Schreibweise bedeutet, daß $\mu$ getestet werden soll, während $\sigma_0^2$ (in Klammern geschrieben) bekannt ist. Die Hypothesen seien $H : \mu \leq \mu_0$ vs. $K : \mu > \mu_0$. Da wir bis dato nur „einfach gegen einfach"-Probleme lösen können, führen wir folgende Einschränkungen durch:

1. Einschränkung: $\tilde{H} : \mu = \mu_0$ vs. $K : \mu > \mu_0$,
2. Einschränkung: $\tilde{H} : \mu = \mu_0$ vs. $\tilde{K} : \mu = \mu_1$ mit $\mu_1 > \mu_0$.

Wir wenden das Fundamentallemma (4.10) auf die 2. Einschränkung an: Die $\lambda^n$-Dichte von $P^{(X_1,\ldots,X_n)}$ ist

$$p_\mu(x) = \left(\frac{1}{\sqrt{2\pi}\sigma_0^2}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu)^2\right\}.$$

Der Dichequotient (DQ) ist

$$\begin{aligned}
\frac{p_{\mu_1}(x)}{p_{\mu_0}(x)} &= \frac{\exp\left\{-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^n x_i^2 - 2\mu_1\sum_{i=1}^n x_i + n\mu_1^2\right)\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^n x_i^2 - 2\mu_0\sum_{i=1}^n x_i + n\mu_0^2\right)\right\}} = \\
&= \exp\left\{-\frac{1}{2\sigma_0^2}\left(-2\Big(\sum_{i=1}^n x_i\Big)(\mu_1-\mu_0) + n(\mu_1^2-\mu_0^2)\right)\right\} = \\
&= \exp\left\{-\frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2) + \frac{\mu_1-\mu_0}{\sigma_0^2}\sum_{i=1}^n x_i\right\}
\end{aligned}$$

Nach dem Fundamentallemma ist dann

$$\varphi^* = \begin{cases} 1 & \frac{p_{\mu_1}(x)}{p_{\mu_0}(x)} \geq k \\ 0 & \frac{p_{\mu_1}(x)}{p_{\mu_0}(x)} < k \end{cases}$$

mit $E_{\mu_0}\varphi^* = \alpha$. Dabei kann $\gamma = 1$ gewählt werden, da $\frac{p_{\mu_1}(x)}{p_{\mu_0}(x)}$ stetig verteilt ist. Nun gilt mit $x_\cdot = \sum x_i$

$$\begin{aligned}
\frac{p_{\mu_1}(x)}{p_{\mu_0}(x)} \geq k \quad &\Longleftrightarrow \quad \underbrace{\exp\left\{-\frac{n(\mu_1^2-\mu_0^2)}{2\sigma_0^2}\right\}}_{>0} \exp\left\{\frac{\mu_1-\mu_0}{\sigma_0^2}x_\cdot\right\} \geq k \\
&\overset{\text{e mon. wachs.}}{\Longleftrightarrow} \quad \exp\left\{\frac{\mu_1-\mu_0}{\sigma_0^2}x_\cdot\right\} \geq k_1 \\
&\Longleftrightarrow \quad \frac{\mu_1-\mu_0}{\sigma_0^2}x_\cdot \geq k_2 \\
&\Longleftrightarrow \quad x_\cdot \geq k_3,
\end{aligned}$$

d.h. wir können den Test schreiben als

$$\varphi^*(x) = \begin{cases} 1 & x_\cdot \geq k_3 \\ 0 & x_\cdot < k_3 \end{cases} \qquad E_{\mu_0}\varphi^* = P_{\mu_0}(X_\cdot \geq k_3) \stackrel{!}{=} \alpha.$$

Unter $\mu = \mu_0$ gilt $X_\cdot \sim \mathrm{N}(n\mu_0, n\sigma_0^2)$ und somit

$$\frac{X_\cdot - n\mu_0}{\sqrt{n\sigma_0^2}} = \sqrt{n}\frac{\overline{X}_\cdot - \mu_0}{\sigma_0} \sim \mathrm{N}(0,1).$$

Insgesamt also

$$\alpha = P_{\mu_0}(X_\cdot \geq k_3) = P_{\mu_0}\Big(\underbrace{\sqrt{n}\frac{\overline{X}_\cdot - \mu_0}{\sigma_0}}_{\sim \mathrm{N}(0,1)} \geq k_4\Big)$$

mit $k_4 = u_\alpha$ ($u_\alpha$ definiert durch $\Phi(u_\alpha) = 1-\alpha$, $u_\alpha$ heißt $\alpha$-Fraktil der $\mathrm{N}(0,1)$-Verteilung). Der Test hat somit folgende Gestalt:

$$\varphi^*(x) = \begin{cases} 1 & \sqrt{n}\frac{\overline{x}_\cdot - \mu_0}{\sigma_0} \geq u_\alpha \\ 0 & \sqrt{n}\frac{\overline{x}_\cdot - \mu_0}{\sigma_0} < u_\alpha \end{cases}$$

$\varphi^*$ ist bester Test zum Niveau $\alpha$ für $\tilde{H} : \mu = \mu_0$ gegen $\tilde{K} : \mu = \mu_1$ für alle $\mu_1 < \mu_0$ (da unabhängig von $\mu_1$). Damit ist $\varphi^*$ gleichmäßig bester Test zum Niveau $\alpha$ für $\tilde{H} : \mu = \mu_0$ gegen $K : \mu > \mu_0$. Weiter ist

$$\begin{aligned} \beta_{\varphi^*}(\mu) &:= E_\mu\varphi^* = P_\mu\Big(\sqrt{n}\frac{\overline{X}_\cdot - \mu_0}{\sigma_0} > u_\alpha\Big) = \\ &= P_0\Big(\frac{\sqrt{n}}{\sigma_0}(\overline{X}_\cdot + \mu - \mu_0) > u_\alpha\Big) = \\ &= P_0\Big(\underbrace{\frac{\sqrt{n}\overline{X}_\cdot}{\sigma_0}}_{\sim \mathrm{N}(0,1)} > u_\alpha - \sqrt{n}\frac{\mu - \mu_0}{\sigma_0}\Big) = \\ &= 1 - \Phi\Big(u_\alpha - \sqrt{n}\frac{\mu - \mu_0}{\sigma_0}\Big) \begin{cases} < \alpha & \mu < \mu_0 \\ = \alpha & \mu = \mu_0 \\ > \alpha & \mu > \mu_0 \end{cases} \end{aligned}$$

Damit ist $\varphi^*$ Test zum Niveau $\alpha$ für $H : \mu \leq \mu_0$ gegen $K : \mu \geq \mu_0$. Sei $\varphi$ ein beliebiger Test zum Niveau $\alpha$ für $H : K$, dann ist auch $\varphi$ Test zum Niveau $\alpha$ für $\tilde{H} : \mu = \mu_0$ vs. $K : \mu > \mu_0$. Dann gilt für alle $\mu > \mu_0$: $E_\mu\varphi \leq E_\mu\varphi^*$, d.h. $\varphi^*$ ist gleichmäßig bester Test zum Niveau $\alpha$ für $H : K$.

Die letzten Überlegungen im Beispiel 4.16 gelten ganz allgemein:

($*$) Für alle $\vartheta \in K$: $\varphi^*$ bester Test zum Niveau $\alpha$ für $\tilde{H} : \{\vartheta\} \iff$

$$\iff \left.\begin{array}{l} \varphi^* \text{ glm. bester Test zum Niveau } \alpha \text{ für } \tilde{H} : K \\ H \supseteq \tilde{H},\ \varphi^* \text{ Test zum Niveau } \alpha \text{ für } H : K \end{array}\right\} \Rightarrow$$

$\Rightarrow \varphi^*$ ist gleichmäßig bester Test zum Niveau $\alpha$ für $H : K$.

Entscheidend für ($*$) war, daß der Dichtequotient $\frac{p_{\mu_1}}{p_{\mu_0}}$ eine spezielle Gestalt hatte, so daß der Neyman-Pearson-Text unabhängig von $\mu_1$ war. Es galt nämlich für alle $\mu \geq \mu_0$: $\forall\, x \in \mathbb{R}^n$

$$\frac{p_{\mu_1}(x)}{p_{\mu_0}(x)} = \underbrace{\exp\Big\{-\frac{n(\mu-\mu_0)^2}{2\sigma_0^2}\Big\}}_{=:C_1(\mu,\mu_0)} \underbrace{\exp\Big\{n\frac{\mu-\mu_0}{\sigma_0}\overline{x}_\cdot\Big\}}_{=:C_2(\mu,\mu_0)} = C_1(\mu,\mu_0)\cdot\exp\{C_2(\mu,\mu_0)\overline{x}_\cdot\} =: H_{\mu\mu_0}\big(T(x)\big)$$

mit $T(x) := \overline{x}$. unabhängig von $\mu$ und

$$H_{\mu\mu_0}(t) := C_1(\mu,\mu_0)\cdot\exp\{C_2(\mu,\mu_0)t\},$$

wobei $H_{\mu\mu_0}(t)$ monoton wachsend in $t$ ist, falls $\mu > \mu_0$. Dies nutzen wir zur Verallgemeinerung.

**Voraussetzungend und Bezeichnungen:**
$(\mathcal{X},\mathcal{B},\mathcal{P}=\{P_\vartheta:\vartheta\in\Theta\})$ statistischer Raum mit $\Theta\subseteq\mathbb{R}^1$, $H:K$ Testproblem ($\emptyset\neq H\subset\Theta$, $K=\Theta\setminus H$) und $\tau\in\mathbb{R}$.

**Definition 4.17:**
$\mathcal{P}$ hat monotonen Dichtequotienten in $\tau$ bzgl. $H:K :\Longleftrightarrow$

(i) $\exists\, T_\tau : (\mathcal{X},\mathcal{B})\to(\mathbb{R}^1,\mathbb{B}^1)$ und

(ii) $\forall\ \vartheta\in K\ (\vartheta\in H) : \exists\ H_{\vartheta\tau} : \mathbb{R}^1\to[0,\infty]$ monoton nicht fallend (wachsend), so daß für die Dichtequotienten gilt:

$$\frac{p_\vartheta(x)}{p_\tau(x)} = H_{\vartheta\tau}\big(T_\tau(x)\big)[P_\vartheta+P_\tau].$$

**Bemerkung:**
Aus $H_{\vartheta\tau}$ monoton folgt $H_{\vartheta\tau}$ meßbar.

**Satz 4.18 (Erweitertes Fundamentallemma):**
Gegeben: $(\mathcal{X},\mathcal{B},\mathcal{P}=\{P_\vartheta:\vartheta\in\Theta\})$, Testproblem $H:K$, $\alpha\in(0,1)$, $\tau\in H$, so daß $\mathcal{P}$ monotonen Dichtequotienten in $\tau$ bzgl. $H:K$ besitzt. Dann ist

$$\varphi^* := \begin{cases} 1 & T_\tau > k \\ \gamma & T_\tau = k \\ 0 & T_\tau < k \end{cases} \qquad \text{mit } E_\tau\varphi^* = \alpha$$

gleichmäßig bester Test zum Niveau $\alpha$ für $H:K$.

Beweis:

(i) Seien $\vartheta\in K$ und

$$H_{\vartheta\tau}\big(T_\tau(x)\big) = \frac{p_\vartheta(x)}{p_\tau(x)}\left\{\begin{array}{c}>\\<\end{array}\right\}\tilde{k} := H_{\vartheta\tau}(k)[P_\vartheta+P_\tau].$$

Wegen $H_{\vartheta\tau}\uparrow$ gilt

$$T_\tau(x)\left\{\begin{array}{c}>\\<\end{array}\right\}k[P_\vartheta+P_\tau] \quad \text{und damit} \quad \varphi^* = \begin{cases} 1 & \frac{p_\vartheta}{p_\tau} > \tilde{k}[P_\vartheta+P_\tau] \\ 0 & \frac{p_\vartheta}{p_\tau} < \tilde{k}[P_\vartheta+P_\tau] \end{cases},$$

also (FL2) und $E_\tau\varphi^* = \alpha$ nach Voraussetzung, also (FL1). Mit FL (b), falls $\tilde{k}<\infty$, ist $\varphi^*$ bester Test zum Niveau $\alpha$ für $P_\vartheta : P_\tau$.

Angenommen $\tilde{k}=\infty$, dann gilt

$$\begin{aligned}\alpha &= E_\tau\varphi^* = P_\tau(T_\tau>k)+\gamma\cdot P_\tau(T_\tau=k)\le P_\tau(T_\tau>k)\le\\ &\le P_\tau\big(H_{\vartheta\tau}(T_\tau)\ge H_{\vartheta\tau}(k)=\tilde{k}\big) = P_\tau\left(\frac{p_\vartheta}{p_\tau}\ge\infty\right) = P_\tau(p_\vartheta>0,p_\tau=0) = 0\end{aligned}$$

Dies ist aber ein Widerspruch zur Voraussetzung $\alpha\in(0,1)$.

(ii) Teil (i) gilt für alle $\vartheta \in K$. Damit ist $\varphi^*$ gleichmäßig bester Test zum Niveau $\alpha$ für $\{\tau\} : K$. Bleibt noch zu zeigen, daß $\varphi \in \Phi_\alpha(H)$ (vgl. Bemerkung nach Beispiel 4.16), d.h. für alle $\vartheta \in H : E_\vartheta \varphi^* \leq \alpha$. Falls nun $H = \{\tau\}$, so gezeigt nach (i). Sei also $H \supset \{\tau\}$. Verwende $1 - \varphi^*$ als Test zum Niveau $1 - \alpha$ für $P_\tau : P_\vartheta$ für $\vartheta \in H \setminus \{\tau\}$ und zeige wie in (i):

$$1 - \varphi^* \text{ ist bester Test zum Niveau } 1 - \alpha \text{ für } P_\tau : P_\vartheta$$

Damit ist $1 - \varphi^*$ gleichmäßig bester Test zum Niveau $1 - \alpha$ für $\{\tau\} : H \setminus \{\tau\}$. Vergleiche dann $1 - \varphi^*$ mit dem konstanten Test $\psi :\equiv 1 - \alpha$. Dann ist $E_\tau \psi = 1 - \alpha$, also $\psi$ Test zum Niveau $1 - \alpha$ für $\tau : H \setminus \{\tau\}$. Dann gilt für alle $\vartheta \in H \setminus \{\tau\}$

$$E_\vartheta (1 - \varphi^*) \geq E_\vartheta \psi = 1 - \alpha$$

und damit für alle $\vartheta \in H \setminus \{\vartheta\} : \quad E_\vartheta \varphi^* \leq \alpha$. Es folgt die Behauptung.

**Bemerkung 4.19:**
$\varphi^*$ aus Satz 4.18 minimiert auch noch die Fehlerwahrscheinlichkeit 1. Art unter $E_\tau \varphi = \alpha$. Denn eine Beweisanalyse zeigt: $\varphi^*$ ist Lösung von

(i) $\forall\, \vartheta \in K : \quad E_\vartheta \varphi = \sup$ unter $E_\tau \varphi = \alpha$

und auch von

(ii) $\forall\, \vartheta \in H \setminus \{\tau\} : \quad E_\vartheta \varphi = \inf$ unter $E_\tau \varphi \geq \alpha$.

Damit ist $\varphi^*$ sowohl Lösung von

(iii) $\forall\, \vartheta \in K : \quad E_\vartheta \varphi = \sup$ und $\forall\, \vartheta \in H : \quad E_\vartheta \varphi \leq \alpha$

als auch von

(iv) $\forall\, \vartheta \in K : \quad E_\vartheta \varphi = \sup$ und $\forall\, \vartheta \in H \setminus \{\tau\} : \quad E_\vartheta \varphi = \inf,\ E_\tau \varphi = \alpha$.

**Beispiel 4.20 (1-seitiger Binomialtest):**
Modell: $X_1, \ldots, X_n \sim \mathrm{B}(1, p)$ stochastisch unabhängig mit $p \in (0, 1)$. Hypothese: $H : p \leq p_0$ vs. $K : p > p_0$ mit $p_0 \in (0, 1)$. Dann ist die Zähldichte

$$f_p(x) = p^{\sum x_i} (1 - p)^{n - \sum x_i}, \qquad x \in \{0, 1\}^n$$

und der Dichtequotient

$$\frac{f_p(x)}{f_{p_0}(x)} = \frac{p^{x_\cdot} (1 - p)^{n - x_\cdot}}{p_0^{x_\cdot} (1 - p_0)^{n - x_\cdot}} = \left( \frac{1 - p}{1 - p_0} \right)^n = \frac{p^{x_\cdot} (1 - p_0)^{n - x_\cdot}}{p_0^{x_\cdot} (1 - p)^{n - x_\cdot}} =: H_{pp_0}\big(T(x)\big)$$

mit $T(x) := x_\cdot$ und

$$H_{pp_0}(t) := \left( \frac{1 - p}{1 - p_0} \right)^n \cdot \exp\left\{ t \cdot \ln \left( \frac{p(1 - p_0)}{p_0 (1 - p)} \right) \right\}$$

ist monoton wachsend (monoton fallend) genau dann, wenn

$$\ln \left( \frac{p(1 - p_0)}{p_0 (1 - p)} \right) > 0 \qquad (\text{„} < \text{“}),$$

d.h. genau dann, wenn $p > p_0$ $(p < p_0)$. Somit ist

$$\varphi^*(x) = \begin{cases} 1 & x_\cdot > k \\ \gamma & x_\cdot = k \\ 0 & x_\cdot < k \end{cases} \qquad \text{mit } E_{p_0} \varphi^* = \alpha$$

ist gleichmäßig bester Test zum Niveau $\alpha \in (0,1)$ für $H : K$. Bestimme $k$ aus $E_{p_0}\varphi^* \stackrel{!}{=} \alpha$, also

$$\sum_{t=k+1}^{n} \binom{n}{t} p_0^t (1-p_0)^{n-t} + \gamma \binom{n}{k} p_0^k (1-p_0)^{n-k} \stackrel{!}{=} \alpha$$

bzw.

$$\sum_{t=0}^{k-1} \binom{n}{t} p_0^t (1-p_0)^{n-t} + (1-\gamma) \binom{n}{k} p_0^k (1-p_0)^{n-k} \stackrel{!}{=} 1-\alpha$$

bzw.

$$F_{p_0}(k-1) + (1-\gamma) f_{p_0}(k) \stackrel{!}{=} 1-\alpha$$

mit $F_{p_0}$ Verteilungsfunktion und $f_{p_0}$ Zähldichte der $\mathrm{B}(n, p_0)$-Verteilung.

Verteilungsfamilien wie $\mathcal{P} = \{\mathrm{B}(n,p) : p \in (0,1)\}$, $\mathcal{P} = \{\mathrm{N}(\mu, \sigma_0^2) : \mu \in \mathbb{R}\}$ oder auch $\mathcal{P} = \{\mathrm{Po}(\lambda) : \lambda > 0\}$ haben eine „ähnliche“ Struktur, so daß die Bestimmung der Tests ohne „lange Herleitung“ möglich wird. Die „ähnliche“ Struktur führt zum Begriff der 1-parametrigen Exponentialfamilie.

**Definition 4.21:**
$(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\})$ statistischer Raum mit eindeutiger Parametrisierung und $\Theta \subseteq \mathbb{R}^k$. $\mathcal{P}$ heißt 1-parametrige Exponentialfamilie in $\xi(\vartheta)$ und $T(x) :\iff \exists\ \mu\ \sigma$-endlich mit $\mu \gg \mathcal{P}$ (d.h. $\forall\ \vartheta \in \Theta : \mu \gg P_\vartheta$) und

$$\forall\ \vartheta \in \Theta : \qquad \frac{dP_\vartheta}{d\mu}(x) = C(\vartheta) \cdot \exp\{\xi(\vartheta) \cdot T(x)\} \cdot h(x) [\mu]$$

mit Funktionen $C : \Theta \to \mathbb{R}_+^1$, $\xi : \Theta \to \mathbb{R}^1$, $T : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}^1, \mathbb{B}^1)$ und $h : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}^1_{(+)}, \mathbb{B}^1 \cap \mathbb{R}^1_{(+)})$.

**Beispiel 4.22:**

(a) (Poisson-Verteilung) $X_1, \ldots, X_n \sim \mathrm{Po}(\lambda)$ stochastisch unabhängig und $\lambda \in \mathbb{R}_+$. Gemeinsame Dichte bzgl. $\mu :=$Zählmaß auf $(\mathbb{N}_0, 2^{\mathbb{N}_0})$

$$f_\lambda(x) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = \underbrace{e^{-n\lambda}}_{=:C(x)} \exp\Big\{ \underbrace{(\ln\lambda)}_{\xi(x)} \underbrace{\sum_{i=1}^{n} x_i}_{T(x)} \Big\} \underbrace{\prod_{i=1}^{n} \frac{1}{x_i!}}_{h(x)}$$

Dann ist $\mathcal{P} = \Big\{ \bigotimes_{i=1}^{n} P_\lambda^{X_i} : \lambda > 0 \Big\}$ 1-parametrige Exponentialfamilie in $\xi(\lambda) = \ln\lambda$ und $T(x) = \sum_{i=1}^{n} x_i$.

(b) (Normal-Verteilung) $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma_0^1)$ stochastisch unabhängig, $\mu \in \mathbb{R}^1$ ($\sigma_0^2 > 0$ fest)

$$\begin{aligned} f_\mu(x) &= \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \left( \frac{x_i - \mu}{\sigma_0} \right)^2 \right\} = \\ &= \underbrace{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{ -\frac{n\mu^2}{2\sigma_0^2} \right\}}_{=C(\mu)} \cdot \exp\Big\{ \underbrace{-\frac{1}{\sigma_0^2}\mu}_{=\xi(\mu)} \underbrace{\sum_{i=1}^{n} x_i}_{=T(x)} \Big\} \underbrace{\exp\left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} x_i^2 \right\}}_{=h(x)} \end{aligned}$$

**Bemerkung 4.23 (Einfache Eigenschaften von 1-parametrigen Exponentialfamilien):**
Gegeben sei eine 1-parametrige Exponentialfamilie $\mathcal{P}$ in $\xi(\vartheta)$ und $T(x)$. Dann gilt

(a) $\dfrac{1}{C(\vartheta)} = \displaystyle\int e^{\xi(\vartheta)T(x)} h(x)\ d\mu(x)$

(b) $\xi : \Theta \to \mathbb{R}^1$ injektiv

(c) $T$ ist nicht $\mathcal{P}$-fast sicher konstant, d.h.

$$\forall\, c \in \mathbb{R}: \quad \exists\, \vartheta \in \Theta: \quad P_\vartheta(T \neq c) > 0$$

(d) $\nu$, definiert durch $\forall\, B \in \mathcal{B}$ mit $\nu(B) := \int\limits_B h(x)\, d\nu(x)$, ist $\sigma$-endliches Maß mit

$$\forall\, \vartheta \in \Theta: \quad \nu \sim P_\vartheta \qquad (\text{d.h. } \nu \gg P_\vartheta \text{ und } P_\vartheta \gg \nu)$$

(e) $\exists\, \nu$ $\sigma$-endlich: $\nu \sim \mathcal{P}$ und

$$\forall\, \vartheta \in \Theta: \quad \frac{dP_\vartheta}{d\nu}(x) = C(\vartheta) e^{\xi(\vartheta)T(x)} [\nu]$$

mit

$$\nu \sim \mathcal{P} \quad :\Longleftrightarrow \quad \nu \gg \mathcal{P} \text{ und } \forall\, N \in \mathcal{B} \text{ gilt } \Big[\forall\, \vartheta \in \Theta:\ P_\vartheta(N) = 0 \Rightarrow \nu(N) = 0\Big]$$

(f) $\forall\, \tau, \vartheta \in \Theta$ gilt: $P_\vartheta \sim P_\tau$

(g) Es ist $\mathcal{Z} := \big\{\xi \in \mathbb{R}^1:\ 0 < \int \exp\{\xi T(x)\} h(x)\ d\mu(x) < \infty\big\}$ konvex, also ein Intervall. $\mathcal{Z}$ heißt natürlicher Parameterraum von $\mathcal{P}$.

(h) $\xi$ und $T$ sind bis auf affine Transformationen bestimmt. Genauer gilt die Äquivalenz folgender zwei Aussagen:

(i) $\mathcal{P}$ ist 1-parametrige Exponentialfamilien in $\tilde{\xi}(\vartheta)$ und $\tilde{T}(x)$

(ii) $\exists\, a, b, c \in \mathbb{R},\ a \neq 0:\quad T = a\tilde{T} + b\,[\mathcal{P}], \quad \xi = \frac{1}{a}\tilde{\xi} + c$

(i) Ist $\mathcal{P}$ 1-parametrige Exponentialfamilie in $\xi(\vartheta)$ und $T(x)$, so ist

$$\mathcal{P}^{(n)} = \Big\{ \bigotimes_{i=1}^{n} P : P \in \mathcal{P} \Big\}$$

1-parametrige Exponentialfamilie in $\xi(\vartheta)$ und $\sum\limits_{i=1}^{n} T(x_i)$

**Satz 4.24:**
Sei $\mathcal{P}$ 1-parametrige Exponentialfamilie in $\xi(\vartheta)$ und $T(x)$, $\tau \in \Theta$, so daß

$$H = \big\{\vartheta \in \Theta : \xi(\vartheta) \leq \xi(\tau)\big\} \neq \emptyset, \quad K = \big\{\vartheta \in \Theta : \xi(\vartheta) > \xi(\tau)\big\} \neq \emptyset.$$

Dann hat $\mathcal{P}$ einen monotonen Dichtequotienten in $\tau$.

Beweis: Es ist

$$\frac{p_\vartheta(x)}{p_\tau(x)} = \frac{C(\vartheta)}{C(\tau)} \cdot \exp\big\{(\xi(\vartheta) - \xi(\tau))T(x)\big\} =: H_{\vartheta\tau}\big(T(x)\big)[\nu]$$

mit

$$H_{\vartheta\tau} \text{ monoton wachsend} \quad \Longleftrightarrow \quad \xi(\vartheta) > \xi(\tau) \quad \Longleftrightarrow \quad \vartheta \in H_1$$

$$H_{\vartheta\tau} \text{ monoton fallend} \quad \Longleftrightarrow \quad \xi(\vartheta) < \xi(\tau) \quad \Longleftrightarrow \quad \vartheta \in H \setminus \{\tau\}$$

Nach dem erweiterten Fundamentallemma existiert ein gleichmäßig bester Test zum Niveau $\alpha$ für $H : \xi \leq \xi_0$ vs. $K : \xi > \xi_0$ und natürlich auch für $\tilde{H} : \xi \geq \xi_0$ vs. $\tilde{K} : \xi < \xi_0$. Ist also $\xi : \Theta \to \mathbb{R}$ monoton wachsend (fallend), so gilt $\xi = \xi(\vartheta) < \xi(\vartheta_0) = \xi_0$ genau dann, wenn $\vartheta < \vartheta_0$ $(\vartheta > \vartheta_0)$.

Somit existiert ein gleichmäßig bester Test zum Niveau $\alpha$ für $H : \vartheta \leq \vartheta_0$ vs. $K : \vartheta > \vartheta_0$ und für $\tilde{H} : \vartheta \geq \vartheta_0$ vs. $\tilde{K} : \vartheta < \vartheta_0$

**Beispiel 4.22 - Fortsetzung:**

(a) (Poisson-Verteilung) $X_1, \ldots, X_n \sim \mathrm{Po}(\lambda)$ stu., $H : \lambda \leq \lambda_0$ vs. $K : \lambda > \lambda_0$. Es war

$$f_\lambda(x) = \underbrace{e^{-n\lambda}}_{=:C(x)} \exp\Big\{ \underbrace{(\ln \lambda)}_{\xi(x)} \underbrace{\sum_{i=1}^n x_i}_{T(x)} \Big\} \underbrace{\prod_{i=1}^n \frac{1}{x_i!}}_{h(x)}$$

Mit $\xi = \xi(\lambda)$ und $\xi_0 = \xi(\lambda_0)$ gilt somit: $\lambda < \lambda_0 \iff \xi \leq \xi_0$. Dann ist

$$\varphi^*(x) = \begin{cases} 1 & T(x) > k^* \\ \gamma^* & T(x) = k^* \\ 0 & T(x) < k^* \end{cases} \qquad \text{mit } E_{\lambda_0}\varphi^* = \alpha$$

gleichmäßig bester Test zum Niveau $\alpha$ für $H : K$. Wegen $T(x) = \sum\limits_{i=1}^n x_i$ und $\sum\limits_{i=1}^n X_i \sim \mathrm{Po}(n\lambda)$ sind $k^*$ und $\gamma^* \in [0,1]$ zu bestimmen aus

$$\sum_{j=1}^{k-1} e^{-n\lambda_0} \frac{(n\lambda_0)^j}{j!} + (1-\gamma^*) e^{-n\lambda_0} \frac{(n\lambda_0)^k}{k!} \stackrel{!}{=} 1-\alpha$$

bzw.

$$F_{n\lambda_0}(k-1) + (1-\gamma^*) f_{n\lambda_0}(k) \stackrel{!}{=} 1-\alpha$$

**Beispiel 4.25 (Stichprobenumfang, Versuchsplanung):**
Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma_0^2)$ stu., $\mu \in \mathbb{R}$ $(\sigma_0^2 > 0)$. $H : \mu \leq \mu_0$ vs. $K : \mu > \mu_0$. Problem: Wie groß muß der Stichprobenumfang $n$ sein, damit man $H$ bei Vorliegen von $\mu \geq \mu_1$ $(> \mu_0)$ mit Wahrscheinlichkeit $\beta$ ablehnt? (z.B. $\mu_0 = 100$, $\mu_1 = 101$, $\sigma_0^2 = 100$, $\beta = 0.9$)

Ziel: $\beta_{\varphi^*}(\mu) \geq \beta$ für alle $\mu \geq \mu_1$. Es war

$$\varphi^*(x) = \begin{cases} 1 & \sqrt{n}\frac{\overline{x}_\cdot - \mu_0}{\sigma_0} > u_\alpha \\ 0 & \sqrt{n}\frac{\overline{x}_\cdot - \mu_0}{\sigma_0} \leq u_\alpha \end{cases}$$

Für alle $\mu \geq \mu_1$ ist dann (vgl. Beispiel 4.16)

$$\begin{aligned} \beta_{\varphi^*}(\mu) = 1 - \Phi\left(u_\alpha - \sqrt{n}\frac{\mu_1-\mu_0}{\sigma_0}\right) \stackrel{!}{\geq} \beta \quad &\stackrel{\beta_{\varphi^*} \text{ mon.}}{\iff} \quad 1 - \Phi\left(u_\alpha - \sqrt{n}\frac{\mu_1-\mu_0}{\sigma_0}\right) \stackrel{!}{=} \beta \\ &\iff \quad u_\alpha - \sqrt{n}\frac{\mu_1-\mu_0}{\sigma_0} = \Phi^{-1}(1-\beta) \\ &\iff \quad n = \left(\frac{\sigma_0}{\mu_1-\mu_0}\big(u_\alpha - \Phi^{-1}(1-\beta)\big)\right)^2 \\ &\stackrel{\alpha=0.05}{\iff} \quad n \approx \left(\frac{10}{101-100}(1.6499 - 1.2816)\right)^2 \approx 13.17 \end{aligned}$$

Wähle also $n = 14$.

**Bemerkung 4.26:**
Tests für Hypothesen der Form

(1) $H : \vartheta \in [\vartheta_1, \vartheta_2]$ vs. $K : \vartheta \notin [\vartheta_1, \vartheta_2]$ $(\vartheta_1 \leq \vartheta_2)$ („innen:außen“) oder

(2) $H : \vartheta \notin [\vartheta_1, \vartheta_2]$ vs. $K : \vartheta \in [\vartheta_1, \vartheta_2]$ $(\vartheta_1 < \vartheta_2)$ („außen:innen“)

wurden bisher nicht behandelt. Während für (2) bei 1-parametrigen Exponentialfamilien sogar gleichmäßig beste Tests zum Niveau $\alpha$ der Form

$$\varphi^*(x) = \begin{cases} 1 & T(x) \in (k_1^*, k_2^*) \\ \gamma_1^* & T(x) = k_i^*,\ i = 1,2 \\ 0 & T(x) \notin (k_1^*, k_2^*) \end{cases} \qquad \text{mit } E_{\vartheta_1}\varphi^* = E_{\vartheta_2}\varphi^* = \alpha$$

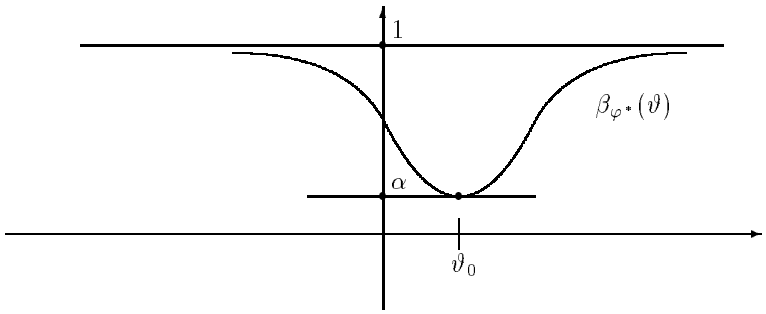
existieren, ist dies bei (1) i.a. nicht mehr der Fall. Man betrachtet deshalb nur noch unverfälschte Tests, d.h. Tests $\varphi$ mit $E_\vartheta \varphi \geq \alpha$ für alle $\vartheta \in K$. Bei 1-parametrigen Exponentialfamilien (mit natürlichem Parameterraum $\Theta$) ist ein gleichmäßig bester Test zum Nivau $\alpha$ für (1) von der Form

$$\varphi^*(x) = \begin{cases} 1 & T(x) \notin (k_1^*, k_2^*) \\ \gamma_i^* & T(x) = k_i^*,\ i = 1,2 \\ 0 & T(x) \in (k_1^*, k_2^*) \end{cases}$$

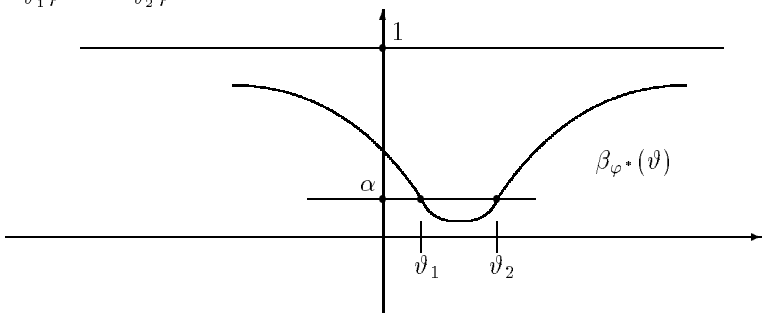
Dabei werden $k_i^*, \gamma_i^*$ bestimmt aus

(a) falls $\vartheta_1 = \vartheta_2 =: \vartheta_0$, d.h. $H : \vartheta = \vartheta_0$ vs. $K : \vartheta \neq \vartheta_0$: $E_{\vartheta_0}\varphi^* = \alpha$

$$\frac{d}{d\vartheta} E_\vartheta \varphi^* \Big|_{\vartheta = \vartheta_0} \stackrel{!}{=} 0 \quad \Longleftrightarrow \quad E_{\vartheta_0}(\varphi^* \cdot T) = \alpha E_{\vartheta_0} T$$




(b) falls $\vartheta_1 < \vartheta_2$: $E_{\vartheta_1}\varphi^* = E_{\vartheta_2}\varphi^* = \alpha$




**Beispiel 4.27 (2-seitiger Gauß-Test):**
$X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma_0^2)$ stu., $\mu \in \mathbb{R}$ $(\sigma_0^2 > 0)$. $H : \mu = \mu_0$ vs. $K : \mu \neq \mu_0$. Dann gilt

$$\varphi^*(x) = \begin{cases} 1 & \sqrt{n}\left|\frac{\overline{x}_\cdot - \mu_0}{\sigma_0}\right| \geq u_{\alpha/2} \\ 0 & \sqrt{n}\left|\frac{\overline{x}_\cdot - \mu_0}{\sigma_0}\right| < u_{\alpha/2} \end{cases}$$

ist UMPV-Test zum Niveau $\alpha$. Es gilt: $\Phi(u_{\alpha/2} - \Phi(-u_{\alpha/2}) = 1 - \alpha$.

**Beispiel 4.28 (2-seitiger Binomialtest):**
$X_1, \ldots, X_n \sim \mathrm{B}(1, p)$ stu., $p \in (0,1)$. $H : p = p_0$ vs. $K : p \neq p_0$. Dann ist

$$\varphi^*(x) = \begin{cases} 1 & T(x) \notin \{k_1^*, \ldots, k_2^*\} \\ \gamma_i^* & T(x) = k_i^*,\ i = 1,2 \\ 0 & T(x) \in \{k_1^* + 1, \ldots, k_2^* - 1\} \end{cases} \qquad \Big(T(x) = x_\cdot\Big)$$

mit $k_i^*, \gamma_i^*, \gamma_i^* \in [0,1]$, bestimmt durch

$$(1) \quad \sum_{x \in \{0,\ldots,k_1-1,k_2+1,\ldots,n\}} \binom{n}{x} p_0^x (1-p_0)^{n-x} + \sum_{j=1}^{2} \gamma_j \binom{n}{k_j} p_0^{k_j} (1-p_0)^{n-k_j} \stackrel{!}{=} \alpha$$

$$(2) \quad \sum_{x \in \{0,\ldots,k_1-1,k_2+1,\ldots,n\}} x \binom{n}{x} p_0^x (1-p_0)^{n-x} + \sum_{j=1}^{2} \gamma_j k_j \binom{n}{k_j} p_0^{k_j} (1-p_0)^{n-k_j} \stackrel{!}{=} \alpha n p_0$$

genau dann, wenn

$$(1') \quad F(k_2-1|n,p_0) - F(k_1|n,p_0) + \sum_{j=1}^{2} (1-\gamma_j) f(k_j|n,p_0) \stackrel{!}{=} 1-\alpha$$

$$(2') \quad F(k_2-2|n-1,p_0) - F(k_1-1|n-1,p_0) + \sum_{j=1}^{2} (1-\gamma_j) f(k_j-1|n-1,p_0) \stackrel{!}{=} 1-\alpha$$

mit $F(\cdot|m,p)$ bzw. $f(\cdot|m,p)$ Verteilungsfunktion bzw. Dichte der $\mathrm{B}(m,p)$-Verteilung.

**Bemerkung 4.29:**
Noch schwieriger wird die Situation, wenn $\vartheta$ mehrdimensional ist. Zum Beispiel

(1) $\vartheta = (\mu, \sigma^2)$ bei Normalverteilungen $\mathrm{N}(\mu, \sigma^2)$,

(2) $\vartheta = (p_1, p_2)$ bei einem 2-Stichproben-Binomialmodell $X_{ij} \sim \mathrm{B}(1, p_i)$, $j = 1, \ldots, n_i$, $i = 1, 2$, stu. Hypothese $p_1 = p_2$ vs. $p_1 \neq p_2$.

Dies führt zu sogenannten „bedingten Tests“.

**Beispiel 4.30 (1-seitiger t-Test):**
$X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ stu., $\mu \in \mathbb{R}$, $\sigma^2 > 0$. $H : \mu \leq \mu_0$ vs. $K : \mu > \mu_0$ $(\mu_0 \in \mathbb{R})$.

Behauptung:

$$T(X) = \sqrt{n} \frac{\overline{X}_\cdot - \mu_0}{S} \stackrel{\mu_0}{\sim} \mathrm{t}_{n-1} \qquad \text{mit } S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_\cdot)^2$$

Beweis: Sei $Q \in \mathbb{R}^{n \times n}$ orthonormal, d.h. $Q^T = Q^{-1}$, also etwa

$$Q = \begin{bmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ & * & \end{bmatrix} \qquad (* = \text{orthonormal ergänzt})$$

$X = (X_1, \ldots, X_n)^T \Rightarrow X \stackrel{\mu_0}{\sim} \mathrm{N}_n(\mu_0 1_n, \sigma^2 I_n)$. Dann ist $Y := QX \sim \mathrm{N}_n(\nu, \sigma^2 I_n)$ mit

$$\nu = \mu_0 Q 1_n = \begin{pmatrix} \sqrt{n}\mu_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

und somit $Y_1 = \sqrt{n}\overline{X}_\cdot$. Es gilt: $Y_1, \ldots, Y_n$ stochastisch unabhängig mit

$$\sum_{i=2}^{n} Y_i^2 = Y^T Y - Y_1^2 = (QX)^T (QX) - n\overline{X}_\cdot^2 \stackrel{Q \text{ orthon.}}{=} \sum_{i=1}^{n} X_i^2 - n\overline{X}_\cdot^2 = \sum_{i=1}^{n} (X_i - \overline{X}_\cdot)^2$$

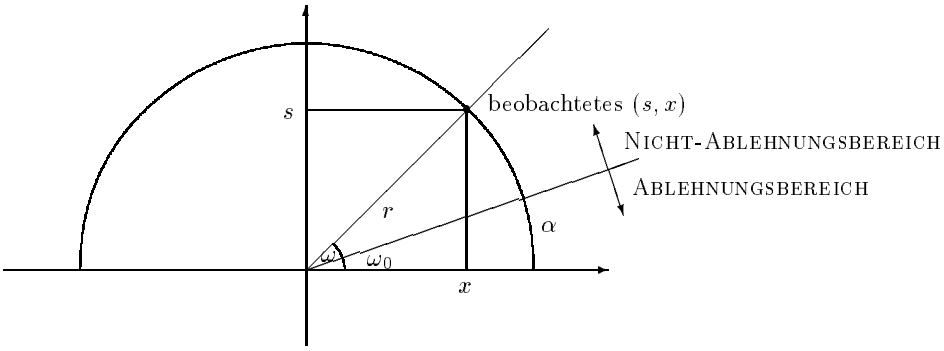
Somit folgt

$$T(X) = \frac{\frac{1}{\sigma}(Y_1 - \sqrt{n}\mu_0)}{\left(\frac{1}{n-1}\sum_{i=2}^{n}\frac{Y_i^2}{\sigma^2}\right)^{\frac{1}{2}}} \begin{cases} \overset{\mu_0}{\sim} \frac{\mathrm{N}(0,1)}{\sqrt{\frac{1}{n-1}\chi^2_{n-1}}} = \mathrm{t}_{n-1} \\ \overset{\mu \neq \mu_0}{\sim} \mathrm{t}_{n-1}\Big(\underbrace{\sqrt{n}\frac{\mu-\mu_0}{\sigma}}_{\text{Nicht-Zentralitätsparameter}}\Big) \qquad \text{(nicht-zentrale } \mathrm{t}_{n-1}\text{-Verteilung)} \end{cases}$$

Es folgt:

$$\varphi^*(x) = \begin{cases} 1 & T(x) \geq \mathrm{t}_{n-1,\alpha} \\ 0 & T(x) < \mathrm{t}_{n-1,\alpha} \end{cases}$$

mit $P(W \leq \mathrm{t}_{n-1,\alpha}) = 1 - \alpha$ $(W \sim \mathrm{t}_{n-1})$ ist UMPV-Test zum Niveau $\alpha$ für $H : K$.




$$\left.\begin{matrix} X \sim \mathrm{N}(0,\sigma^2) \\ \frac{S^2}{\sigma^2} \sim \chi^2_\nu \end{matrix} \quad \text{stu.} \right\} \; R^2 = S^2 + X^2, \quad \tan\Omega = \frac{S}{X} \Rightarrow R, \Omega \text{ stu.}$$

Somit ist $P^\Omega$ unabhängig von $\sigma^2$ und

$$P(\Omega \leq \omega) = \frac{1}{2} F_{\mathrm{Beta}\left(\frac{\nu}{2},\frac{1}{2}\right)}(\sin^2\omega), \qquad \omega \in \left[0, \frac{\pi}{2}\right]$$

und es gilt

$$P(\Omega \leq \omega | R = r) \overset{R,\Omega \text{ stu}}{=} P(\Omega \leq \omega).$$

Wählt man $\forall\, r > 0$ $\omega_0$ so, daß $P(\Omega \leq \omega_0) = \alpha$, ist

$$\begin{aligned} P(\Omega \leq \omega_0) &= \int P(\Omega \leq \omega_0 | R = r)\, dP^R(r) = \\ &= \int P(\Omega \leq \omega_0)\, dP^R(r) = \alpha \int 1\, dP^R(r) = \alpha \end{aligned}$$

Es gilt

$$\begin{aligned} \Omega \leq \omega_0 &\iff \tan\Omega \leq \tan\omega_0 \\ &\iff \frac{S}{X} \leq \tan\omega_0 \\ &\iff \frac{X}{\frac{1}{\sqrt{\nu}}S} \geq \frac{\sqrt{\nu}}{\tan\omega_0} \end{aligned}$$

Also

$$\begin{aligned} P(\Omega \leq \omega_0) \overset{!}{=} \alpha \quad &\overset{\alpha < \frac{1}{2}}{\iff} \quad \sin^2\omega_0 = F^{-1}_{\mathrm{Beta}\left(\frac{\nu}{2},\frac{1}{2}\right)}(2\alpha) \\ &\iff \quad \omega_0 = \arcsin\left(\sqrt{F^{-1}_{\mathrm{Beta}\left(\frac{\nu}{2},\frac{1}{2}\right)}(2\alpha)}\right) \end{aligned}$$

Es gilt: $\frac{\sqrt{\nu}}{\tan \omega_0} = \mathrm{t}_{\nu,\alpha}$.

**Beispiel 4.31 (2-Stichproben t-Test):**
Modell: $X_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$, $j = 1, \ldots, n_i$, $i = 1, 2$, stu., $\mu_i \in \mathbb{R}$, $\sigma^2 > 0$ $(n_i \in \mathbb{N} : n := n_1 + n_2 \geq 3)$. Somit

$$\varphi^*(x) = \begin{cases} 1 & \sqrt{\frac{n_1 n_2}{n}} \left| \frac{\overline{x}_{1\cdot} - \overline{x}_{2\cdot}}{s} \right| > \mathrm{t}_{n-2,\frac{\alpha}{2}} \\ 0 & \sqrt{\frac{n_1 n_2}{n}} \left| \frac{\overline{x}_{1\cdot} - \overline{x}_{2\cdot}}{s} \right| \leq \mathrm{t}_{n-2,\frac{\alpha}{2}} \end{cases}$$

mit $s^2 = \frac{1}{n-2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_{i\cdot})^2$ und $\mathrm{t}_{n-2,\frac{\alpha}{2}}$ $\frac{\alpha}{2}$-Fraktil der $\mathrm{t}_{n-2}$-Verteilung. Es gilt:

$$\sqrt{n_1 n_2} n \frac{\overline{X}_{1\cdot} - \overline{X}_{2\cdot}}{S} \overset{\mu_1 = \mu_2}{\sim} \mathrm{t}_{n-2}$$

**Beispiel 4.32 (2-Stichproben Binomialtest, $(2 \times 2)$-Tafel, exakter Test von Fischer):**
$X_{ij} \sim \mathrm{B}(1, p_i)$, $j = 1, \ldots, n_i$, $i = 1, 2$, stu., $p_i \in (0, 1)$. Hypothese $H : p_1 \leq p_2$ vs. $K : p_1 > p_2$ (oder auch $\tilde{H} : p_1 = p_2$ vs. $\tilde{K} : p_1 \neq p_2$), zum Beispiel

$$X_{ij} = \begin{cases} 1 & \text{Medikament } i \text{ bei Patient } j \text{ in Gruppe } i \text{ wirksam} \\ 0 & \text{Medikament } i \text{ bei Patient } j \text{ in Gruppe } i \text{ nicht wirksam} \end{cases}$$

Klar: $X_{i\cdot} \sim \mathrm{B}(n_i, p_i)$, $i = 1, 2$, stu., $X_{\cdot\cdot} \sim \mathrm{B}(n_\cdot, p)$, falls $p := p_1 = p_2$

| | Anz. Erfolge | Anz. Mißerfolge | $\sum$ |
|---|---|---|---|
| Gruppe 1 | $X_{1\cdot}$ | $n_1 - X_{1\cdot}$ | $n_1$ |
| Gruppe 2 | $X_{2\cdot}$ | $n_2 - X_{2\cdot}$ | $n_2$ |
| | $X_{\cdot\cdot}$ | $n_\cdot - X_{\cdot\cdot}$ | $n_\cdot$ |

Idee: Wäre $X_{1\cdot} + X_{2\cdot} = v$ fest gegeben, so sollte man $H$ (bzw. $\tilde{H}$) ablehnen, wenn $X_{1\cdot}$ „groß“ ist (bzw. „groß“ oder „klein“).

Nun gilt für $u \in \{0, 1, \ldots, v\}$, $v \in \{0, 1, \ldots, u\}$, $U := X_{1\cdot}$, $V := X_{\cdot\cdot}$

$$\begin{aligned} & P_{(p_1,p_2)}(U = u, V = v) = \\ &= P_{(p_1,p_2)}(X_{1\cdot} = u, X_{\cdot\cdot} = v) = P_{(p_1,p_2)}(X_{1\cdot} = u, X_{2\cdot} = v - u) = \\ &= \binom{n_1}{u} p_1^u (1-p_1)^{n_1-u} \binom{n_2}{v-u} p_2^{v-u} (1-p_2)^{n_2-(v-u)} = \\ &= \underbrace{\binom{n_1}{u}\binom{n_2}{v-u}}_{=:h(u,v)} \underbrace{(1-p_1)^{n_1}(1-p_2)^{n_2}}_{=:C(\eta,\xi)} \exp\left\{ u \cdot \underbrace{\ln\left(\frac{p_1(1-p_2)}{(1-p_1)p_2}\right)}_{=:\eta} + v \cdot \underbrace{\ln\left(\frac{p_2}{1-p_2}\right)}_{=:\xi} \right\} = \\ &=: P_{\eta,\xi}(U = u, V = v) \end{aligned}$$

und

$$\begin{array}{lcl} p_1 < p_2 & & \eta > 0 \\ p_1 = p_2 & \Longleftrightarrow & \eta = 0 \\ p_1 > p_2 & & \eta < 0 \end{array}$$

$$
\begin{aligned}
P_{\eta,\xi}(U=u|V=v) &= \frac{P_{\eta,\xi}(U=u,V=v)}{P_{\eta,\xi}(V=v)} = \frac{P_{\eta,\xi}(X_1.=u,X_2.=v-u)}{\sum_{u'=0}^{n_1} P_{\eta,\xi}(X_1.=u',V=v)} = \\
&= \frac{C(\eta,\xi)\exp\{u\eta+v\xi\}h(u,v)}{\sum_{u'=0}^{n_1} C(\eta,\xi)\exp\{u'\eta+v\xi\}h(u',v)} = \underbrace{\frac{1}{\sum_{u'=0}^{n_1}\exp\{u'\eta\}h(u',v)}}_{\tilde{C}_v(\eta)} \cdot \exp\{u\eta\}h(u,v) = \\
&=: P_\eta^{U|V=v}(\{u\})
\end{aligned}
$$

bei festem $v$ unabhängig von $\xi$ und 1-parametriger Exponentialfamilie in $\eta$ und $u$. D.h. wir haben einen monotonen Dichtequotienten, können das erweiterte Fundamentallemma anwenden und erhalten

$$
\psi_v^*(u) = \begin{cases} 1 & u > k(v) \\ \gamma(v) & u = k(v) \\ 0 & u < k(v) \end{cases} \qquad \text{mit } E_{\eta=0}\psi_v^* = \int \psi_v^*(u)\, dP_\eta^{U|V=v}(u) \stackrel{!}{=} \alpha
$$

als gleichmäßig besten Test zum Niveau $\alpha$ für $\tilde{H}_v : \eta \le 0$ vs. $\tilde{K}_v : \eta > 0$. Setze $\psi^*(u,v) := \psi_v^*(u)$. Dann ist $\forall\, \eta \le 0$

$$
\begin{aligned}
E_\eta\psi^*(\cdot,\cdot) &= \int \psi^*(u,v)\, dP_{\eta,\xi}^{(U,V)}(u,v) = \sum_u\sum_v \psi^*(u,v) P_{\eta,\xi}^{(U,V)}(\{(u,v)\}) = \\
&= \sum_{v=0}^{n.} \underbrace{\sum_{u=0}^{v} \psi^*(u,v) P_\eta(U=u|V=v)}_{\le\alpha}\, P_{\eta,\xi}(V=v) \le \alpha
\end{aligned}
$$

und „=“ für $\eta = 0$. Damit ist $\psi^*(\cdot,\cdot)$ Test zum Niveau $\alpha$ für $H : K$. Weiter kann man zeigen: $\psi^*(\cdot,\cdot)$ ist gleichmäßig bester unverfälschter Test zum Niveau $\alpha$ für $H : K$.

Zur Berechnung des kritischen Wertes $k(v)$ benötigen wir noch $P_{\eta=0}^{U|V=v}$. Es ist (s.o.)

$$
\begin{aligned}
P_{\eta=0}^{U|V=v}(\{u\}) &= \tilde{C}_v(0)e^{u\cdot 0}h(u,v) = \left(\sum_{u'=0}^{n_1} h(u',v)\right)^{-1} h(u,v) = \\
&= \Big(\underbrace{\sum_{u'=0}^{n_1}\binom{n_1}{u'}\binom{n_2}{v-u'}}_{=\binom{n_1+n_2}{v}}\Big)^{-1}\binom{n_1}{u}\binom{n_2}{v-u} = \frac{\binom{n_1}{u}\binom{n_2}{v-u}}{\binom{n_1+n_2}{v}} = \\
&\hat{=} \quad \text{Dichte der hypergeometrischen Verteilung } \mathrm{H}(n.,v,n_1)
\end{aligned}
$$

d.h.

$$
\int \psi_v^*(u)\, dP_{\eta=0}^{U|V=v} \stackrel{!}{=} \alpha \quad \Longleftrightarrow \quad \sum_{u>k} \frac{\binom{n_1}{u}\binom{n_2}{v-u}}{\binom{n_1+n_2}{v}} + \gamma\frac{\binom{n_1}{k}\binom{n_2}{v-k}}{\binom{n_1+n_2}{v}} \stackrel{!}{=} \alpha
$$

Entsprechend berechnet man $k_i, \gamma_i$ für den 2-seitigen Test $\tilde{\psi}_v^*$ für $\tilde{H} : p_1 = p_2$ vs. $\tilde{K} : p_1 \ne p_2$ aus

(1) $\displaystyle\sum_{u<k_1} g(u) + \gamma_1 g(k_1) + \gamma_2 g(k_2) + \sum_{u>k_2} g(u) = \alpha$

(2) $\displaystyle\sum_{u<k_1-1} \tilde{g}(u) + \gamma_1\tilde{g}(k_1-1) + \gamma_2\tilde{g}(k_2-1) \sum_{u>k_2-1} \tilde{g}(u) = \alpha$

mit $g$ bzw. $\tilde{g}$ Zähldichte der $\mathrm{H}(n.,v,n_1)$- bzw. $\mathrm{H}(n.-1,v-1,n_1-1)$-Verteilung.

**Bemerkung 4.33:**
Bei 2-seitigen Hypothesen werden häufig nicht die exakten Tests, sondern einfach zwei einseitige nichtrandomisierte Tests verwendet (in der Hoffnung, daß diese sich nicht allzusehr von den exakten Tests unterscheiden.

Im Beispiel 4.32 bestimmt man für $\tilde{H} : p_1 = p_2$ die kritischen Werte $k_1, k_2$ einfach aus

$$\sum_{u \le k_1} P_{\eta=0}(U = u | V = v) \le \frac{\alpha}{2} \qquad k_1 \text{ möglichst „groß“}$$

$$\sum_{u \ge k_2} P_{\eta=0}(U = u | V = v) \le \frac{\alpha}{2} \qquad k_2 \text{ möglichst „klein“}$$

Lehne $\tilde{H}$ ab, falls $X_{1\cdot} \le k_1$ oder $X_{1\cdot} \ge k_2$.

**Bemerkung 4.34:**
In statistischen Programmpaketen (z.B. SAS, SPSS) werden häufig sogenannte $p$-Werte (Überschreitungswahrscheinlichkeiten, $p$-values) zum Testen einer Hypothese ausgegeben. Ist z.B. $T$ eine Teststatistik zum Testen von $H : K$, die unter $K$ zu „größren Werten neigt“ und $x \in \mathcal{X}$ beobachtet wurde, so heißt

$$p(x) := \sup_{\vartheta \in H} P_\vartheta(T \ge T(x))$$

$p$-Wert (zum Testen von $H : K$). Es gilt:

$$\forall\, \vartheta \in H : \quad P_\vartheta(p(x) \le \alpha) \le \alpha,$$

d.h. $\varphi(x) = \begin{cases} 1 & p(x) \le \alpha \\ 0 & p(x) > \alpha \end{cases}$ ist Test zum Niveau $\alpha$ für $H : K$.

## 4.4 Konfidenzbereiche

Generelle Voraussetzungen: $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\})$ statistischer Raum, $\gamma : \Theta \to \Gamma$, z.B. für $\varphi = (\mu, \sigma^2)$: $\gamma(\vartheta) = \mu$, $\gamma := \gamma(\vartheta)$. Schätzproblem: $\hat{\gamma} =?$ bzw. $\gamma \in ?$

(i) Punktschätzer: $\hat{\gamma} : (\mathcal{X}, \mathcal{B} \to (\tilde{\Gamma}, \cdots)$, $\tilde{\Gamma} \supseteq \Gamma$. Nachteil von Punktschätzern: Häufig gilt:

$$P_\vartheta\big(\hat{\gamma}(\cdot) = \gamma(\vartheta)\big) = 0$$

(ii) Bereichsschätzer: $\hat{\gamma} : (\mathcal{X}, \mathcal{B}) \to (2^\Gamma, \mathcal{A}_\Gamma)$ mit $\mathcal{A}_\Gamma$ geeigneter $\sigma$-Algebra.

Achtung: $\hat{\gamma}$ ist Schätzer (als Zufallsvariable), $\hat{\gamma}(x)$ ist Schätzung (nach Beobachtung von $x$).

**Definition 4.35:**

(a) $C : \mathcal{X} \to 2^\Gamma$ heißt Bereichsschätzer (BS) für $\gamma \in \Gamma :\iff$

$$\forall\, \gamma' \in \Gamma : \quad A(\gamma') := \{x : C(x) \ni \gamma'\} \in \mathcal{B}$$

(b) Sei $C : \mathcal{X} \to 2^\Gamma$ Bereichsschätzer für $\gamma \in \Gamma$ und sei $\alpha \in [0, 1]$. Dann heißt $C := (C(x) : x \in \mathcal{X})$ Familie von Konfidenzbereichen (KB) zum Konfidenzniveau $1 - \alpha$ für $\gamma \in \Gamma :\iff$

$$\forall\, \vartheta \in \Theta : \quad P_\vartheta\Big(\{x : C(x) \ni \gamma(\vartheta)\}\Big) \ge 1 - \alpha$$

Kurzsprechweise: $C$ ist $(1 - \alpha)$-Konfidenzbereich. Es bezeichne $\varphi_{1-\alpha}(\Gamma)$ die Menge aller $(1 - \alpha)$-Konfidenzbereiche $C$ für $\gamma \in \Gamma$.

**Beispiel 4.36:**
$X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ stu., $\mu \in \mathbb{R}$, $\sigma^2 > 0$, $\vartheta = (\mu, \sigma^2)$, $\Gamma(\vartheta) := \mu$. Dann ist

(i) für alle $(\mu, \sigma^2)$

$$1 - \alpha = P_{\mu,\sigma^2}\left(\sqrt{n}\frac{\overline{X}_\cdot - \mu}{S} \leq u_\alpha\right) = P_{\mu,\sigma^2}\left(\mu \geq \overline{X}_\cdot - \frac{S u_\alpha}{\sqrt{n}}\right)$$

Damit ist $C = \left(C(x) = \left[\overline{x}_\cdot - \frac{s u_\alpha}{\sqrt{n}}, \infty\right) : x \in \mathbb{R}^n\right)$ ein $(1-\alpha)$-Konfidenzbereich für $\mu$. Es heißt $C(x)$ 1-seitiges Konfidenzintervall für $\mu$. $\underline{\mu}(x) := \overline{x}_\cdot - \frac{s u_\alpha}{\sqrt{n}}$ untere $(1-\alpha)$-Konfidenzschranke.

(ii)

$$1 - \alpha = P_{\mu,\sigma^2}\left(\sqrt{n}\frac{|\overline{X}_\cdot - \mu|}{S} \leq u_{\frac{\alpha}{2}}\right) = P_{\mu,\sigma^2}\left(\overline{X}_\cdot - \frac{s u_{\frac{\alpha}{2}}}{\sqrt{n}} \leq \mu \leq \overline{X}_\cdot + \frac{s u_{\frac{\alpha}{2}}}{\sqrt{n}}\right)$$

Damit ist

$$C = \left(C(x) = \left[\overline{x}_\cdot - \frac{s u_{\frac{\alpha}{2}}}{\sqrt{n}}, \overline{x}_\cdot + \frac{s u_{\frac{\alpha}{2}}}{\sqrt{n}}\right] : x \in \mathbb{R}^n\right)$$

$(1-\alpha)$-Konfidenzbereich für $\mu$. $C(x)$ heißt 2-seitiges Konfidenzintervall.

Eine allgemeine Konstruktionsmethode für Konfidenzbereiche beruht auf foldendem Satz. Dazu vorher eine Bezeichnung:

$$\begin{aligned}\Phi_\alpha(\Gamma) \quad := \quad & \Big\{\varphi = (\varphi_\gamma : \gamma \in \Gamma) : \forall\, \gamma \in \Gamma : \ \varphi_\gamma : (\mathcal{X}, \mathcal{B}) \to (\{0,1\}, 2^{\{0,1\}}), \\ & \varphi_\gamma \text{ Test z. N. } \alpha \text{ für } H_\gamma := \{\vartheta \in \Theta : \gamma(\vartheta) = \gamma\} \text{ vs. } K_\gamma := \Theta \setminus H_\gamma\Big\}\end{aligned}$$

**Satz 4.37 (Korrespondenzsatz, Dualität von Tests und Konfidenzbereichen):**

(a) Sei $\varphi = (\varphi_\gamma : \gamma \in \Gamma) \in \Phi_\alpha(\Gamma)$ und sei $\forall\, x \in \mathcal{X}$ $C(x) := \{\gamma \in \Gamma : \varphi_\gamma(x) = 0\}$, so daß $C = (C(x) : x \in \mathcal{X})$ Bereichsschätzer, dann gilt $C \in \varphi_{1-\alpha}(\Gamma)$.

(b) Sei $C \in \varphi_{1-\alpha}(\Gamma)$ und $\forall\, x \in \mathcal{X} : \ \forall\, \gamma \in \Gamma : \ \varphi_\gamma(x) = 1 - I_{C(x)}(\gamma)$. Dann gilt

$$\varphi = (\varphi_\gamma : \gamma \in \Gamma) \in \Phi_\alpha(\Gamma).$$

Beweis:

(a) Für alle $\vartheta \in \Theta$:

$$P_\vartheta\Big(\{x : C(x) \ni \Gamma(\vartheta)\}\Big) \overset{\text{Def. } \varphi}{=} P_\vartheta\Big(\{x : \varphi_{\gamma(\vartheta)}(x) = 0\}\Big) \overset{(*)}{\geq} 1 - \alpha$$

$(*)$ da $\varphi_\gamma$ Test zum Niveau $\alpha$ für $H_\gamma = \{\vartheta \in \Theta : \gamma(\vartheta) = \gamma\}$ vs. $K_\gamma$.

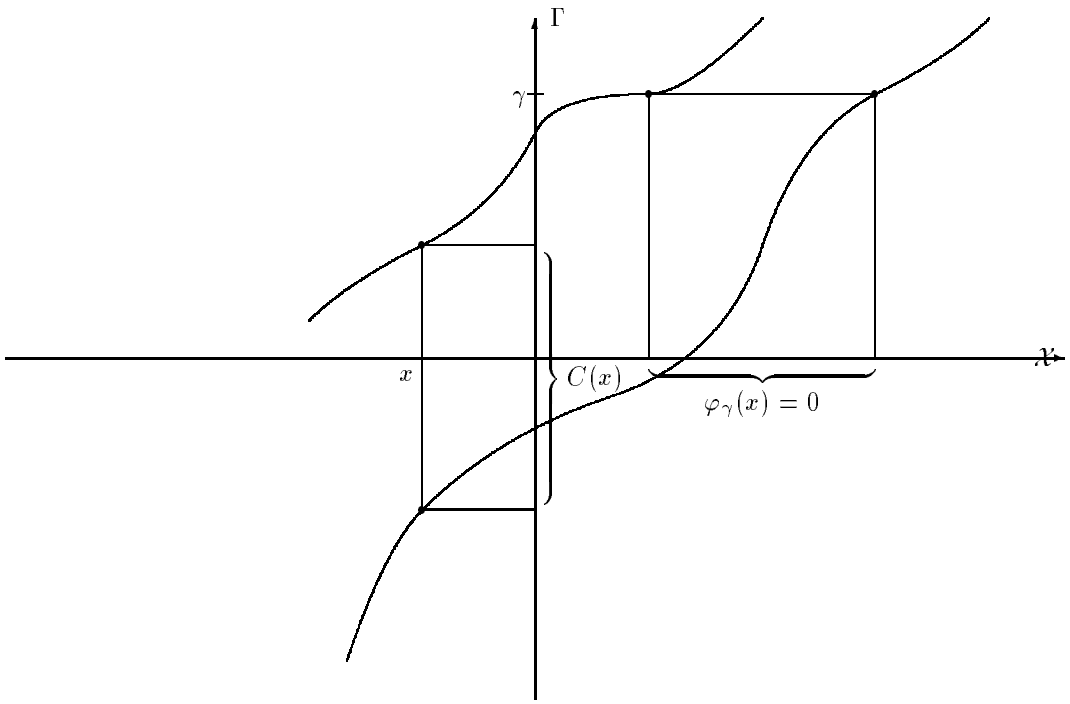
(b) Für alle $\vartheta \in H_\vartheta$:

$$P_\vartheta(\varphi_\gamma = 0) = P_\vartheta\Big(\{x : 1 - I_{C(x)}(\gamma) = 0\}\Big) = P_\vartheta\Big(\{x : C(x) \ni \gamma\}\Big) \overset{(*)}{\geq} 1 - \alpha$$

$(*)$ da $C \in \varphi_{1-\alpha}(\Gamma)$.

**Bemerkung 4.38:**
Die Dualität von Tests und Konfidenzbereichen wird häufig wie folgt beschrieben:

$$C(x) \ni \gamma \quad \Longleftrightarrow \quad \varphi_\gamma(x) = 0$$




**Bemerkung 4.39**
ACHTUNG: Nach Beobachtung von $x \in \mathcal{X}$ bedeutet die Angabe von $C(x)$ nicht

„$\gamma$ liegt mit Wahrscheinlichkeit $1-\alpha$ in $C(x)$"

Über $\Gamma$ ist kein Wahrscheinlichkeitsmaß gegeben! $C(x)$ hängt offensichtlich von der aktuellen Beobachtung von $x$ ab, ist also zufällig. Das Konfidenzniveau $1-\alpha$ besagt nur, daß bei $N$-maliger unabhängiger Wiederholung des Experimentes der Anteil mit $C(x) \ni \gamma$ ($\gamma$ der wahre Parameter) gegen $1-\alpha$ (meist noch $\geq 1-\alpha$) strebt.

$$\frac{1}{N}\Big|\big\{x_i : C(x_i) \ni \gamma,\ i \in \{1,\ldots,N\}\big\}\Big| \overset{N\to\infty}{\longrightarrow} 1-\alpha \qquad (\geq 1-\alpha) \text{ f.s.}$$

Die gleiche Interpretation gilt natürlich auch für Niveau-$\alpha$-Tests für $H : K$. Wird das Experiment $N$-mal unabhängig wiederholt mit $\vartheta \in H$ und den Ausgängen $x_1, \ldots, x_N$, so gilt:

$$\frac{1}{N}\Big|\big\{x_i : \varphi(x_i) = 1\big\}\Big| \overset{N\to\infty}{\longrightarrow} \alpha \qquad (\leq \alpha) \text{ f.s.}$$

(frequentistischer Ansatz der Statistik)

**Beispiel 4.40 (Konfidenzellipsoide):**
$X \sim \mathrm{N}(\mu, \Sigma_0)$, $\mu \in \mathbb{R}^p$ ($\Sigma_0 \in \mathbb{R}^{p\times p}_+ > 0$). Hypothese $H_{\mu'} : \mu = \mu'$ vs. $K_{\mu'} : \mu \neq \mu'$ mit $\mu' \in \mathbb{R}^p$ ($\Theta = \mathbb{R}^p$).

Test für $H_{\mu'}$:

$$\varphi_{\mu'}(x) = \begin{cases} 1 & T(x) \geq \chi^2_{p,\alpha} \\ 0 & T(x) < \chi^2_{p,\alpha} \end{cases}$$

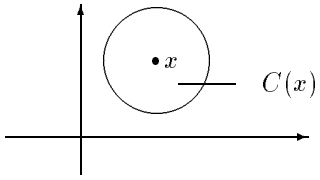
mit $T(x) = (X-\mu')^T\Sigma_0^{-1}(X-\mu')$ und $P(\chi_p^2 \le \chi_{p,\alpha}^2) = 1-\alpha$. Es gilt $T(X) \overset{\mu=\mu'}{\sim} \chi_p^2$, denn

$$Y \sim \Sigma_0^{-\frac{1}{2}}(X-\mu') \overset{\mu=\mu'}{\sim} \mathrm{N}_p(0, I_p) \qquad Y^TY = \sum_{i=1}^{p} Y_i^2 \overset{\mu=\mu'}{\sim} \chi_p^2$$

Somit ist $C = \big(C(x) : x \in \mathbb{R}^p\big) \in \varphi_{1-\alpha}(\Theta)$ mit

$$C(x) = \big\{\mu' : \varphi_{\mu'}(x) = 0\big\} = \big\{\mu' : (x-\mu')^T\Sigma_0^{-1}(x-\mu') \le \chi_{p,\alpha}^2\big\}$$

$(1-\alpha)$-Konfidenzellipsoid, z.B. $p = 2$ und $\Sigma_0 = I_2$




**Bemerkung 4.41:**

(a) Konfidenzbereiche sind „informativer“ als Tests. Ist zum Beispiel $\varphi$ ein (nicht-randomisierter) Test zum Niveau $\alpha$ für $H : K$, so faßt man das Testergebnis formal als Entscheidung für $\Theta$, falls $\varphi(x) = 0$ bzw. $K$, falls $\varphi(x) = 1$. Setzt man

$$C_\varphi(x) = \begin{cases} \Theta & \varphi(x) = 0 \\ K & \varphi(x) = 1 \end{cases}$$

so ist $C_\varphi = \big(C_\varphi(x) : x \in \mathcal{X}\big)$ ein $(1-\alpha)$-Konfidenzbereich für $\vartheta \in \Theta$, denn

$$P_\vartheta\big(\{x : C_\varphi(x) \ni \vartheta\}\big) = \left\{ \begin{array}{ll} P_\vartheta\big(\{x : \varphi(x) = 0\}\big) \ge 1-\alpha & \vartheta \in H \\ P_\vartheta\big(\{x : \varphi(x) = 0 \vee \varphi(x) = 1\}\big) = 1 & \vartheta \in K \end{array} \right\} \ge 1-\alpha$$

Beobachtet man ein $x \in \mathcal{X}$, so daß $C_\varphi(x) = \Theta$, so ist dies natürlich wenig „informativ“.

(b) Will man zusätzlich zur Entscheidung eines Tests $\varphi$ einen $(1-\alpha)$-Konfidenzbereich $C$ angeben, so sollten $\varphi$ und $C$ „kompatibel“ sein, d.h. man fordert, daß für alle $x \in \mathcal{X}$ gilt:

$$C_\varphi(x) \supseteq C(x).$$

Sind $\varphi$ und $C$ kompatibel und $C \in \varphi_{1-\alpha}(\Gamma)$, so gilt:

$$\forall\, \vartheta \in \Theta : \quad P_\vartheta\Big(\underbrace{\{x : C_\varphi(x) \ni \vartheta\}}_{\text{d.h. kein Fehler 1. Art}} \cap \{x : C(x) \ni \vartheta\}\Big) \ge 1-\alpha$$

Beispiel hierzu: 1-seitiger t-Text für $\mu \le 0$ gegen $\mu > 0$ ist kompatibel mit dem entsprechenden 1-seitigen Konfidenzintervall.

## 4.5 Maximum-Likelihood-Schätzer (ML-Schätzer)

Generelle Voraussetzungen: $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\})$ statistischer Raum, $p_\vartheta$ $\mu$-Dichte von $P_\vartheta$, $\vartheta \in \Theta$.

**Definition 4.42:**

(a) $l(\vartheta|x) := p_\vartheta(x)$ heißt Likelihood-Funktion (bei gegebenem $x \in \mathcal{X}$)

(b) $L(\vartheta|x) := \log p_\vartheta(x)$ heißt Log-Likelihood-Funktion (bei gegebenem $x \in \mathcal{X}$)

(c) Für $x \in \mathcal{X}$ heißt jeder Wert $\hat{\vartheta}(x) \in \overline{\Theta}$ Maximum-Likelihood-Schätzer für $\vartheta \in \overline{\Theta} :\Longleftrightarrow$ $\hat{\vartheta}(x)$ ist eine Lösung $t$ der Gleichung

$$l(t|x) = \sup_{\vartheta \in \overline{\Theta}} l(\vartheta|x).$$

Es heißt $\hat{\vartheta}$ ML-Schätzer.

**Beispiel 4.43:**

(a) $X_1, \ldots, X_n \sim \mathrm{B}(1, \vartheta)$ stu., $\vartheta \in (0,1)$. Dann ist $l(\vartheta|x) = \vartheta^{x_\cdot}(1-\vartheta)^{n-x_\cdot}$

$$\begin{aligned} \frac{d}{d\vartheta} l(\vartheta|x) &= x_\cdot \vartheta^{x_\cdot - 1}(1-\vartheta)^{n-x_\cdot} - (n-x_\cdot)\vartheta^{x_\cdot}(1-\vartheta)^{n-x_\cdot-1} = \\ &= \vartheta^{x_\cdot-1}(1-\vartheta)^{n-x_\cdot-1}\Big(\underbrace{x_\cdot(1-\vartheta) - (n-x_\cdot)\vartheta}_{x_\cdot = n\vartheta \iff \vartheta = \overline{x}_\cdot}\Big) \stackrel{!}{=} 0 \end{aligned}$$

Dann ist $\hat{\vartheta}(x) = \overline{x}_\cdot$ ML-Schätzer für $\vartheta$, falls $0 < x_\cdot < n$ ist. Für $x_\cdot = 0$ ist $l(\vartheta|x) = (1-\vartheta)^n$, maximal für $\hat{\vartheta}(x) = 0 \notin \Theta$, für $x_\cdot = n$ ist $l(\vartheta|x) = \vartheta^n$, maximal für $\hat{\vartheta}(x) = 1 \notin \Theta$, aber jeweils $\in \overline{\Theta}$.

(b) $X \sim \mathrm{B}(n, \vartheta)$, $\vartheta \in [0,1] =: \Theta$

$$l(\vartheta|x) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x}$$

$\hat{\vartheta}$ definiert durch $\hat{\vartheta}(x) = \frac{x}{n}$ ist ML-Schätzer für $\vartheta \in \Theta$ (s.o.)

(c) $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ stochastisch unabhängig, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ und $\vartheta = (\mu, \sigma^2)$. Da $l(\vartheta|x) \to \log l(\vartheta|x)$ (nur) eine monotone Transformation ist, können wir auch eine Lösung suchen von

$$\begin{aligned} \log l(\vartheta|x) &\longrightarrow \sup_\vartheta \\ \log l(\vartheta|x) = \log\left(\frac{1}{\sqrt{(2\pi)^n}\sigma^n}\right) - \frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2} &\longrightarrow \sup_{\mu,\sigma^2} \\ \alpha - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2} &\longrightarrow \sup_{\mu,\sigma^2} \end{aligned}$$

Es gilt:

$$\sum_{i=1}^n (x_i - \mu)^2 \to \inf_\mu \quad \Longleftrightarrow \quad \mu = \overline{x}_\cdot.$$

Also noch zu bilden:

$$-\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^n \frac{(x_i - \overline{x}_\cdot)^2}{\sigma^2} \longrightarrow \sup_{\sigma^2}$$

Setze $h(\tau) := -\frac{n}{2}\log\tau - \frac{1}{2}\frac{c}{\tau} \to \sup_\tau$ und bilde $h'(\tau) = 0$. Dies gilt für $\tau = \frac{c}{n}$. Also ist

$$\big(\hat{\mu}(x), \hat{\sigma}^2(x)\big) = \Big(\overline{x}_\cdot, \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x}_\cdot)^2\Big)$$

ML-Schätzer für $(\mu, \sigma^2)$.

(d) $X_1, \ldots, X_n \sim \mathrm{Po}(\lambda)$ stochastisch unabhängig, $\lambda > 0$

$$\begin{aligned} \log l(\lambda|x) &= \log\Big(\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\Big) = -n\lambda + \Big(\sum_{i=1}^{n} x_i\Big)\log\lambda - \sum_{i=1}^{n}\log(x_i!) \longrightarrow \sup_{\lambda} \\ \frac{d}{d\lambda}\log l(\lambda|x) &= -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i \stackrel{!}{=} 0 \quad \Rightarrow \hat{\lambda}(x) = \overline{x}. \end{aligned}$$

# A Diskrete Verteilungen

Angegeben werden jeweils die Zähldichten $p(\omega)$ für $\omega \in \tilde{\Omega}$, wobei $\tilde{\Omega}$ höchtens abzählbar ist und $\Omega_0 \subseteq \tilde{\Omega} \subseteq \Omega$ erfüllt, also nicht notwendig gleich dem Träger $\Omega_0$ der Verteilung ist.

1. **Dirac- oder 1-Punkt-Verteilung $\delta_\omega$:**
$$\tilde{\Omega} = \{\omega\}, \quad p(\omega) = 1.$$

2. **Gleich-Verteilung $\mathrm{G}(\tilde{\Omega})$:**
$$0 < |\tilde{\Omega}| < |\mathbb{N}|, \quad p(\omega) = \frac{1}{|\tilde{\Omega}|} \quad \text{für } \omega \in \tilde{\Omega}.$$

3. **Binominal-Verteilung $\mathrm{B}(n,p)$:**
$n \in \mathbb{N}, \quad p \in [0,1], \quad \tilde{\Omega} = \{0,1,\ldots,n\}$. Dann ist
$$p(i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad \text{für } i = 0,1,\ldots,n.$$

4. **Multinominal-Verteilung $\mathrm{M}(n;p_1,\ldots,p_k)$:**
$n,k \in \mathbb{N},\ k \geq 2,\ p_j \in [0,1]$ mit $\sum\limits_{j=1}^{k} p_j = 1$ und $\tilde{\Omega} = \{0,1,\ldots,n\}^k$.
$$p(i_1,\ldots,i_k) = \begin{cases} \binom{n}{i_1 \cdots i_k} \prod\limits_{j=1}^{k} p_j^{i_j} & \text{falls } \sum\limits_{j=1}^{k} i_j = n \\ 0 & \text{sonst} \end{cases}$$
mit dem Multinomialkoeffizienten $\binom{n}{i_1 \cdots i_k} := \dfrac{n!}{\prod\limits_{j=1}^{k} i_j!}$.

5. **Hypergeometrische Verteilung $\mathrm{H}(n,m,k)$:**
$m \leq n,\ k \leq n,\ \tilde{\Omega} = \{0,1,\ldots,n\}$.
$$p(i) = \begin{cases} \dfrac{\binom{m}{i}\binom{n-m}{k-i}}{\binom{n}{k}} & \text{falls } \max\{0, k+m-n\} \leq i \leq \min\{m,k\} \\ 0 & \text{sonst} \end{cases}$$

6. **Polya-Verteilung:**
$b,c,n \in \mathbb{N},\ s \in \mathbb{Z},\ \min\{b,c\} + (n-1)\cdot s \geq 0,\ \tilde{\Omega} = \{0,\ldots,n\}$.
$$p(t) = \binom{n}{t} \prod_{i=1}^{t-1}(b+is) \prod_{j=0}^{n-t-1}(c+js) \Big( \prod_{\nu=0}^{n-1}(b+c+\nu s) \Big) \qquad \text{für } t = 0,\ldots,n.$$

7. **Geometrische Verteilung $\mathrm{NB}(1,p)$:**
$$p \in ]0,1],\ \tilde{\Omega} = \mathbb{N}_0, \quad p(i) = p(1-p)^i \quad \text{für } i \in \mathbb{N}_0.$$

8. **Negative Binominal-Verteilung $\mathrm{NB}(m,p)$:**
$$m \in \mathbb{N},\ p \in ]0,1],\ \tilde{\Omega} = \mathbb{N}_0, \quad p(i) = \binom{m-1+i}{m-1} p^m (1-p)^i \quad \text{für } i \in \mathbb{N}_0.$$

9. **Poisson-Verteilung $\mathrm{Po}(\lambda)$:**
$$\lambda \geq 0,\ \tilde{\Omega} = \mathbb{N}_0, \quad p(i) = e^{-\lambda}\frac{\lambda^i}{i!} \quad \text{für } i \in \mathbb{N}_0.$$

# B Absolut stetige Verteilungen

Angegeben werden jeweils die Lebesgue-Dichten $p(x)$ für $x \in \Omega_0 \subseteq \Omega$, wobei $\Omega_0$ ein Träger der jeweiligen Verteilung ist.

1. **(Stetige) GleichVerteilung** $\mathrm{G}(\langle a,b\rangle_{(n)})$**:**
   $\Omega_0 = [a,b]_{(n)},\ \langle a,b\rangle_{(n)} \in \left\{ ]a,b]_{(n)}, ]a,b[_{(n)}, [a,b[_{(n)}, [a,b]_{(n)} \right\},\ a,b \in \mathbb{R}^n,\ a<b,\ n \in \mathbb{N}.$

$$p(x) = \prod_{i=1}^{n} \frac{1}{b_i - a_i}.$$

2. **Exponential-Verteilung** $\mathrm{EXP}(\lambda)$**:**

$$\Omega_0 = \mathbb{R}_{(+)},\ \lambda > 0, \quad p(x) = \lambda \cdot \exp(-\lambda x).$$

3. **Doppelte Exponential-Verteilung** $\mathrm{DEXP}(\lambda)$**:**

$$\Omega_0 = \mathbb{R},\ \lambda > 0, \quad p(x) = \frac{\lambda}{2} \cdot \exp(-\lambda \cdot |x|).$$

4. **Weibull-Verteilung** $\mathrm{WEI}(\sigma,\beta,\mu)$**:**

$$\Omega_0 = ]\mu,\infty[,\ \sigma,\beta > 0,\ \mu \in \mathbb{R}, \quad p(x) = \frac{\beta}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{\beta-1} \exp\left(\left[-\left(\frac{x-\mu}{\sigma}\right)\right]^{\beta}\right).$$

   Beachte: $\mathrm{WEI}(\sigma,1,0) = \mathrm{EXP}\left(\frac{1}{\sigma}\right)$.

5. **Cauchy-Verteilung** $\mathrm{CAU}(\sigma,\mu)$**:**

$$\Omega_0 = \mathbb{R},\ \sigma > 0,\ \mu \in \mathbb{R}, \quad p(x) = \frac{\sigma}{\pi}\frac{1}{\sigma^2 + (x-\mu)^2}.$$

6. **Logistische Verteilung** $\mathrm{LOG}(\sigma,\mu)$**:**

$$\Omega_0 = \mathbb{R},\ \sigma > 0,\ \mu \in \mathbb{R}, \quad p(x) = \frac{\exp\left(-\frac{x-\mu}{\sigma}\right)}{\sigma\left(1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right)^2}.$$

7. **Gamma-Verteilung** $\Gamma(\kappa,\sigma,\mu)$**:**

$$\Omega_0 = ]\mu,\infty[,\ \kappa,\sigma > 0,\ \mu \in \mathbb{R}, \quad p(x) = \frac{1}{\sigma^{\kappa}\Gamma(\kappa)}(x-\mu)^{\kappa-1}\exp\left(-\frac{x-\mu}{\sigma}\right).$$

   Beachte: $\Gamma(1,\sigma,0) = \mathrm{EXP}\left(\frac{1}{\sigma}\right)$.

8. **(Zentrale)** $\chi^2_n$**-Verteilung** $\chi^2_n$**:**

$$\Omega_0 = \mathbb{R}_+,\ n \in \mathbb{N}, \quad p(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}\exp\left(-\frac{x}{2}\right).$$

   Beachte: $\chi^2_n = \Gamma\left(\frac{n}{2},2,0\right),\ \chi^2_2 = \Gamma(1,2,0) = \mathrm{EXP}\left(\frac{1}{2}\right)$.

9. **(Zentrale)** $\mathrm{t}_n$**-Verteilung** $\mathrm{t}_n$**:**

$$\Omega_0 = \mathbb{R},\ n \in \mathbb{N}, \quad p(x) = \frac{1}{\sqrt{\pi n}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

   Beachte: $\mathrm{t}_1 = \mathrm{CAU}(1,0)$.

10. **(Zentrale) $\mathrm{F}_{n,m}$-Verteilung $\mathrm{F}_{n,m}$:**

$$\Omega_0 = \mathbb{R}_+,\ n, m \in \mathbb{N}, \quad p(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}}{\left(1+\frac{n}{m}x\right)^{\frac{n+m}{2}}}.$$

11. **Beta-Verteilung** $\mathrm{Beta}(p,q)$**:**

$$\Omega_0 = ]0,1[,\ p,q>0, \quad p(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}.$$

Beachte: $\mathrm{Beta}(1,1) = \mathrm{G}\left(\langle 0,1\rangle_{(1)}\right)$.

12. $n$**-dimensionale Normalverteilung** $\mathrm{N}(\mu, V)$**:**
$\Omega_0 = \mathbb{R}^n,\ \mu \in \mathbb{R}^n,\ V \in \mathbb{R}_+^{n\times n}$, d.h. $V$ positiv definit ( $\iff V' = V,\ V > 0$).

$$p(x) = \frac{1}{\sqrt{(2\pi)^n \det V}} \exp\left(-\frac{1}{2}(x-\mu)'V^{-1}(x-\mu)\right).$$

13. **Lognormale Verteilung** $\mathrm{LOGN}(\mu,\sigma)$**:**

$$\Omega_0 = \mathbb{R}_+,\ \mu \in \mathbb{R},\ \sigma > 0, \quad p(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right).$$

Beachte: $X \sim \mathrm{LOGN}(\mu,\sigma) \iff \log X \sim N(\mu,\sigma^2)$.

# C Mengen und Indikatoren

Es seien $\Omega$ eine beliebige nicht-leere Menge und $T$ eine beliebige nicht-leere Indexmenge. $A, B, C$ (mit oder ohne Index) seien Teilmengen von $\Omega$.

**Rechenregeln für Mengen:**

(a) $\Big(\bigcup_{\tau\in T} A_\tau\Big)^c = \bigcap_{\tau\in T} A_\tau$ und $\Big(\bigcap_{\tau\in T} A_\tau\Big)^c = \bigcup_{\tau\in T} A_\tau$.

(b) $\Big(\bigcap_{\tau\in T} A_\tau\Big) \cup B = \bigcap_{\tau\in T} (A_\tau \cup B)$ sowie $\Big(\bigcup_{\tau\in T} A_\tau\Big)B = \bigcup_{\tau\in T} (A_\tau B)$. Wir setzen $AB := A \cap B$.

(c) $\bigcup_{\tau\in T} (A_\tau \cup B_\tau) = \Big(\bigcup_{\tau\in T} A_\tau\Big) \cup \Big(\bigcup_{\tau\in T} B_\tau\Big)$ und $\bigcup_{\tau\in T} A_\tau B_\tau \subseteq \Big(\bigcup_{\tau\in T} A_\tau\Big)\Big(\bigcup_{\sigma\in T} B_\sigma\Big)$.

(d) $\sum_{\tau\in T} (A_\tau + B_\tau) = \sum_{\tau\in T} A_\tau + \sum_{\tau\in T} B_\tau$. $A + B$ ist genau dann definiert, wenn $A$ und $B$ disjunkt sind. In diesem Falle gilt $A + B := A \cup B$. Entsprechend ist $\sum_{\tau\in T} A_\tau$ definiert.

(e) $\bigcup_{i=1}^{\infty} A_i = A_1 + A_2 A_1^c + \ldots + A_n A_{n-1}^c \cdots A_1^c + \ldots = \sum_{j=1}^{\infty} A_j \bigcap_{k=1}^{j-1} A_k^c$. Es gelten die Konventionen

$$\bigcap_{k\in\emptyset} A_k = \Omega \text{ und } \bigcup_{k\in\emptyset} A_k = \emptyset.$$

(f) $\bigcup_{i=1}^{n} \bigcap_{j=1}^{m} A_{ij} \subseteq \bigcap_{j=1}^{m} \bigcup_{i=1}^{n} A_{ij}$. Gleichheit gilt i.a. nicht.

(g) $A \cup B = A + A^c B = A\Delta B + AB$. Dabei ist $A\Delta B := AB^c + A^c B$.

(h) $C(A\Delta B) = (CA)\Delta(CB)$.

(i) $\Big(\bigcup_{\tau\in T} A_\tau\Big)\Delta\Big(\bigcup_{\tau\in T} B_\tau\Big) \subseteq \bigcup_{\tau\in T} (A_\tau \Delta B_\tau)$.

**Rechenregeln für Indikatorfunktionen:**
Bei festem $A \subseteq \Omega$ heißt $I_A : \Omega \to \{0, 1\} \subseteq \mathbb{R}$ mit

$$I_A(\omega) := \begin{cases} 1 & \omega \in A \\ 0 & \omega \in A^c \end{cases}$$

Indikatorfunktion (kurz: Indikator) von $A$. Sämtliche mengenalgebraischen Relationen lassen sich übertragen auf Eigenschaften von Indikatorfunktionen. Es gilt

(a) $I_\Omega = 1,\ I_\emptyset = 0$,

(b) $I_{A^c} = 1 - I_A$,

(c) $I_{A\cap B} = I_{AB} = I_A \cdot I_B = \min\{I_A, I_B\}$,

(d) $I_{A\cup B} = \max\{I_A, I_B\}$ und $I_{A+B} = I_A + I_B$,

(e) $I_{A-B} = I_A - I_B$. Dabei ist $A - B$ genau dann definiert, wenn gilt $B \subseteq A$. In diesem Falle ist $A - B := A \cap B^c$.

(f) $I_{A\Delta B} = |I_A - I_B|$,

(g) $I_{\bigcup_{n=1}^{\infty} A_n} = \sup I_{A_n}$,

(h) $I_{\bigcap_{n=1}^{\infty} A_n} = \inf I_{A_n}$,

(i) $I_{\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k} = \limsup I_{A_n}$,

(j) $I_{\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k} = \liminf I_{A_n}$.

# Index


asymptotisches Modell, 95

Bayes'sche Formel, 20
van Beek, 93
Beispiel
  Beta-Verteilung, 66
  Cauchy-Verteilung, 52
  $\chi^2_1$-Verteilung, 62
  F-Verteilung, 65
  Faltung von Poisson-/Normalverteilung, 62
  $\Gamma$-Verteilung, 62
  Normalverteilung, 52
  Normalverteilung (2-dimensional), 55
  Poisson-Verteilung, 53
  t-Verteilung, 64
  Taxiproblem, 96
Binominalverteilung, 23
Bonterroni-Ungleichung, 11
Borel-Cantelli, 71
Borel-$\sigma$-Algebra, 8

Cauchy-Ungleichung, 56

Definition
  1-parametrige Exponentialfamilie, 110
  $\mathcal{A}, \mathcal{B}$-meßbar, 27
  Äquivalenz von Maßen, 43
  bedingte, 18
  Bereichsschätzer, 118
  Dichtequotient, 101
    monotoner, 108
  diskreter Wahrscheinlichkeitsraum, 24
  Dynkin-System, 69
  Erwartungswert, 51
  Faltung, 61
  Konfidenzbereich, 118
  Kovarianz, 55
  Maß, 7
  Momente (absolute/zentrale), 54
  $\mu$-fast sicher, 41
  $\mu$-Integral, 37
  $\mu$-integrierbar, 40
  $\mu$-quasiintegrierbar, 40
  $\mu$-stetig, totalstetig bzgl. $\mu$, 43
  $P$-fast sichere Konvergenz, 74
  $P$-stochastische Konvergenz, 74
  Permutation/Kombination, 15
  Produkt von Maßräumen, 31
  Produkt-$\sigma$-Algebra, 31
  Radon-Nikodym-Dichte, 44
  schwache Konvergenz, 84
  schwaches Gesetz der großen Zahlen (SGGZ), 74
  $\sigma$-Algebra, 6
    der terminalen Ereignisse, 69
  Standardabweichung, 53
  starkes Gesetz der großen Zahlen (StGGZ), 74
  stochastische Unabhängigkeit, 21
  stochastische Unabhängigkeit (Familien), 29
  stochastische Unabhängigkeit von Mengensystemen, 28
  Test, 100
    gleichmäßig bester, 100
    unverfälschter, 100
  Varianz, 53
  Verteilungsfunktion, 33
  von $X$ erzeugte $\sigma$-Algebra, 29
  Zufallsvariable, 26
Dirac, 7
  -Maß, 7
diskrete Wahrscheinlichkeitsverteilungen, 24
Dualität
  von Tests und Konfidenzbereichen, 119

Ereignisse
  stochastische Unabhängigkeit von, 21
Erinnerungsprozesse, 82
Ersetzungsmethode von Lindeberg/Levy, 90

Faltungsformel, 61
Fatou, 41
Fehler
  1. Art, 96
  2. Art, 96
Feller-Bedingung, 88
Fubini, 46
Fundamentallemma
  erweitertes, 108
Fundamentallemma von Neyman und Pearson, 102

Grundannahme der Statistik, 99

Hilberth, 2
Hölder-Ungleichung, 56

Integrationstheorie, 26, 37
iterierter Logarithmus, 68

Jensen-Ungleichung, 58

Kettenregel für Maße, 44
Kolmogoroff, 4

Null-Eins-Gesetz von, 70
Kolmogorov
-Bedingung, 76
-Ungleichung, 76
Kombinatorik, 15
Konfidenzellipsoide, 120
konjugierte Exponenten, 56
Konvergenz
$P$-fast sichere, 74
$P$-stochastische, 74
schwache, 83
Konvergenz von Mengen, 12

Laplace
-Experimente, 14
Rechenregeln für, 14
Lebesgue
-Maß, 8
Satz von, 42
Lemma
erweitertes Fundamental-, 108
Fundamental- von Neyman und Pearson, 102
Kronecker-, 77
von Borel-Cantelli, 71
von Fatou, 41
Lindeberg-Bedingung, 88
Lyapunov-Bedingung, 91

Maße
diskretes Wahrscheinlichkeits-, 27
Produktwahrscheinlichkeits-, 31
Rechenregeln, 10
Stetigkeit, 12
Maximum-Likelihood-Schätzer, 122
Mengen
Konvergenz von, 12
Meßbarkeit, 28
Minkowsky-Ungleichung, 56
ML-Schätzer, 122
Momente, 54
absolute, 54
zentrale, 54
de Morgan, 9
$\mu$-Integral
Isotonie, 39
Linearität, 39
Multinominalkoeffizient, 16
Multiplikationssatz für bedingte Ereignisse, 19

Nikodym, 44
Niveau
Test zum, 100
Null-Eins-Gesetz, 68
von Kolmogoroff, 70

Paradoxon
Bertrand'sches, 3
von Chevalier de Meré, 3
Poincare
Siebformel von, 10
Poisson, 25
Poisson'scher Grenzwertsatz, 25
Portmanteau-Theorem, 86
Produktwahrscheinlichkeitsmaß, 31
Projektionsabbildung, 31

Radon, 44
Recente-Problem, 11, 20

Satz
Bayes'sche Formel, 20
Eigenschaften der Verteilungsfunktion, 33
Eigenschaften von Erwartungswerten, 51
Korrespondenz-, 119
Multiplikationssatz für Erwartungswerte, 54
Poisson'scher Grenzwert-, 25
Transformationsformel, 46
Verteilung des Produktes von stochastisch unabhängigen Zufallsvariablen, 63
von Berry-Esseen, 93
von der beschränkten Konvergenz (Lebesgue), 42
von der monotonen Konvergenz, 41
von der totalen Wahrscheinlichkeit, 20
von der totalen Zerlegung, 20
von Fubini für Maße, 46
von Glirenko-Cantelli, 82
von Lebesgue, 42
von Radon/Nikodym, 44
Zentral- der Statistik, 82
Zweireihen-, 77
Schätzer
Lücken-, 97
Median-, 96
SGGZ, 74
von Markov, 75
Siebformel
Anwendung der, 11
Siebformel von Sylvester/Poincare, 10
Signifikanzniveau, 98
Statistik
Grundannahme der, 99
statistischer Raum, 99
statistisches Experiment, 99
StGGZ, 74
von Etemadi, 76
von Kolmogorov, 76
Stochastische Unabhängigkeit, 21
von endlich vielen Ereignissen, 22
stochastische Unabhängigkeit

von Mengensystemen, 28
Sub-$\sigma$-Additivität, 10
subjektive Wahrscheinlichkeit, 4
Sylvester
Siebformel von, 10

Technik des „Stutzens“, 76
Test, 98, 100
„einfach gegen einfach“, 100
$2 \times 2$-Tafel, 116
„außen:innen“, 112
1-seitiger Binomialtest, 109
1-seitiger Gauß-Test, 106
1-seitiger t-Test, 114
exakter Test von Fischer, 116
gleichmäßig bester, 98
„innen:außen“, 112
2-Stichproben t-, 116
2-Stichproben Binomial-, 116
2-seitiger Binomialtest, 113
2-seitiger Gauß-Test, 113
Transformationsformel, 46
Transformationsformel für $\lambda^n$-Dichten, 46
Transformationsformel für $\lambda^n$-Integral, 46

Ungleichung
Bonterroni-, 11
Tschebyschev-Markov-, 59

Verteilung
Bernoulli-, 24
Beta-, 66
Binominal-, 23, 24
Cantor-, 50
$\chi^2_1$-, 62
diskrete Wahrscheinlichkeits-, 24
Exponential-, 44
F-, 65
$\Gamma$-, 62
geometrische, 24
Gleich-, 24
hypergeometrische, 25
Multinominal-, 25
negative Binominal-, 25
nicht zentrale t-, 115
Poisson-, 25
$t$-, 64
Verteilungsannahme, 99
Verteilungsfamilien, 110

Wahrscheinlichkeit
-sbegriff, 2
axiomatische, 4
bedingte, 18
Multiplikationssatz für, 19
diskrete, 24
subjektive, 4

Zentralsatz der Statistik, 82
ZGWS
multivariater, 93
Lindeberg/Levy, 93
von de Moivre-Laplace, 88
von Lindeberg, 89
von Lindeberg/Feller, 92
von Lindeberg/Levy, 91
von Lyapunov, 92
Zufallsvariable, 26
Bild/Urbild, 26
Bildmaß, 27
diskrete, 27
ganzzahlige, 27
$k$-dimensionale, 27
numerische, 27
Operationstreue der Umkehrabbildung, 26
reelle, 27
Urbildabbildung einer Komposition, 27
vektorwertige, 27