ELEMENTARE STOCHASTIK

Vorlesung gehalten von
Martin Aigner

Wintersemester 2008/9

Inhaltsverzeichnis Elementare Stochastik

Vorlesung gehalten von Martin Aigner, Wintersemester 2008/9

1 Grundbegriffe

1.1 Zufall

Stochastik ist die Lehre von den Gesetzmäßigkeiten des Zufalls. Dies scheint zunächst ein Widerspruch in sich zu sein. Trotzdem wird jeder annehmen, dass bei wiederholtem Münzwurf in etwa der Hälfte der Fälle Kopf bzw. Zahl erscheint. Dies ist genau so eine Gesetzmäßigkeit, die wir studieren wollen.

Die Stochastik besteht aus zwei Teilen. Die *Wahrscheinlichkeitstheorie* beschreibt zufällige Vorgänge mit Modellen, die *Statistik* zieht Schlußfolgerungen aus den Beobachtungen auf den realen Vorgang und das Modell.

Beispiel. Die W-Theorie modelliert das Werfen einer Münze mit einem sogenannten Bernoulli-Modell. Die Statistik schließt aus den Beobachtungen, ob die Münze fair ist.

Grundlegend für die Behandlung zufälliger Vorgänge sind folgende Bedingungen:

A. Das Zufallsexperiment ist beliebig oft wiederholbar.

B. Es gibt keine äußeren Informationen über den Ausgang.

C. Die Erfahrung über bisherige Ausgänge nützt nichts.

1.2 Wahrscheinlichkeitsräume

Die Stochastik, wie wir sie heute betreiben, beginnt mit den Axiomen, die Kolmogorov 1933 postuliert hat.

Zunächst ist eine Menge Ω von *Elementarereignissen* gegeben. Zum Beispiel ist $\Omega = \{K(\text{opf}), Z(\text{ahl})\}$ beim Münzwurf oder $\Omega = \{1, 2, \dots, 6\}$ beim Würfeln. Jede Teilmenge $A \subseteq \Omega$ heißt ein *Ereignis*. Zum Beispiel $A = \{\text{mindestens } k \text{ Mal Kopf bei } n \text{ Münzwürfen}\}$ oder $A = \{\text{gerade Zahl}\}$ beim Würfeln.

Definition. $\mathcal{E} \subseteq 2^\Omega$ ist eine *σ-Algebra über* Ω, falls gilt:

1. $\varnothing \in \mathcal{E}$,

2. $A \in \mathcal{E} \Rightarrow A^c = \Omega \setminus A \in \mathcal{E}$,

3. $A_1, A_2, A_3, \ldots \in \mathcal{E} \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{E}$.

Beispiel. $\mathcal{E} = \{\varnothing, \Omega\}$, $\mathcal{E} = 2^\Omega$. Sind $\mathcal{F}_i$ σ-Algebren über Ω, dann auch $\bigcap \mathcal{F}_i$. Ist also $\mathcal{E}_0 \subseteq \Omega$. beliebig, so gibt es eine kleinste σ-Algebra $[\mathcal{E}_0]$, die $\mathcal{E}_0$ enthält, nämlich $[\mathcal{E}_0] = \bigcap \mathcal{F}$, $\mathcal{E}_0 \subseteq \mathcal{F}$, $\mathcal{F}$ σ-Algebra. $[\mathcal{E}_0]$ heißt die von $\mathcal{E}_0$ *erzeugte* σ-Algebra.

Beispiel. Die *Borelmengen* in $\mathbb{R}$ (oder $\mathbb{R}^n$) sind die von den offenen Intervallen erzeugte σ-Algebra.

Wir kennen die Gesetze von de Morgan:

$$(\bigcup A_i)^c = \bigcap A_i^c, \quad (\bigcap A_i)^c = \bigcup A_i^c\,.$$

Aus der Definition können wir sofort ein paar Folgerungen ziehen.

1. $\mathcal{E}$ enthält alle endlichen Vereinigungen und Durchschnitte, denn mit $A_1, \ldots, A_n \in \mathcal{E}$ ist

$$A_i \cup \ldots \cup A_n \cup \varnothing \cup \varnothing \cup \ldots = A_1 \cup \ldots \cup A_n \in \mathcal{E}\,.$$

Ferner ist mit $A_i \in \mathcal{E}$ auch $A_i^c \in \mathcal{E}$, also

$$(A_1 \cap \ldots \cap A_n)^c = A_1^c \cup \ldots \cup A_n^c \in \mathcal{E}\,,$$

und somit $A_1 \cap \ldots \cap A_n \in \mathcal{E}$.

2. Mit $A, B \in \mathcal{E}$ ist auch $A \setminus B \in \mathcal{E}$. Es gilt

$$A \setminus B = A \cap B^c \in \mathcal{E}\,.$$

3. Alle abzählbaren Vereinigungen und Durchschnitte von $\mathcal{E}$-Mengen A_i sind in $\mathcal{E}$.

Für die Vereinigung ist dies Axiom 3), und für die Durchschnitte verwenden wir de Morgans Gesetz.

Definition. Ein *Wahrscheinlichkeitsraum* ist ein Tripel $(\Omega, \mathcal{E}, P)$, wobei

1. $\mathcal{E} \subseteq 2^\Omega$ ist σ-Algebra,
2. $P : \mathcal{E} \to [0, 1]$ heißt das *Wahrscheinlichkeitsmaß* und erfüllt

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

ii) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, falls $A_1, A_2, \ldots$ paarweise disjunkte Mengen in $\mathcal{E}$ sind.

Bemerkung. Falls Ω endlich oder abzählbar ist, so nehmen wir immer die σ-Algebra $\mathcal{E} = 2^{\Omega}$. Für überabzählbare Mengen Ω (wie z.B. die Menge der Ausgänge beim unendlich oft wiederholten Münzwurf) funktioniert 2^{Ω} nicht. Das heißt, es gibt kein geeignetes W-Maß. (Beweis siehe Georgii).

Wir ziehen wiederum ein paar unmittelbare Folgerungen.

1. $P(\varnothing) = 0$.

Es ist $P(\varnothing) = P(\varnothing \cup \varnothing \cup \varnothing \cup \ldots) = \sum P(\varnothing)$, und dies ist nur für $P(\varnothing) = 0$ erfüllt.

2. $P(A) + P(A^c) = 1$.

Es ist $\Omega = A \cup A^c \cup \varnothing \cup \varnothing \cup \ldots$, also $1 = P(\Omega) = P(A) + P(A^c)$.

3. $A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$.

Das folgt aus $A \cup B = A \cup B \cup \varnothing \cup \varnothing \ldots$

4. $A \subseteq B \Rightarrow P(B) = P(A) + P(B \setminus A) \geq P(A)$.

Das W-Maß ist also monoton steigend.

5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Wir haben $(A \cup B) \setminus A = B \setminus (A \cap B)$, also nach 4)

$$P(A \cup B) - P(A) = P(B) - P(A \cap B)\,.$$

6. $P(\bigcup_{n \geq 1} A_n) \leq \sum_{n \geq 1} P(A_n)$.

Wir stellen $\bigcup_{n \geq 1} A_n$ als Vereinigung disjunkter Mengen dar:

$$\bigcup_{n \geq 1} A_n = \bigcup_{n \geq 1} (A_n \setminus \bigcup_{j<n} A_j)\,,$$

und erhalten mit 4)

$$P(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} P(A_n \setminus \bigcup_{j<n} A_j) \leq \sum_{n\geq 1} P(A_n)\,.$$

Beispiel. 1. Werfen einer Münze. $\Omega = \{0,1\}$, $1 =$"Kopf", $0 =$"Zahl", $P(1) = p$, $P(0) = 1-p$, $0 \leq p \leq 1$.

2. Würfeln. $\Omega = \{1,2,\ldots,6\}$, $P(k) = \frac{1}{6}$.

3. $\Omega = [0,1] \subseteq \mathbb{R}$, $P([a,b]) = b-a$.

Wir werden öfters Produkte von W-Räumen betrachten. Seien $\mathcal{E}_i$ σ-Algebren auf Ω_i, $i = 1,\ldots,n$, dann ist die Produktalgebra $\mathcal{E}^{\otimes}$ auf $\Omega_1 \times \ldots \times \Omega_n$ die von $\mathcal{E}_1 \times \ldots \times \mathcal{E}_n$ erzeugte σ-Algebra.

Die ersten interessanten Ergebnisse sind die folgenden.

Lemma 1.1. a) *Sei $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ eine aufsteigende Kette von Ereignissen, dann gilt*

$$P(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n)\,.$$

b) *Ist $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$ eine absteigende Kette, so gilt*

$$P(\bigcap_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n)\,.$$

Beweis. Wir haben

$$A_n = A_1 \cup (A_2 \setminus A_1) \cup \ldots \cup (A_n \setminus A_{n-1})\,,$$

also

$$P(A_n) = P(A_1) + P(A_2 \setminus A_1) + \ldots + P(A_n \setminus A_{n-1})\,,$$

und es folgt mit $A_0 = \varnothing$

$$\begin{aligned} P(\bigcup_{n\geq 1} A_n) &= P(\bigcup_{n\geq 1}(A_n \setminus A_{n-1})) = \sum_{n\geq 1} P(A_n \setminus A_{n-1}) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i \setminus A_{i-1}) = \lim_{n\to\infty} P(A_n)\,. \end{aligned}$$

Aussage b) folgt mit de Morgan. $\square$

Folgerung 1.2. a. *Ist* $A_1 \subseteq A_2 \subseteq \ldots$ *mit* $P(A_n) = 0$ *für alle* n*, so gilt* $P(\bigcup A_n) = 0$.

b. *Ist* $A_1 \supseteq A_2 \supseteq \ldots$ *mit* $P(A_n) = 1$ *für alle* n*, so gilt* $P(\bigcap_{n\geq 1} A_n) = 1$.

Wir kennen $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Die Verallgemeinerung auf n Mengen ist das *Prinzip der Inklusion - Exklusion*.

Lemma 1.3. *Es gilt*

$$P(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\leq i_1<\cdots<i_k\leq n} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Beweis. Für $n = 1, 2$ ist das richtig, nun verwenden wir Induktion:

$$\begin{aligned} P(A_1 \cup \ldots \cup A_n) &= P((A_1 \cup \ldots \cup A_{n-1}) \cup A_n) \\ &= P(A_1 \cup \ldots \cup A_{n-1}) + P(A_n) - P((A_1 \cup \ldots \cup A_{n-1}) \cap A_n) \\ &= P(A_1 \cup \ldots \cup A_{n-1}) + P(A_n) - P(\bigcup_{i=1}^{n-1}(A_i \cap A_n)). \end{aligned}$$

Der erste Summand ergibt alle Summen mit $1 \leq i_1 < \ldots < i_k \leq n-1$ nach Induktion, der zweite die Summe mit $k = 1$, $i_1 = n$, und für den dritten gilt wieder nach Induktion

$$P(\bigcup_{i=1}^{n-1}(A_i \cap A_n)) = \sum_{k=2}^{n-1} (-1)^{k-2} \sum_{1\leq i_1<\cdots<i_{k-1}\leq n-1} P(A_{i_1} \cap \ldots \cap A_{i_{k-1}} \cap A_n),$$

und mit dem Minuszeichen folgt die Formel. $\square$

Schreiben wir Ω für den leeren Durchschnitt, so können wir die Formel auch auf folgende Form bringen:

$$P(\Omega \setminus (A_1 \cup \ldots \cup A_n) = \sum_{k=0}^{n} (-1)^{k} \sum_{1\leq i_1<\cdots<i_k\leq n} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

1.3 Diskrete Wahrscheinlichkeitsräume

Wenn Ω endlich oder abzählbar ist, so heißt $(\Omega, 2^{\Omega}, P)$ ein *diskreter* Wahrscheinlichkeitsraum. Da $\sum_{\omega\in A} P(\omega) \leq 1$ ist für jedes $A \subseteq \Omega$, so konvergiert

$\sum_{\omega \in A} P(\omega) = P(A)$. Ein disjunkter W-Raum ist also durch Angabe von $P(\omega)$, $\omega \in \Omega$, vollkommen determiniert, wobei $\sum_{\omega \in \Omega} P(\omega) = 1$ ist.

Beispiele.
1. **Gleichverteilung.** Ω endlich, $P(\omega) = \frac{1}{|\Omega|}$, also $P(A) = \frac{|A|}{|\Omega|}$. Man spricht von einem *Laplace Modell*, und interpretiert $\frac{|A|}{|\Omega|}$ als "günstige" durch "mögliche" Fälle. Als Beispiele haben wir eine faire Münze, Würfel oder Lotto.

2. **Binomialverteilung.** Wir führen ein Experiment mit den Ausgängen 1 und 0 n Mal aus, wobei die W-keit von 1 gleich p ist, von 0 gleich $1-p$. Wir interpretieren 1 als "Erfolg" und 0 als "Mißerfolg" und sprechen von einem *Bernoulli Modell*. Was uns interessiert, ist die Anzahl k der Erfolge bei n Versuchen. Es ist also $\Omega = \{0, 1, 2, \ldots, n\}$, $P(k) = \binom{n}{k} p^k (1-p)^k$, und wir setzen

$$b(k, n; p) = \binom{n}{k} p^k (1-p)^{n-k} .$$

3. **Poissonverteilung.** Hier ist $\Omega = \{0, 1, 2, 3, \ldots\}$,

$$P(n) = p(n; \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}, \ \lambda > 0 .$$

Wir sprechen von einem *Poisson Modell*.

4. **Geometrische Verteilung.** Wir führen wie beim Bernoulli Modell ein Experiment beliebig oft aus, mit $p = W$-keit für 1, $q = 1 - p = W$-keit für 0. Was uns interessiert, ist, in welchem Versuch Erfolg zum ersten Mal auftritt. Also $\Omega = \mathbb{N}$,

$$P(n) = q^{n-1}(1-q), \quad 0 \leq q < 1 .$$

Als Beispiel sei A_0 das Ereignis, dass der erste Erfolg in einem Versuch mit gerader Nummer auftritt. Dann ist

$$\begin{aligned} P(A_0) &= q(1-q) + q^3(1-q) + \ldots = q(1-q) \sum_{i \geq 0} q^{2i} \\ &= q(1-q) \frac{1}{1-q^2} = \frac{q}{1+q} . \end{aligned}$$

Für das komplementäre Ereignis $A_1 = \mathbb{N} \setminus A_0$ haben wir daher

$$P(A_1) = 1 - \frac{q}{1+q} = \frac{1}{1+q} > \frac{q}{1+q} = P(A_0) .$$

5. **Urnenmodelle.** Dies sind die klassischen Modelle. Gegeben seien N Kugeln in einer Urne, von denen n gezogen werden. Wir numerieren die Kugeln $\{1, 2, \ldots, N\}$. Eine Ziehung ist also ein Wort $a_1 a_2 \ldots a_n$, $a_i \in \{1, 2, \ldots, N\}$, und Ω sei die Menge der Ziehungen. Wir haben vier Klassen:

(A) Mit Zurücklegen:

- (A1) geordnet, das heißt die Reihenfolge spielt eine Rolle. Dann ist $|\Omega| = N^n$.
- (A2) ungeordnet, die Reihenfolge, in der die Kugeln gezogen werden, ist belanglos. Dann ist $|\Omega| = \binom{N+n-1}{n} = \frac{N(N+1)\ldots(N+n-1)}{n!}$.

(B) Ohne Zurücklegen:

- (B1) geordnet, $|\Omega| = N(N-1)\cdots(N-n+1)$,
- (B2) ungeordnet, $|\Omega| = \binom{N}{n} = \frac{N(N-1)\cdots(N-n+1)}{n!}$.

Beweis von (A2). Jede Ziehung entspricht einem Wort $1 \leq a_1 \leq a_2 \leq \ldots \leq a_n \leq N$. Zu $a_1 a_2 \ldots a_n$ assoziieren wir das Wort $1 \leq a_1 < a_2 + 1 < a_3 + 2 < \ldots < a_n + n - 1 \leq N + n - 1$. Dies ergibt eine Bijektion auf die Menge aller n-Untermengen von $\{1, 2, \ldots, N + n - 1\}$, also ist $|\Omega| = \binom{N+n-1}{n}$.

Beispiel. Geburtstagsparadox. Bei einer Party sind n Personen. Was ist die W-keit, dass keine zwei Personen am selben Tag Geburtstag haben? Wir schließen Schaltjahre aus, also ist $N = 365$. Die Geburtstage entsprechen dann einer Ziehung nach Modell (B1). Unter der Voraussetzung der Gleichverteilung aller Tage haben wir also für die W-keit "keine zwei Personen haben am selben Tag Geburtstag"

$$\frac{365 \cdot 364 \cdot \cdots \cdot (365 - n + 1)}{365^n} = 1(1 - \frac{1}{365})(1 - \frac{2}{365}) \cdots (1 - \frac{n-1}{365})$$

und das ist $< \frac{1}{2}$ für $n \geq 23$.

Von 1936 - 1978 gab es 24 Fieldsmedaillenpreisträger, und tatsächlich haben Ahlfors (1936) und Fefferman (1978) beide am 18. April Geburtstag.

6. **Hypergeometrische Verteilung.** Gegeben sind n Kugeln, davon r rote und $n - r$ weiße. Wir ziehen m Kugeln ohne Zurücklegen, und interessieren uns für die W-keit, dass genau k rote unter den m gezogenen Kugeln sind.

Wir haben Modell (B2), die Anzahl der möglichen Ziehungen ist $\binom{n}{m}$ und die Anzahl der günstigen $\binom{r}{k}\binom{n-r}{m-k}$. Also ist die W-keit

$$h(k,m;r,n) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}\,.$$

Wir sprechen von einem *hypergeometrischen* Modell.

Beispiel. Eine übliche Lottoausspielung 6 aus 49 entspricht dieser Situation mit $n = 49$, $r = 6$ (richtige), $m = 6$ (angekreuzte Zahlen). Also ist die W-keit für k richtige

$$h(k,6;6,49) = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}\,.$$

Die W-keit für 6 richtige ist $\frac{1}{\binom{49}{6}} \sim 7 \cdot 10^{-8} \sim \frac{1}{14 \text{ Mill}}$. Die W-keit für ≥ 3 richtige ist $\sim 0{,}018$, und die W-keit für 0 richtige ist $\frac{\binom{43}{6}}{\binom{49}{6}} \sim 0{,}436$, also immer noch fast 50%.

1.4 Wahrscheinlichkeitsräume mit Dichtefunktion

Die andere wichtige Klasse von W-Räumen betrifft kontinuierliche Prozesse. Ω ist $\mathbb{R}$ oder eine "einfache" Teilmenge wie $[a,b]$, oder allgemein $\mathbb{R}^n$ bzw. einfache Teilmengen davon. Die σ-Algebra enthält alle Borelmengen in Ω.

Wir erklären das W–Maß P durch eine sogenannte *Dichtefunktion* $f(x)$. Dabei wird vorausgesetzt:

1. $f(x) \geq 0$,
2. $f(x)$ ist stückweise stetig,
3. $\int\limits_{\Omega} f(x)dx = 1$.

Die entsprechenden Forderungen gelten für $\mathbb{R}^n$.

Wir erklären nun für $A \in \mathcal{E}$

$$P(A) = \int\limits_A f(x)dx\,.$$

Zum Beispiel ist $P([a,b]) = \int\limits_a^b f(x)dx\,.$

Nach den Regeln der Integrationsrechung erfüllt P die Forderungen an ein W-Maß. Wir können $f(x)$ auf ganz $\mathbb{R}$ (oder $\mathbb{R}^n$) erklären, indem wir $f(x) = 0$ für $x \notin \Omega$ setzen. Dies wird in den uns interessierenden Fällen keine Schwierigkeiten bereiten.

Die *Verteilungsfunktion* ist

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

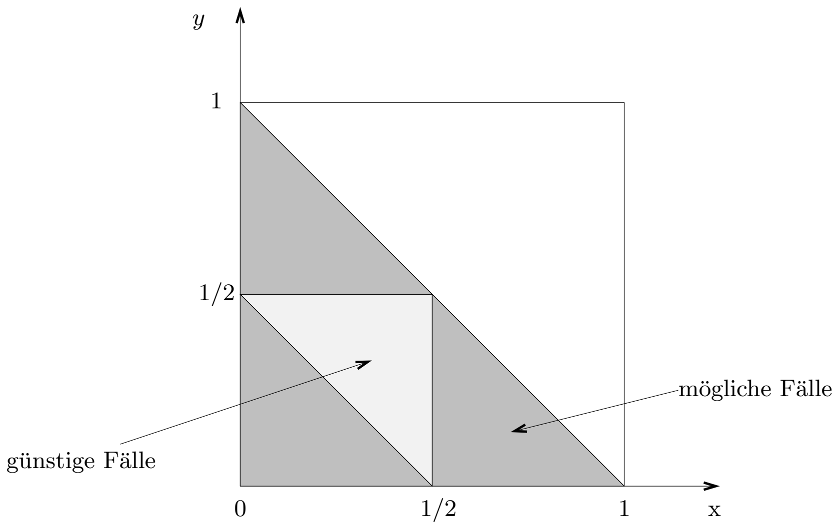
$F(x)$ ist eine monoton wachsende differenzierbare Funktion, und es gilt

$$F'(x) = f(x)\,.$$

Beispiele.

1. Gleichverteilung. Es sei $\Omega = [a, b]$, $f(x) = \frac{1}{b-a}$ für alle $x \in \Omega$. Offenbar ist $\int\limits_a^b f(x)dx = 1$. Die Verteilungsfunktion ist $F(x) = \int\limits_a^x \frac{1}{b-a}dt = \frac{x-a}{b-a}$. Als Beispiel haben wir $P([\alpha, \beta]) = \frac{\beta-\alpha}{b-a}$ für $[\alpha, \beta] \subseteq [a, b]$. Für allgemeines Ω setzen wir $I = \int\limits_{\Omega} 1dx \in (0, \infty)$ und erklären die Gleichverteilung durch $f(x) = \frac{1}{I}$ für alle x.

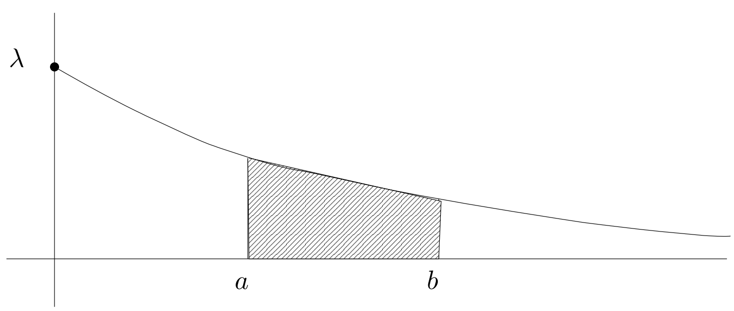
Beispiel. Der Weihnachtsmann bricht seinen Stab in drei Teile. Was ist die W-keit p, dass die drei Teile ein Dreieck bilden? Wir nehmen an, der Stab hat Länge 1, und die Teile können auch Länge 0 haben (die W-keit dafür ist natürlich 0). Seien x, y zwei Längen, dann ist $\Omega = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}$. Aufgrund der Dreiecksungleichung können wir genau dann ein Dreieck bilden, wenn $x + y \geq \frac{1}{2}$, $x \leq \frac{1}{2}$, $y \leq \frac{1}{2}$ gilt. Damit haben wir folgende Situation



Unter der Annahme der Gleichverteilung ist somit $p = \frac{1}{4}$.

2. Exponentialverteilung. Hier ist $\Omega = [0, \infty)$, $\lambda > 0$, und $f(x) = \lambda e^{-\lambda x}$. Wir haben $\int\limits_0^\infty \lambda e^{-\lambda x} dx = -e^{\lambda x}\Big|_0^\infty = 1$, also ist $f(x)$ Dichtefunktion. Die Verteilungsfunktion ist

$$F(x) = \int\limits_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}\,.$$



$P([a, b])$

Die Exponentialverteilung ist wichtig für Wartezeitenmodelle, die wir später behandeln werden.

3. Normalverteilung. Hier ist $\Omega = \mathbb{R}$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. Es ist nicht unmittelbar einsichtig, dass $\int\limits_{-\infty}^{\infty} f(x)dx = 1$ ist. Zum Beweis verwenden wir den folgenden Trick.

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy = \int\limits_{-\infty}^{\infty} e^{-\frac{y^2}{2}} \left(\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) dy = \left(\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right)^2 .$$

Transformation auf Polarkoordinaten $x = r\cos\vartheta$, $y = r\sin\vartheta$, $dxdy = rdrd\vartheta$ ergibt

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy = \int\limits_{0}^{2\pi} \left(\int\limits_{0}^{\infty} re^{-\frac{r^2}{2}} dr \right) d\theta = \int\limits_{0}^{2\pi} \left(-e^{-\frac{r^2}{2}} \Big|_0^{\infty} \right) d\vartheta$$

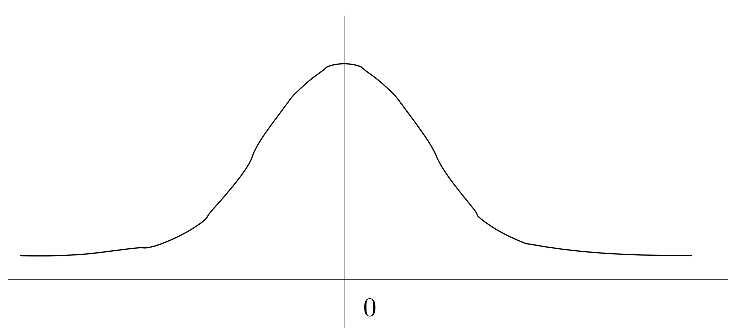
$$= \int\limits_{0}^{2\pi} d\vartheta = 2\pi ,$$

und es folgt $\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$.

Wir sprechen von der *Standard Normalverteilung* $N(0,1)$, und bezeichnen die Verteilungsfunktion mit

$$\phi(x) = \int\limits_{\infty}^{x} f(t)dt .$$

Die Dichte ist die berühmte *Gaußsche Glockenkurve.*



Sie wird uns in den fundamentalen Grenzwertsätzen wieder begegnen.

Die Normalverteilung $N(\mu, \sigma^2)$ ist erklärt durch die Dichte

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

4. Gammaverteilung. Die Funktion

$$\Gamma(r) = \int\limits_0^\infty x^{r-1} e^{-x} dx \quad (r \geq 0)$$

heißt die Γ-*Funktion*. Sie ist eine der fundamentalen Funktionen der Analysis. Wir stellen einige Eigenschaften zusammen:

1. $\Gamma(1) = 1$. Dies folgt durch Integration $\int\limits_0^\infty e^{-x} dx = -e^{-x}\Big|_0^\infty = 1$.

2. $\Gamma(r+1) = r\Gamma(r)$. Wir erhalten durch partielle Integration

$$\begin{aligned}\Gamma(r+1) &= \int\limits_0^\infty x^r e^{-x} dx = \underbrace{x^r(-e^{-x})\Big|_0^\infty}_{=0} + r \int\limits_0^\infty x^{r-1} e^{-x} dx \\ &= r\Gamma(r) .\end{aligned}$$

Aus $\Gamma(1) = 1$ folgt insbesondere $\Gamma(n+1) = n!$ für $n \in \mathbb{N}$.

Die $\gamma_{\alpha,r}$-Verteilung ist erklärt durch die Dichtefunktion

$$f(x) = \frac{\alpha^r x^{r-1} e^{-\alpha x}}{\Gamma(r)} \text{ auf } [0, \infty), \; \alpha, r > 0 .$$

Mit der Transformation $y = \alpha x$ haben wir

$$\int\limits_0^\infty \alpha^r x^{r-1} e^{-\alpha x} dx = \alpha \int\limits_0^\infty y^{r-1} e^{-y} \frac{1}{\alpha} dy = \Gamma(r) ,$$

also ist $f(x)$ Dichtefunktion.

Der Spezialfall $r = 1$ ergibt $f(x) = \alpha e^{-\alpha x}$, also die Exponentialverteilung.

Beispiel. Was ist $\Gamma(\frac{1}{2})$? Wir haben für $r = \alpha = \frac{1}{2}$

$$\Gamma(\frac{1}{2}) = \frac{1}{\sqrt{2}} \int_0^\infty x^{-\frac{1}{2}} e^{-\frac{x}{2}} dx \,.$$

Mit der Transformation $x = y^2$ erhalten wir mit $dx = 2ydy$

$$\int_0^\infty x^{-\frac{1}{2}} e^{-\frac{x}{2}} dx = \frac{1}{2} \int_{-\infty}^\infty \frac{1}{y} e^{-\frac{y^2}{2}} 2ydy = \sqrt{2\pi}\,,$$

und somit

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}\,.$$

Die Γ-Verteilung wird in der Statistik eine große Rolle spielen.

1.5 Bedingte Wahrscheinlichkeit

Es sei ein W-Raum $(\Omega, \mathcal{E}, P)$ gegeben, $A, B \in \mathcal{E}$.

Definition. Die *bedingte W-keit von A unter der Annahme* von B ist

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

wobei $P(B) > 0$ vorausgesetzt wird.

Definition. Die Ereignisse A und B heißen *unabhängig*, falls

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

gilt oder äquivalent

$$P(A|B) = P(A), \quad P(B|A) = P(B)\,.$$

Die Gleichungen bedeuten, dass wir aus der Kenntnis von B keine zusätzliche Information über A erhalten, und umgekehrt.

Beispiel. 1. Würfeln. Sei $A = \{\text{gerade Zahl}\}$, $B = \{x \geq 3\}$. Dann ist $P(A \cap B) = \frac{1}{3}$, $P(A) = \frac{1}{2}$, $P(B) = \frac{2}{3}$, also $P(A|B) = \frac{1}{2}$, $P(B|A) = \frac{2}{3}$. Die Ereignisse A, B sind unabhängig. Wählt man $B = \{x \geq 4\}$, so ist $P(A \cap B) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, also $P(A|B) = P(B|A) = \frac{2}{3}$. Die Ereignisse sind nicht unabhängig.

2. Geometrische Verteilung. Es sei $A = \{1,2\}$, $B = \{\text{gerade}\}$. Wir haben $P(A) = (1-q) + q(1-q) = 1-q^2$, $P(B) = \frac{q}{1+q}$, $P(A \cap B) = q(1-q)$, somit $P(A|B) = 1-q^2 = P(A)$. A und B sind unabhängig.

Satz 1.4. *Es sei $\Omega = \bigcup_k B_k$ disjunkte Vereinigung, $P(B_k) > 0$ für alle k.*

a. $P(A) = \sum_k P(B_k) P(A|B_k)$.

b. *Formel von Bayes: Sei $P(A) > 0$, dann gilt*

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_k P(B_k)P(A|B_k)}\,.$$

Beweis. a. Wir haben

$$\begin{aligned} P(A) &= P(A \cap \bigcup_k B_k) = P(\bigcup_k (A \cap B_k)) = \sum_k P(A \cap B_k) \\ &= \sum_k \frac{P(A \cap B_k)P(B_k)}{P(B_k)} = \sum_k P(B_k)P(A|B_k)\,. \end{aligned}$$

b. Hier ist

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A \cap B_i)}{P(A)} \cdot \frac{P(B_i)}{P(B_i)} = \frac{P(B_i)P(A|B_i)}{P(A)}\,. \qquad \square$$

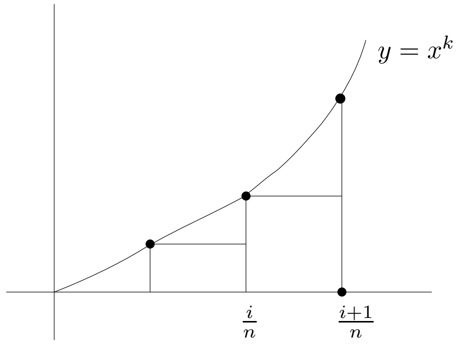
Beispiele. 1. Es sind n Urnen mit je n Kugeln gegeben. In Urne i befinden sich i rote und $n-i$ weiße.

Problem I. Eine Urne wird uniform mit W-keit $\frac{1}{n}$ gewählt. Was ist die W-keit p_k, dass bei k-maligem Ziehen mit Zurücklegen aus der gewählten Urne k Mal eine rote Kugel gezogen wird?

Problem II. Eine Urne wird mit W-keit $\frac{1}{n}$ gewählt. Es wird k Mal gezogen mit Zurücklegen und jedes Mal erscheint rot. Was ist die W-keit q_k, dass auch beim $k+1$-ten Mal rot gezogen wird?

Lösung von Problem I. Sei $A = \{k \text{ Mal rot}\}$, $B_i = \{\text{Urne } i\}$. Wir haben $P(B_i) = \frac{1}{n}$, $P(A|B_i) = (\frac{i}{n})^k$, also $P(A) = \sum_{i=1}^{n} (\frac{i}{n})^k \frac{1}{n}$. Schauen wir auf die

Kurve $y = x^k$



so sieht man $p_k = P(A) \sim \int\limits_0^1 x^k dx = \frac{1}{k+1}$ für große n.

Lösung von Problem II. Sei $B = \{\text{die ersten } k \text{ Kugeln rot}\}$, $A = \{\text{alle } k + 1 \text{ Kugeln rot}\}$. Wir haben

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

wegen $A \subseteq B$, also $q_k = P(A|B) \sim \frac{\frac{1}{k+2}}{\frac{1}{k+1}} = \frac{k+1}{k+2}$.

2. Es wird ein medizinischer Test zu einer Krankheit durchgeführt. Sei $B = \{\text{Krankheit}\}$, $A = \{\text{Test positiv}\}$. Der Patient ist interessiert an der W-keit, dass er bei einem positiven Test tatsächlich krank ist, also $P(B|A)$. Angenommen, wir haben die folgenden Daten:

$$P(A|B) = 0,95, \;\; P(A|B^c) = 0,10, \;\; P(B) = 0,06\,.$$

Dann ist nach Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0,95 \cdot 0,06}{0,95 \cdot 0,06 + 0,10 \cdot 0,94} = 0,378\,.$$

3. Ziegenproblem. In einer Fernsehshow sind drei Türen, hinter einer befindet sich ein Auto (das der Kandidat gewinnen möchte), hinter den beiden anderen Ziegen. Der Kandidat zeigt auf Tür 1. Der Moderator (der weiß, wo das Auto ist) öffnet Tür 3, hinter der sich eine Ziege befindet. Der Kandidat hat ein zweite Chance. Soll er wechseln?

Sei $B_i = \{\text{Auto hinter Tür } i\}$, $P(B_i) = \frac{1}{3}$. $A = \{\text{Quizmaster zeigt, dass hinter Tür 3 eine Ziege ist}\}$. Der Kandidat ist also an $P(B_1|A)$ und $P(B_2|A)$ interessiert. Es ist $P(A|B_1) = \frac{1}{2}$, $P(A|B_2) = 1$, $P(A|B_3) = 0$, also $P(A) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$. Mit Bayes erhalten wir

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{1}{3},$$

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{2}{3}.$$

Der Kandidat sollte also wechseln.

Seien nun Ereignisse $A_1, A_2, \ldots, A_n$ gegeben. Mit Induktion beweist man sofort

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Definition. Die Ereignisse $A_1, \ldots, A_n$ heißen *unabhängig*, falls $P(A_{i_1} \cap \ldots \cap A_{i_r}) = \prod_{j=1}^{r} P(A_{i_j})$ für alle $I = \{i_1, \ldots, i_r\} \subseteq \{1, \ldots, n\}$. Falls $A_1, \ldots, A_n$ unabhängig sind, so sind sie auch paarweise unabhängig, aber nicht umgekehrt.

Beispiele. 1. Sei $\Omega = \{1, 2, 3, 4\}$ mit Gleichverteilung, $A_1 = \{1, 2\}$, $A_2 = \{2, 3\}$, $A_3 = \{1, 3\}$. Dann gilt $P(A_i \cap A_j) = P(A_i)P(A_j)$ für alle $i \neq j$, aber $P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3)$.
2. Wir werfen eine Münze drei Mal, $\Omega = \{K, Z\}^3$. $A = \{\text{mindestens 2 Mal Kopf}\}$, $B = \{\text{1. Wurf ist Kopf}\}$, $C = \{\text{der 2. und 3. Wurf zeigen dasselbe}\}$. Wir haben $|\Omega| = 8$, $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = \frac{3}{8}$ (KKK, KKZ, KZK), $P(A \cap C) = \frac{1}{4}$ (KKK, ZKK), $P(B \cap C) = \frac{1}{4}$ (KKK, KZZ), $P(A \cap B \cap C) = \frac{1}{8}$ (KKK). A, B, C sind nicht unabhängig.

Lemma 1.5. *Sind $A_1, \ldots, A_n$ unabhängig, so auch die komplementären Ereignisse $A_1^c, \ldots, A_n^c$.*

Beweis. Sei $A_{i_1}, \ldots, A_{i_r}$ gegeben, wobei wir uns auf $A_1, \ldots, A_r$ beschränken können. Nach Inklusion-Exklusion haben wir

$$P(A_1^c \cap \ldots \cap A_r^c) = P((A_1 \cup \ldots \cup A_r)^c) = \sum_{k=0}^{n} (-1)^k \sum_{1 \le i_1 < \cdots < i_k \le r} P(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Andererseits ist

$$P(A_1^c)\cdots P(A_r^c) = \prod_{i=1}^{r}(1 - P(A_i)) = \sum_{k=0}^{r}(-1)^k \sum_{1\le i_1<\cdots<i_k\le r} P(A_{i_1})\cdots P(A_{i_k})$$

$$= \sum_{k=0}^{r}(-1)^k \sum_{1\le i_1<\cdots<i_k\le r} P(A_{i_1}\cap\ldots\cap A_{i_k})$$

wegen der Unabhängigkeit der A_i. $\square$

2 Zufallsvariablen

2.1 Induzierter Raum und Verteilung

Wir kommen zum wichtigsten Begriff der W-Theorie. In den meisten Situationen sind wir nicht an der gesamten Verteilung interessiert, sondern nur an bestimmten Daten, die vom Experiment abhängen.

Beispiele. Im Lotto interessiert die Anzahl der richtigen Zahlen, beim zweimaligen Würfeln zum Beispiel die Augensumme, und beim Münzwurf, im wievielten Wurf zum ersten Mal Kopf kommt.

Definition. Sei $(\Omega, \mathcal{E}, P)$ ein W-Raum, $X : \Omega \to W$ (Wertebereich), $\mathcal{F}$ σ-Algebra auf W. Falls $X^{-1}(F) \in \mathcal{E}$ für alle $F \in \mathcal{F}$, so heißt X *Zufallsvariable* auf Ω, genauer $(W, \mathcal{F})$-wertige Zufallsvariable.

Wenn W endlich oder abzählbar ist, so sei $\mathcal{F} = 2^W$. Wir werden fast ausschließlich $W \subseteq \mathbb{R}$ betrachten, $\mathcal{F}$ = Borelmengen. X heißt dann *reelle* Zufallsvariable.

Definition. Der durch X *induzierte* W-Raum ist $(W, \mathcal{F}, P_X)$ mit $P_X(F) = P(X^{-1}(F))$. Wir schreiben kurz $P(X \in F)$ für $P(\{\omega \in \Omega : X(\omega) \in F\})$. Zum Beispiel $P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$ oder $P(X \le x) = P(\{\omega \in \Omega : X(\omega) \le x\}$, falls $W = \mathbb{R}$.

Satz 2.1. *P_X ist W-Maß auf $(W, \mathcal{F})$.*

Beweis. $P_X(W) = P(X^{-1}(W)) = P(\Omega) = 1$, und ferner $P_X(\dot{\bigcup\limits_i} F_i) = P(X^{-1}(\dot{\bigcup\limits_i} F_i)) = P(\dot{\bigcup\limits_i} X^{-1}(F_i)) = \sum\limits_i P(X^{-1}(F_i)) = \sum\limits_i P_X(F_i)$. $\square$

Beispiele. 1. Münzwurf, $\Omega = \{K, Z\}$, $P(K) = p$, $P(Z) = 1 - p$. Sei X = Nummer des ersten Kopfwurfs, $W = \mathbb{N}$. Es ist $P(X = n) = (1-p)^{n-1}p$, und wir erhalten die geometrische Verteilung für X.

2. Zweimaliges Würfeln, $\Omega = \{1, 2, \dots, 6\}^2$ mit Gleichverteilung. Sei X= Augensumme, $W = \{1, 2, \dots, 12\}$. Dann ist $P(X = s) = \frac{1}{36} \cdot \#\{(i,j) \in \Omega : i + j = s\}$, zum Beispiel $P(X = 10) = \frac{3}{36} = \frac{1}{12}$.

Definition. Sei $X : \Omega \to \mathbb{R}$. Die *Verteilungsfunktion* von X ist $F_X : \mathbb{R} \to [0, 1]$ mit

$$F_X(x) = P(X \le x) = P(\{\omega \in \Omega : X(\omega) \le x\}).$$

Offenbar gilt

$$\lim_{x \to -\infty} F(x) = 0, \ \lim_{x \to \infty} F(x) = 1\,,$$
$$x < y \Rightarrow F(x) \le F(y)\,.$$

Ferner ist $F(x)$ *rechtsstetig*, das heißt mit $h \searrow 0$ geht $F(x+h) \to F(x)$.

Beispiele. 1. Münzwurf. $\Omega = \{K, Z\}$, $X(K) = 1$, $X(Z) = 0$, $P(K) = p$, $P(Z) = 1 - p$. Es ist

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1\,. \end{cases}$$

2. Indikatorvariable. Sei $A \in \mathcal{E}$, $I_A : \Omega \to \mathbb{R}$ mit

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A\,. \end{cases}$$

Es ist

$$F_{I_A}(x) = \begin{cases} 0 & x < 0 \\ 1-P(A) & 0 \le x < 1 \\ 1 & x \ge 1\,. \end{cases}$$

Angenommen, X und Y sind zwei reelle Zufallsvariablen, $X, Y : \Omega \to \mathbb{R}$. Wir fassen (X, Y) als Zufallsvariable auf mit $(X,Y) : \Omega^2 \to \mathbb{R}^2$, $(X,Y)(\omega_1, \omega_2) = (X(\omega_1), Y(\omega_2))$. Allgemein seien $X_1, \ldots, X_n : \Omega^n \to \mathbb{R}$, dann ist $(X_1, \ldots, X_n) : \Omega^n \to \mathbb{R}^n$. Die *gemeinsame* Verteilung ist

$$F(x_1, \ldots, x_n) = P(X_1 \le x_1 \wedge X_2 \le x_2 \wedge \ldots \wedge X_n \le x_n)\,.$$

2.2 Diskrete Zufallsvariablen

Eine (reelle) Zufallsvariable $X : \Omega \to W \subseteq \mathbb{R}$ heißt *diskret*, falls W endlich oder abzählbar ist.

Seien $X, Y : \Omega \to W \subseteq \mathbb{R}$ reelle Zufallsvariablen. Wir werden X und Y als unabhängig ansehen, falls

$$P(X \in A \wedge Y \in B) = P(X \in A)P(X \in B)$$

für alle $A, B \subseteq W$ gilt. Das folgende (leicht zu beweisende) Resultat zeigt, dass man sich bei diskreten Zufallsvariablen auf einelementige Mengen A, B beschränken kann.

Lemma 2.2. *Zwei diskrete Zufallsvariablen $X, Y : \Omega \to W \subseteq \mathbb{R}$ sind genau dann unabhängig, wenn*

$$P(X = x \wedge Y = y) = P(X = x)P(Y = y) \quad \textit{für alle } x, y \in W$$

gilt.

Mit Induktion gilt analog das allgemeine Resultat: Die Zufallsvariablen $X_1, \ldots, X_n : \Omega \to W \subseteq \mathbb{R}$ sind genau dann unabhängig, wenn

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

ist für alle $(x_1, \ldots, x_n) \in W^n$.

Beispiel. Betrachten wir den Münzwurf, $\Omega = \{K, Z\}$, $P(K) = p$, $P(Z) = 1 - p$, $0 < p < 1$. Es sei $X = \#K$, $Y = \#Z$ bei einmaligem Wurf. Dann ist

$$P(X = 1 \wedge Y = 1) = 0,\ P(X = 1)P(Y = 1) = p(1 - p)\,,$$

also sind, wie zu erwarten, X und Y nicht unabhängig.

Nehmen wir nun an, die Münze wird N Mal geworfen, wobei N Poisson verteilt ist, mit $\lambda > 0$. Das heißt, die W-keit N Mal zu werfen, ist $\frac{\lambda^N}{N!}e^{-\lambda}$. Dann ist

$$\begin{aligned} P(X = x \wedge Y = y) = {} & P(X = x \wedge Y = y \big| N = x + y)P(N = x + y) \\ = {} & \binom{x+y}{x} p^x (1-p)^y \frac{\lambda^{x+y}}{(x+y)!} e^{-\lambda} = \frac{(\lambda p)^x}{x!} \frac{(\lambda(1-p))^y}{y!} e^{-\lambda}\,, \end{aligned}$$

und

$$\begin{aligned} P(X = x) &= \sum_{n \geq x} P(X = x \big| N = n)P(N = n) = \sum_{n \geq x} \binom{n}{x} p^x (1-p)^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \frac{(\lambda p)^x}{x!} \sum_{n \geq x} \frac{(\lambda(1-p))^{n-x}}{(n-x)!} = e^{-\lambda} \frac{(\lambda p)^x}{x!} e^{\lambda(1-p)} = \frac{(\lambda p)^x}{x!} e^{-\lambda p}\,. \end{aligned}$$

Analog erhalten wir $P(Y = y) = \frac{(\lambda(1-p))^y}{y!} e^{-\lambda(1-p)}$, und wir sehen, dass X und Y unabhängig sind.

Nun führen wir die wichtigsten Maßzahlen ein.

Definition. Sei $X : \Omega \to \mathbb{R}$ diskrete Zufallsvariable. Der *Erwartungswert* $\mathrm{E}[X]$ ist

$$\mathrm{E}[X] = \sum_x xP(X = x) = \sum_{\omega \in \Omega} X(\omega)P(\omega)\,,$$

falls die Summe absolut konvergiert.

Der Erwartungswert gibt also an, was für ein Wert von X im Mittel zu erwarten ist. Es gibt also stets ω, ω' mit

$$X(\omega) \leq \mathrm{E}[X],\ \ X(\omega') \geq \mathrm{E}[X]\,.$$

Für endliche Mengen Ω oder endlichem Wertebereich existiert $\mathrm{E}[X]$ natürlich immer.

Im Fall der Gleichverteilung auf endlichem Ω, $P(\omega) = \frac{1}{|\Omega|}$ für $\omega \in \Omega$, haben wir

$$\mathrm{E}[X] = \frac{1}{|\Omega|} \sum_\omega X(\omega)\,.$$

Ist $\Omega = \{\omega_1, \ldots, \omega_n\}$, $X(\omega_i) = x_i$, so können wir

$$\mathrm{E}[X] = \frac{x_1 + \cdots + x_n}{n}$$

schreiben. $\mathrm{E}[X]$ ist somit der übliche Durchschnittswert.

Lemma 2.3. *Ist $X : \Omega \to \mathbb{R}$ Zufallsvariable, $g : \mathbb{R} \to \mathbb{R}$ beliebig, so gilt*

$$\mathrm{E}[g(X)] = \sum_x g(x)P(X = x) = \sum_\omega g(X(\omega))P(\omega)\,,$$

falls die Summe konvergiert.

Beweis. Wir setzen $Y = g \circ X : \Omega \to \mathbb{R}$, $\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$. Dann ist

$$E[Y] = \sum_\omega Y(\omega)P(\omega) = \sum_\omega g(X(\omega))P(\omega) = \sum_x \sum_{\omega : X(\omega) = x} g(x)P(\omega) =$$

$$= \sum_x g(x)P(X = x). \qquad \square$$

Künftig werden wir den Zusatz, "falls die Summe existiert" weglassen. Alle Sätze sind mit dieser Einschränkung zu verstehen.

Beispiel. X nehme die Werte $-2,-1,1,3$ mit W-keiten $\frac{1}{4},\frac{1}{8},\frac{1}{4},\frac{3}{8}$ an. Es sei $Y=X^2$. Dann ist

$$\mathrm{E}[X^2]=4\cdot\frac{1}{4}+1\cdot\frac{1}{8}+1\cdot\frac{1}{4}+9\cdot\frac{3}{8}=\frac{38}{8}=\frac{19}{4}\,.$$

Definition. Sei $k\in\mathbb{N}$. Das k-te *Moment* von X ist

$$m_k=\mathrm{E}[X^k]\,.$$

Das k-te *zentrale Moment* ist

$$\sigma_k=E[(X-\mathrm{E}[X])^k]\,.$$

Besonders wichtig sind

$$\mathrm{E}[X]=m_1,\ \mathrm{Var}[X]=\sigma_2=E[(X-\mathrm{E}[X])^2]\,.$$

$\mathrm{Var}[X]$ heißt die *Varianz*, $\sqrt{\mathrm{Var}[X]}$ die *Streuung*.

Die Varianz gibt also an, welche (quadratische) *Abweichung* vom Erwartungswert im Mittel zu erwarten ist. Ist zum Beispiel X eine konstante Funktion, $X(\omega)=c$ für alle ω, so haben wir $\mathrm{E}[X]=c$, $\mathrm{Var}[X]=0$.

Satz 2.4 (Linearität des Erwartungswertes)**.** *Seien* $X_1,\dots,X_n:\Omega\to\mathbb{R}$, $X=\lambda_1X_1+\cdots+\lambda_nX_n$. *Dann ist* $\mathrm{E}[X]=\sum\limits_{i=1}^{n}\lambda_i\mathrm{E}[X_i]$.

Beweis. Wir haben

$$\mathrm{E}[X]=\sum_{\omega}X(\omega)P(\omega)=\sum_{\omega}(\sum_{i=1}^{n}\lambda_iX_i(\omega))P(\omega)=\sum_{i=1}^{n}\lambda_i\mathrm{E}[X_i]\,. \qquad \square$$

Folgerung 2.5. *Es ist* $\mathrm{Var}[X]=\mathrm{E}[X^2]-(\mathrm{E}[X])^2$.

Beweis. Es gilt

$$\begin{aligned}\mathrm{Var}[X]&=\mathrm{E}[(X-\mathrm{E}[X])^2]=E[X^2-2\mathrm{E}[X]X+(\mathrm{E}[X])^2]\\&=\mathrm{E}[X^2]-2\mathrm{E}[X]\mathrm{E}[X]+(\mathrm{E}[X])^2\\&=\mathrm{E}[X^2]-(\mathrm{E}[X])^2\,. \qquad \square\end{aligned}$$

Satz 2.6. *Es seien* $X:\Omega\to T\subseteq\mathbb{R}$, $Y:\Omega\to U\subseteq\mathbb{R}$ *unabhängige Zufallsvariablen. Dann gilt* $\mathrm{E}[XY]=\mathrm{E}[X]\mathrm{E}[Y]$.

Beweis. Wir haben

$$\begin{aligned}
\mathrm{E}[XY] &= \sum_{\omega} X(\omega)Y(\omega)P(\omega) = \sum_{(x,y)\in T\times U} xyP(X = x \wedge Y = y) \\
&= \sum_{(x,y)} xyP(X = x)P(Y = y) = \sum_{x} xP(X = x) \sum_{y} yP(Y = y) \\
&= \mathrm{E}[X]\mathrm{E}[Y]\,. \qquad \square
\end{aligned}$$

Satz 2.7. *Für die Varianz gilt*

a. $\mathrm{Var}[X + c] = \mathrm{Var}[X]\,,$

b. $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]\,,$

c. $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$, *falls X, Y unabhängig sind.*

Beweis. a. $\mathrm{Var}[X+c] = \mathrm{E}[(X+c-\mathrm{E}[X+c])^2] = E[(X-\mathrm{E}[X])^2] = \mathrm{Var}[X]$.

b.
$$\begin{aligned}
\mathrm{Var}[cX] &= \mathrm{E}[(cX - \mathrm{E}[cX])^2] = \mathrm{E}[(cX - c\mathrm{E}[X])^2] \\
&= \mathrm{E}[c^2(X - \mathrm{E}[X])^2] = c^2\mathrm{Var}[X]\,.
\end{aligned}$$

c.
$$\begin{aligned}
\mathrm{Var}[X + Y] &= \mathrm{E}[(X + Y)^2] - (\mathrm{E}[X + Y])^2 \\
&= \mathrm{E}[X^2] + 2\mathrm{E}[XY] + \mathrm{E}[Y^2] - (\mathrm{E}[X])^2 - 2\mathrm{E}[X]\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\
&= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 + \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y]\,. \qquad \square
\end{aligned}$$

Beispiele. 1. Bernoulli Verteilung. $\Omega = \{K, Z\}$, $P(K) = p$, $P(Z) = 1 - p$, $X : \Omega \to \{0, 1\}$, $X(K) = 1$, $X(Z) = 0$. Dann ist $P(X = 1) = p$, $P(X = 0) = 1 - p$ und somit

$$\mathrm{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p,\ \mathrm{E}[X^2] = p,\ \mathrm{Var}[X] = p - p^2 = p(1 - p)\,.$$

Die größte Varianz erhalten wir also für die faire Münze $p = \frac{1}{2}$.

2. Binomialverteilung. Wir führen unabhängig n Würfe durch mit $P(X = k) = \binom{n}{k}p^k(1 - p)^{n-k}$, $X =$ Anzahl Kopf, $0 \le k \le n$. Sei $X_i = \begin{cases} 1 \\ 0 \end{cases}$ die Indikatorvariable im i-ten Wurf Kopf oder Zahl. Dann ist $X = X_1 + \cdots + X_n$, also

$$\mathrm{E}[X] = np,\ \mathrm{Var}[X] = np(1 - p)\,.$$

3. Poissonverteilung. Sei $P(X = n) = \frac{\lambda^n}{n!}e^{-\lambda}$, $n \geq 0$. Dann ist

$$\mathrm{E}[X] = \sum_{n\geq 0} n\frac{\lambda^n}{n!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{n\geq 1}\frac{\lambda^{n-1}}{(n-1)!} = \lambda\,,$$

$$\begin{aligned}\mathrm{E}[X^2] &= \sum_{n\geq 0} n^2\frac{\lambda^n}{n!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{n\geq 1} n\frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda}\sum_{n\geq 0}(n+1)\frac{\lambda^n}{n!} \\ &= \lambda e^{-\lambda}\left(\lambda\sum_{n\geq 1}\frac{\lambda^{n-1}}{(n-1)!} + e^{\lambda}\right) = \lambda^2 + \lambda\,.\end{aligned}$$

Also ist

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

Definition. Seien $X, Y : \Omega \to \mathbb{R}$. Die *Kovarianz* von X und Y ist

$$\mathrm{cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

Ausmultiplizieren ergibt

$$\mathrm{cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]\,,$$

also insbesondere $\mathrm{cov}[X, X] = \mathrm{Var}[X]$.

X und Y heißen *unkorreliert*, falls $\mathrm{cov}[X, Y] = 0$ ist. Wegen Satz 2.6 sind unabhängige Variablen X, Y stets unkorreliert.

Der *Korrelationskoeffizient* ist

$$\rho(X, Y) = \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}\,.$$

Lemma 2.8. *Es gilt für Konstanten a, b*

$$\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] + 2ab\mathrm{cov}[X, Y]\,.$$

Beweis. Ausrechnen. $\square$

Lemma 2.9. *Sei $X : \Omega \to \mathbb{R}$, dann gilt*

a. $\mathrm{E}[X^2] = 0 \Leftrightarrow P(X = 0) = 1$,

b. $\mathrm{Var}[X] = 0 \Leftrightarrow P(X = \mathrm{E}[X]) = 1$.

Beweis. a. Es gilt $\mathrm{E}[X^2] = \sum_x x^2 P(X = x)$. Ist also $\mathrm{E}[X^2] = 0$, so muß $P(X = x) = 0$ sein für alle $x \neq 0$, also $P(X \neq 0) = 0$ und somit $P(X = 0) = 1$. Die Umkehrung ist ebenso klar.
b. Mit $\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2]$ brauchen wir nur a) auf $X - \mathrm{E}[X]$ anzuwenden. $\square$

Satz 2.10 (Cauchy-Schwarz Ungleichung)**.** *Es gilt*

$$(\mathrm{E}[XY])^2 \leq \mathrm{E}[X^2]\mathrm{E}[Y^2]\,,$$

mit Gleichheit genau dann, wenn $P(Y = aX) = 1$ für ein $a \in \mathbb{R}$, oder $P(X = a'Y) = 1$ für $a' \in \mathbb{R}$.

Beweis. Angenommen $\mathrm{E}[X^2] = 0$, dann ist $P(X = 0) = 1$ nach Lemma 2.9, somit $P(X = x) = 0$ für $x \neq 0$. Daraus folgt $P(X = x \wedge Y = y) = 0$ für $x \neq 0, y$, also

$$\mathrm{E}[XY] = \sum_{(x,y)} xyP(X = x \wedge Y = y) = 0,$$

und wir haben $P(X = 0 \cdot Y) = 1$. Ist umgekehrt $P(X = 0) = 1$, so gilt $\mathrm{E}[X^2] = 0, \mathrm{E}[XY] = 0$, und wir erhalten Gleichheit. Wir können also $\mathrm{E}[X^2] > 0$ annehmen. Wir betrachten die Zufallsvariable $Z = Y - aX$ für $a \neq 0$. Dann gilt

$$0 \leq \mathrm{E}[Z^2] = \mathrm{E}[(Y - aX)^2] = \mathrm{E}[Y^2] - 2a\mathrm{E}[XY] + a^2\mathrm{E}[X^2]\,.$$

Als Polynom in a hat diese quadratische Gleichung höchstens eine reelle Nullstelle, also ist die Diskriminante

$$4\mathrm{E}[XY]^2 - 4\mathrm{E}[X^2]\mathrm{E}[Y^2] \leq 0\,,$$

und das ist die Cauchy-Schwarz Ungleichung. Gleichheit gilt genau für $E[(Y - aX)^2] = 0$, also genau für $P(Y = aX) = 1$ nach Lemma 2.9. $\square$

Folgerung 2.11. *Für den Korrelationskoeffizienten $\rho(X, Y)$ gilt $|\rho(X, Y)| \leq 1$, mit Gleichheit genau dann, wenn $P(Y = aX + b) = 1$, $a, b \in \mathbb{R}$, oder $P(X = a'Y + b') = 1$, $a', b' \in \mathbb{R}$.*

Beweis. Wir setzen $U = X - \mathrm{E}[X]$, $V = Y - E[Y]$, dann ist $\mathrm{cov}[X, Y] = \mathrm{E}[UV]$, $\mathrm{Var}[X] = \mathrm{E}[U^2]$, $\mathrm{Var}[Y] = \mathrm{E}[V^2]$. Satz 2.10 ergibt somit

$$(\mathrm{cov}[X, Y])^2 \leq \mathrm{Var}[X]\mathrm{Var}[Y]$$

also

$$|\rho(X, Y)| \leq 1.$$

Gleichheit gilt für $\mathrm{Var}[X] = 0$ oder $P(V = aU) = 1$ für $a \in \mathbb{R}$. Im ersten Fall ist nach Lemma 2.9, $P(X = \mathrm{E}[X]) = 1$, also $P(X = 0 \cdot Y + \mathrm{E}[X]) = 1$, und umgekehrt. Im zweiten Fall haben wir $P(Y - E[Y] = a(X - \mathrm{E}[X])) = 1$, also $P(Y = aX + b) = 1$ mit $b = -a\mathrm{E}[X] + E[Y]$, und umgekehrt. $\square$

Wir besprechen noch zwei weitere fundamentale Ungleichungen.

Satz 2.12. *Es gilt*

a. *Ungleichung von Markov: Sei $X \geq 0$, dann ist*

$$P(X \geq a) \leq \frac{\mathrm{E}[X]}{a} \quad (a > 0)\,.$$

b. *Ungleichung von Tschebyschev:*

$$P(|X - \mathrm{E}[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2} \quad (a > 0)\,.$$

Beweis. a. Wir haben

$$\begin{aligned} \mathrm{E}[X] = \sum_{\omega} X(\omega)P(\omega) &= \sum_{X(\omega)\geq a} X(\omega)P(\omega) + \underbrace{\sum_{X(\omega)<a} X(\omega)P(\omega)}_{\geq 0} \\ &\geq a \sum_{X(\omega)\geq a} P(\omega) = aP(X \geq a). \end{aligned}$$

b. Sei $Y = |X - \mathrm{E}[X]|^2$, dann ist $P(Y \geq a^2) \leq \frac{\mathrm{E}[Y]}{a^2}$ nach a), somit

$$P(|X - \mathrm{E}[X]| \geq a) = P(|X - \mathrm{E}[X]|^2 \geq a^2) \leq \frac{\mathrm{E}[Y]}{a^2} = \frac{\mathrm{Var}[X]}{a^2}\,. \qquad \square$$

Beispiel. Es seien $X_1, \ldots, X_n$ unabhängige Kopien von X (zum Beispiel n Münzwürfe), $Y = X_1 + \cdots + X_n$. Dann ist $\mathrm{E}[Y] = n\mathrm{E}[X]$, $\mathrm{Var}[Y] = n\mathrm{Var}[X]$, also

$$\mathrm{E}[\frac{X_1 + \cdots + X_n}{n}] = \mathrm{E}[X], \ \mathrm{Var}[\frac{X_1 + \cdots + X_n}{n}] = \frac{\mathrm{Var}[X]}{n} .$$

Tschebyschev ergibt

$$P(\Big|\frac{X_1 + \cdots + X_n}{n} - \mathrm{E}[X]\Big| \geq a) \leq \frac{1}{n}\frac{\mathrm{Var}[X]}{a^2}$$

und dies geht gegen 0 für $n \to \infty$.

Ebenso wie bedingte W-keiten können wir auch bedingte Erwartungswerte betrachten. Es seien $X, Y : \Omega \to \mathbb{R}$, dann sei $Y\big|X = x$ die Zufallsvariable Y unter der Annahme $X = x$. Wir setzen $\psi(x) = \mathrm{E}[Y\big|X = x]$, $\psi(X)$ ist also wieder eine Zufallsvariable.

Satz 2.13. *Es gilt* $\mathrm{E}[\psi(X)] = \mathrm{E}[Y]$, *das heißt also*

$$\mathrm{E}[Y] = \sum_x \mathrm{E}[Y\big|X = x)P(X = x) .$$

Beweis. Wir haben

$$\begin{aligned}
\mathrm{E}[\psi(X)] &= \sum_x \psi(x)P(X = x) = \sum_x \mathrm{E}[Y\big|X = x]P(X = x) \\
&= \sum_x \sum_y yP(Y = y\big|X = x)P(X = x) \\
&= \sum_x \sum_y yP(Y = y \wedge X = x) = \sum_y y \sum_x P(Y = y \wedge X = x) \\
&= \sum_y yP(Y = y) = \mathrm{E}[Y] . \qquad \square
\end{aligned}$$

Beispiel. Sei X die Augensumme bei zweimaligem Würfeln mit Gleichverteilung, $Y = \#$ gerade Ziffern. Dann ist

$$\mathrm{E}[Y\big|X = 2] = 0, \ \mathrm{E}[Y\big|X = 3] = 1, \ E[Y\big|X = 4] = \frac{2}{3}, \ldots$$

Offenbar ist $\mathrm{E}[Y] = 1$.

2.3 Wahrscheinlichkeitserzeugende Funktionen

Für diskrete Zufallsvariablen $X : \Omega \to \mathbb{N}_0$ können wir eine kompakte Form finden.

Definition. Sei $X : \Omega \to \mathbb{N}_0$. Dann heißt die formale Reihe

$$G_X(z) = \sum_{n\geq 0} P(X = n) z^n$$

die *wahrscheinlichkeitserzeugende Funktion.*

Lemma 2.14. *Falls die jeweiligen Reihen konvergieren, so gilt:*

a. $\mathrm{E}[X] = G'(1)$,

b. $\mathrm{Var}[X] = G''(1) + G'(1) - (G'(1))^2$.

Beweis. a. Wir haben

$$\mathrm{E}[X] = \sum_{n\geq 0} nP(X = n) = G'(1)\,.$$

b. $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \sum_{n\geq 0} n^2 P(X = n) - (\mathrm{E}[X])^2$. Nun ist

$$G''(z) = \sum_{n\geq 1} n(n-1)P(X = n)z^{n-2} = \sum_{n\geq 2} n^2 P(X = n)z^{n-2} - \sum_{n\geq 2} nP(X = n)z^{n-2}$$

also

$$\sum_{n\geq 1} n^2 P(X = n) = G''(z)\big|_{z=1} + G'(z)\big|_{z=1}$$

somit

$$\mathrm{Var}[X] = G''(1) + G'(1) - (G'(1))^2\,. \qquad \square$$

Beispiele. 1. Binomialverteilung, $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Hier ist $G(z) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = (pz + (1-p))^n$, also $G'(z) = np(pz+(1-p))^{n-1}$, $G''(z) = n(n-1)p^2(pz + (1-p))^{n-2}$. Es folgt

$$\mathrm{E}[X] = np,\ \mathrm{Var}[X] = n(n-1)p^2 + np - n^2p^2 = np(1-p)\,.$$

2. Poissonverteilung, $P(X = n) = \frac{\lambda^n}{n!}e^{-\lambda}$. Hier ist $G(z) = \sum\limits_{n\geq 0} \frac{\lambda^n}{n!}e^{-\lambda}z^n = e^{-\lambda}e^{\lambda z} = e^{\lambda(z-1)}$, $G'(z) = \lambda e^{\lambda(z-1)}$, $G''(z) = \lambda^2 e^{\lambda(z-1)}$. Es folgt

$$\mathrm{E}[X] = \lambda,\ \mathrm{Var}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda\,.$$

Sind $A(z) = \sum\limits_{n\geq 0} a_n z^n$, $B(z) = \sum\limits_{n\geq 0} b_n z^n$, dann können wir das Produkt bilden. Der Koeffizient von z^n in $A(z)B(z)$ ist $\sum\limits_{k=0}^{n} a_k b_{n-k}$. Wir nennen

$$A(z)B(z) = \sum_{n\geq 0}(\sum_{k=0}^{n} a_k b_{n-k})z^n$$

die *Konvolution* von $A(z)$ und $B(z)$. Wir schreiben $[z^n]A(z)$ für den Koeffizienten von z^n in $A(z)$.

Beispiel. Hypergeometrische Verteilung. $X : \Omega \to \{0, 1, \ldots, r\}$, $P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$, n, m, r fest.

Sei $G_m(z) = \sum\limits_{k\geq 0} \binom{r}{k}\binom{n-r}{m-k} z^k$, und

$$H(z,t) = (1 + tz)^r(1 + t)^{n-r}\,.$$

Dann ist

$$[t^m]H(z,t) = \sum_{k=0}^{m} \binom{r}{k} z^k \binom{n-r}{m-k} = G_m(z)\,,$$

also

$$H(z,t) = \sum_{m\geq 0} G_m(z)t^m\,.$$

Es folgt

$$\begin{aligned}
G'_m(z) &= [t^m]\frac{\partial}{\partial z}H(z,t) \\
&= [t^m]rt(1 + tz)^{r-1}(1 + t)^{n-r} \\
&= r[t^{m-1}](1 + tz)^{r-1}(1 + t)^{n-r} \\
&= r\sum_{k=0}^{m-1} \binom{r-1}{k} z^k \binom{n-r}{m-1-k}\,,
\end{aligned}$$

und mit $z = 1$,

$$G'_m(1) = r \sum_{k=0}^{m-1} \binom{r-1}{k} \binom{n-r}{m-1-k} = r \binom{n-1}{m-1}.$$

Daraus erhalten wir

$$\mathrm{E}[X] = \frac{1}{\binom{n}{m}} r \binom{n-1}{m-1} = \frac{rm}{n}.$$

Beispiel. Ein berühmtes Problem betrifft Zufallswege auf $\mathbb{Z}$. Ein Wanderer startet in 0 und geht immer mit W-keit $1/2$ einen Schritt nach rechts oder links. Was ist die W-keit p_0, dass er letztlich wieder nach 0 zurückkommt?

Es sei a_n die Anzahl der Wege, die im n-ten Schritt zum *ersten Mal* nach 0 zurückkehren. Weiter sei b_n die Anzahl der Wege, die im n-ten Schritt nach 0 kommen. Also zum Beispiel $a_0 = 0$, $a_1 = 0$, $a_2 = 2$, $b_0 = 1$, $b_1 = 0$, $b_2 = 2$. Sei $A(z) = \sum\limits_{n \geq 0} a_n z^n$, $B(z) = \sum\limits_{n \geq 0} b_n z^n$. Das Ereignis $A = \{\text{kehrt zurück}\}$ ist daher die disjunkte Summe $A = \bigcup A_n$, $A_n = \{\text{kehrt im } n\text{-ten Schritt zum ersten Mal zurück}\}$ mit $P(A_n) = \frac{a_n}{2^n}$. Wir haben also

$$p_0 = \sum_{n \geq 0} \frac{a_n}{2^n} = A(\frac{1}{2}).$$

Nun ist

$$b_n = \sum_{k=0}^{n} a_k b_{n-k}$$

durch Klassifikation, wann er zum ersten Mal zurückkehrt. Wir erhalten also

$$B(z) = A(z)B(z) + 1,$$

somit

$$A(z) = 1 - \frac{1}{B(z)},$$

und insbesondere

$$p_0 = 1 \Leftrightarrow B(\frac{1}{2}) = \infty.$$

Zur Analyse von $B(z)$ sehen wir zunächst, das der Wanderer eine *gerade* Anzahl $2n$ von Schritten tun muß, um zu 0 zurückzukehren, n in jeder Richtung. Also ist

$$b_{2n} = \binom{2n}{n}, \; b_{2n+1} = 0 \quad (n \geq 0)\,,$$

und somit

$$B(\frac{1}{2}) = \sum_{n \geq 0} \frac{\binom{2n}{n}}{2^{2n}}\,.$$

Durch Induktion verifiziert man leicht

$$\binom{2n}{n} \geq \frac{2^{2n}}{2n}\,,$$

also

$$B(\frac{1}{2}) \geq 1 + \sum_{n \geq 1} \frac{1}{2n} = 1 + \frac{1}{2} \sum_{n \geq 1} \frac{1}{n} = \infty\,,$$

da die harmonische Reihe $\sum\limits_{n \geq 1} \frac{1}{n}$ divergiert.

Ergebnis: Der Wanderer kehrt mit W-keit 1 zum Ausgangspunkt 0 zurück.

Übrigens gilt im Fall, dass n mit W-keit p nach rechts geht und mit W-keit $1-p$ nach links geht, stets $p_0 < 1$ für alle $p \neq \frac{1}{2}$. Er kehrt also nur im Fall der Gleichverteilung fast sicher zu 0 zurück.

2.4 Binomialverteilung und ihre Approximation

Wir betrachten den besonders wichtigen Fall, dass die Zufallsvariable $X : \Omega \to \mathbb{R}$ binomialverteilt ist. Es sei $p = W$-keit Erfolg, $1 - p = W$-keit Mißerfolg, $X =$ Anzahl der Erfolge bei n Versuchen. Dann ist

$$P(X = k) = b(k, n; p) = \binom{n}{k} p^k (1-p)^{n-k}\,,$$

$$\mathrm{E}[X] = np, \; \mathrm{Var}[X] = np(1-p)\,.$$

Es ergibt sich das Problem, dass die Binomialkoeffizienten $\binom{n}{k}$ wegen des raschen Wachstums von $n!$ schwer zu berechnen sind. Man wird also versuchen, gute Approximationen zu bestimmen.

Betrachten wir zunächst die hypergeometrische Verteilung

$$\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}\,.$$

Angenommen wir haben die Situation:

$$n, n-r, r \gg m, k\,,$$

das heißt $n, n-r, r$ sind "viel größer" als m und k.

Beispiel. Wir haben 500 Schrauben, 60 davon sind defekt, $n = 500$, $r = 60$, und wir testen $m = 5$ Schrauben.

Mit Zurücklegen haben wir die Binomialverteilung, ohne Zurücklegen ist die W-keit für Erfolg $\sim \frac{r}{n}$. Die Idee ist, dass für große $n, r, n-r$ die Ziehungen von m Schrauben sich wenig unterscheiden, ob wir mit oder ohne Zurücklegen ziehen.

Aufgrund der Annahmen machen wir die folgenden Approximationen:

1. $k \ll r \Rightarrow r(r-1)\cdots(r-k+1) \sim r^k$
2. $m \ll r \Rightarrow n(n-1)\cdots(n-m+1) \sim n^m$
3. $m-k \ll n-r \Rightarrow (n-r)(n-r-1)\cdots(n-r-m+k+1) \sim (n-r)^{m-k}$.

Wir erhalten

$$\begin{aligned} h(k, m; r, n) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} &= \tfrac{m!}{k!(m-k)!}\,\cdot \\ &\quad \tfrac{r(r-1)\cdots(r-k+1)(n-r)(n-r-1)\cdots(n-r-m+k+1)}{n(n-1)\cdots(n-m+1)} \\ &\sim \tbinom{m}{k}\tfrac{r^k(n-r)^{m-k}}{n^m} = \tbinom{m}{k}(\tfrac{r}{n})^k(1-\tfrac{r}{n})^{m-k} \\ &= b(k, m; \tfrac{r}{n})\,. \end{aligned}$$

Im Erwartungswert stimmen die Ausdrücke überein.

$$\mathrm{E}[X_{\text{hyp}}] = \frac{rm}{n}, \quad \mathrm{E}[X_{\text{bin}}] = m\frac{r}{n}\,.$$

Beispiel. $n = 60$, $r = 25$, $m = 4$. Die Tabelle zeigt, dass wir gute Approximationen erhalten:

	$k = 0$	1	2	3	4
$b(k, 4; \frac{25}{60})$	$0,107$	$0,336$	$0,366$	$0,165$	$0,026$
$h(k, 4; 25, 60)$	$0,116$	$0,331$	$0,354$	$0,167$	$0,032$

Natürlich benötigen wir auch für die hypergeometrische Verteilung Binomialkoeffizienten. In der Praxis verwendet man als Approximation die Poisson Verteilung

$$p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} .$$

Wir machen die Voraussetzungen

1. $n \gg 0 \Rightarrow (1 + \frac{x}{n})^n \sim e^x$,
2. $p \ll 1$, k klein $\Rightarrow (1 - p)^k \sim 1$,
3. $k \ll n \Rightarrow n(n-1)\cdots(n-k+1) \sim n^k$.

Daraus erhalten wir

$$\begin{aligned} b(k, n; p) &= \tbinom{n}{k} p^k (1-p)^{n-k} = \tfrac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\ &\sim \tfrac{n^k}{k!} p^k (1-p)^{n-k} = \tfrac{1}{k!}(np)^k \tfrac{(1-p)^n}{(1-p)^k} \\ &\sim \tfrac{1}{k!}(np)^k (1-p)^n = \tfrac{1}{k!}(np)^k (1 - \tfrac{np}{n})^k \\ &\sim \tfrac{1}{k!}(np)^k e^{-np} = p(k; np) , \end{aligned}$$

also

$$b(k, n; p) \sim p(k; \lambda) \text{ mit } \lambda = np .$$

Wiederum sehen wir Gleichheit für den Erwartungswert

$$\mathrm{E}[X_{\text{bin}}] = np, \quad \mathrm{E}[X_{\text{pois}}] = \lambda = np .$$

Beispiel. Jemand spielt 3 Jahre im Lotto. Mit welcher W-keit kann man mindestens 3 richtige erwarten? Wir haben $p(X \geq 3) = 0,018$, also ist die W-keit

$$1 - b(0, 3 \cdot 52; 0,018) \sim 1 - p(0; 3 \cdot 52 \cdot 0,018) = 0,93 .$$

2.5 Zufallsvariablen mit Dichtefunktion

Definition. Die Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt *stetig verteilt* mit Dichtefunktion, falls die Verteilungsfunktion $F(x) = P(X \leq x)$ von der Form

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

ist, wobei $f(x)$ integrierbar ist, $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$. Es gilt dann

$$\lim_{x\to-\infty} F(x) = 0,\ \lim_{x\to\infty} F(x) = 1,\ F'(x) = f(x)\,.$$

Für $A \in \mathcal{E}$ haben wir

$$P(X \in A) = \int_A f(t)dt\,,$$

insbesondere

$$P(a \leq X \leq b) = \int_a^b f(t)dt\,.$$

Achtung: Es ist stets $P(X = x) = 0$.

Es seien $X, Y : \Omega \to \mathbb{R}$ stetig verteilte Zufallsvariablen mit Verteilungsfunktionen $F_X(x)$, $F_Y(y)$ und Dichten $f_X(x)$, $f_Y(y)$. Die *gemeinsame Verteilung* $F(x, y)$, falls sie existiert, ist

$$F(x, y) = P(X \leq x \wedge Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)dvdu\,.$$

Definition. $X, Y : \Omega \to \mathbb{R}$ heißen *unabhängig*, falls die Ereignisse $\{X \leq x\}$, $\{Y \leq y\}$ unabhängig sind für alle x, y, also wenn

$$F(x, y) = F_X(x)F_Y(y)\,,$$

gilt. Allgemein für n Variablen.

Lemma 2.15. *Sind* $X, Y : \Omega \to \mathbb{R}$ *stetig verteilt mit* $F(x,y)$, $f(x,y)$ *und* $F_X(x), f_X(x), F_Y(y), f_Y(y)$ *wie zuvor, so gilt*

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \; f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

X *und* Y *sind unabhängig genau dann, wenn*

$$f(x,y) = f_X(x)f_Y(y) \, .$$

Beweis. Wir haben

$$F_X(x) = P(X \le x) = \lim_{y\to\infty} P(X \le x \wedge Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,v)dvdu$$

$$= \int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f(u,y)dy\right)du \, ,$$

und somit

$$f_X(x) = F_X'(x) = \int_{-\infty}^{\infty} f(x,y)dy \, ,$$

analog für $f_Y(y)$. X und Y unabhängig bedeutet $F(x,y) = F_X(x)F_Y(y)$ für alle x, y, somit

$$\frac{\partial}{\partial x}F(x,y) = F_X'(x)F_Y(y), \; \frac{\partial^2}{\partial x \partial y}F(x,y) = F_X'(x)F_Y'(y) = f_X(x)f_Y(y) \, ,$$

also

$$f(x,y) = f_X(x)f_Y(y) \, .$$

Umgekehrt folgt aus $f(x,y) = f_X(x)f_Y(y)$,

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dvdu = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)dvdu = F_X(x)F_Y(y) \, . \qquad \square$$

Allgemein gilt: $X_1, \dots, X_n : \Omega \to \mathbb{R}$ unabhängig

$$\Leftrightarrow F(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \text{ für alle } (x_1, \dots, x_n)$$

$$\Leftrightarrow f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \text{ für alle } (x_1, \dots, x_n)\,.$$

Wir erklären wieder den Erwartungswert und die Varianz. Alle Sätze für diskrete Variablen gelten auch hier, die Beweise sind allerdings etwas subtiler.

Definition. Der *Erwartungswert* und die *Varianz* von X sind

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx, \ \mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f(x) dx.$$

Beispiel. Gleichverteilung auf $[a, b]$. Hier ist $f(x) = \frac{1}{b-a}$,

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

Wir berechnen

$$\mathrm{E}[X] = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}\,,$$

$$\mathrm{Var}[X] = \int_a^b (x - \frac{a+b}{2})^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{(x - \frac{a+b}{2})^3}{3}\Big|_a^b = \frac{(b-a)^2}{12}\,.$$

Insbesondere erhalten wir für das Einheitsintervall $[0, 1]$, $\mathrm{E}[X] = \frac{1}{2}$, $\mathrm{Var}[X] = \frac{1}{12}$.

Der folgende Satz gibt eine einfache Methode an, falls die Dichte $f(x) = 0$ ist für $x < 0$.

Lemma 2.16. *Sei $X : \Omega \to \mathbb{R}$ mit $f(x) = 0$ für $x < 0$. Dann gilt*

$$\mathrm{E}[X] = \int_0^{\infty} (1 - F(x)) dx\,,$$

wobei $F(x)$ die Verteilungsfunktion ist.

Beweis. Wir haben

$$\int_0^\infty (1 - F(x))dx = \int_0^\infty P(X > x)dx = \int_{x=0}^\infty \left(\int_{y=x}^\infty f(y)dy \right) dx$$

$$= \int_{y=0}^\infty \left(\int_{x=0}^y dx \right) f(y)dy = \int_{y=0}^\infty y f(y)dy = \mathrm{E}[X]\,. \qquad \square$$

Beispiel. Exponentialverteilung. $X : [0, \infty) \to [0, \infty)$ mit $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$ für $x \geq 0$, $f(x) = 0$ für $x < 0$, $F(x) = 1 - e^{-\lambda x}$. Aus dem Lemma folgt

$$\mathrm{E}[X] = \int_0^\infty e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^\infty = \frac{1}{\lambda}\,.$$

Wir wollen nun die entsprechenden Sätze für $\mathrm{E}[X]$ und $\mathrm{Var}[X]$ herleiten.

Satz 2.17. *Sei $X : \Omega \to \mathbb{R}$ und $Y = g(X)$, wobei $g : \mathbb{R} \to \mathbb{R}$ stetig ist und $g(x) \geq 0$ für alle x. Dann gilt*

$$\mathrm{E}[g(X)] = \int_{-\infty}^\infty g(x) f(x)dx\,.$$

Beweis. Da $Y \geq 0$ ist, gilt nach dem Lemma

$$\mathrm{E}[Y] = \int_0^\infty (1 - F_Y(y))dy\,,$$

wobei

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y),\ 1 - F_Y(y) = P(g(X) > y)\,.$$

Sei A das Ereignis $A = \{x : g(x) > y\}$, dann ist

$$P(g(X) > y) = \int_A f(x)dx\,,$$

und somit

$$\mathrm{E}[Y] = \int_0^\infty \left(\int_A f(x)dx \right) dy = \int_{-\infty}^\infty \left(\int_{y=0}^{g(x)} dy \right) f(x)dx = \int_{-\infty}^\infty g(x) f(x)dx\,. \qquad \square$$

Bemerkung. Der Satz gilt auch ohne die Einschränkung $g(x) \geq 0$.

Folgerung 2.18. *Wir haben für* $X : \Omega \to \mathbb{R}$

a. $\mathrm{E}[aX+b] = a\mathrm{E}[X]+b$,

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

c. $\mathrm{Var}[aX+b] = a^2\mathrm{Var}[X]$.

Beweis. Mit $Y = aX+b$, $g(x) = ax+b$ gilt

$$\mathrm{E}[aX+b] = \int_{-\infty}^{\infty} (ax+b)f(x)dx = a\mathrm{E}[X]+b\,.$$

b. Sei $Y = (X-\mathrm{E}[X])^2$, $g(x) = (x-\mathrm{E}[X])^2$, dann ist

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} (x-\mathrm{E}[X])^2 f(x)dx = \mathrm{Var}[X]\,.$$

Ausrechnen ergibt

$$\begin{aligned}\mathrm{Var}[X] &= \int_{-\infty}^{\infty} (x^2-2x\mathrm{E}[X]+(\mathrm{E}[X])^2)f(x)dx \\ &= \mathrm{E}[X^2]-2(\mathrm{E}[X])^2+(\mathrm{E}[X])^2 = \mathrm{E}[X^2]-(\mathrm{E}[X])^2\,.\end{aligned}$$

c. $$\begin{aligned}\mathrm{Var}[aX+b] &= \int_{-\infty}^{\infty} (ax-a\mathrm{E}[X])^2 f(x)dx \\ &= a^2\int_{-\infty}^{\infty} (x-E[X])^2 f(x)dx = a^2\mathrm{Var}[X]\,. \qquad \square\end{aligned}$$

Beispiel. Standardnormalverteilung. Sei $X : \mathbb{R} \to \mathbb{R}$ mit Dichte $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Dann ist mit der Substitution $y = -x$

$$\mathrm{E}[X] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}}dx = -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ye^{-\frac{y^2}{2}}dy = -\mathrm{E}[X]\,,$$

also $\mathrm{E}[X] = 0$. Für die Varianz erhalten wir mit partieller Integration

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \mathrm{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot x e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \left(-x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) = 1\,.$$

Wir haben noch nicht die allgemeine Linearität des Erwartungswertes für stetige Zufallsvariablen bewiesen. Dazu müssen wir uns überlegen, wie die Dichte der Summe $X + Y$ aussieht.

Problem I. Gegeben X mit $F_X(x)$, $f_X(x)$ und $Y = g(X)$. Was ist $F_Y(y)$, $f_Y(y)$?

Problem II. Gegeben X und Y mit $F_X(x), f_X(x)$, $F_Y(y), f_Y(y)$. Was ist $F_{X+Y}(z)$, $f_{X+Y}(z)$?

Satz 2.19. *Sei $X : \Omega \to \mathbb{R}$ mit Dichte $f(x)$, $Y = g(X)$, wobei $g(x)$ monoton steigend ist mit $g'(x) \neq 0$ für alle x. Dann ist die Dichtefunktion $h(y)$ von Y gegeben durch*

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}\,, \quad g^{-1} \text{ Umkehrfunktion.}$$

Beweis. Es ist $F_X(x) = \int_{-\infty}^{x} f(t)dt = P(X \leq x)$, also

$$F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f(t)dt\,.$$

Mit der Substitution $t = g^{-1}(u)$ haben wir $dt = \frac{1}{g'(g^{-1}(u))} du$ und erhalten

$$F_Y(y) = \int_{-\infty}^{y} \frac{f(g^{-1}(u))}{g'(g^{-1}(u))} du\,,$$

und somit

$$h(y) = F_Y(y)' = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))}\,. \qquad \square$$

Bemerkung. Der Satz gilt allgemein, so lange $g'(x) \neq 0$ ist für alle x.

Beispiel. Sei X $N(0,1)$-verteilt $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$, $Y = \sigma X + \mu$, $\sigma > 0$. Sei $g(x) = \sigma x + \mu$, dann ist $g^{-1}(y) = \frac{y-\mu}{\sigma}$, $g'(x) = \sigma$. Für die Dichte von Y erhalten wir demnach

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

das heißt Y ist $N(\mu, \sigma^2)$-verteilt. Ferner ist

$$\mathrm{E}[Y] = \sigma \mathrm{E}[X] + \mu = \mu, \mathrm{Var}[X] = \sigma^2 \mathrm{Var}[X] = \sigma^2 .$$

Beispiel. Sei X $N(0,1)$-verteilt. Was ist $h(y)$ für $Y = X^2$? Wir haben $g(x) = x^2$, $g^{-1}(y) = \pm\sqrt{y}$, $g'(x) = 2x$. Indem wir die positive und negative Wurzel in Betracht ziehen, erhalten wir mit $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$$h(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, \ y \geq 0 .$$

Diese Dichtefunktion sollte bekannt vorkommen, es ist die Γ-Verteilung mit $r = \alpha = \frac{1}{2}$. Also erhalten wir das Resultat:

$$X \sim N(0,1) \Rightarrow X^2 \sim \gamma_{\frac{1}{2},\frac{1}{2}} .$$

Betrachten wir nun Problem II.

Satz 2.20. *Seien $X, Y : \Omega \to \mathbb{R}$ und $f(x,y)$ die gemeinsame Dichte. Dann hat $Z = X + Y$ die Dichte*

$$f_Z(z) = \int_{-\infty}^{\infty} f(t, z-t) dt .$$

Beweis. Sei A das Ereignis

$$A = \{(x,y) : x + y \leq z\} = \{(x,y) : Z \leq z\} .$$

Dann ist

$$P(Z \leq z) = \iint_A f(u,v) du dv = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{z-u} f(u,v) dv du .$$

Die Substitution $x = u$, $y = u + v$ mit $u = x$, $v = y - x$ ergibt die Jacobi Matrix

$$\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = 1$$

also $dudv = dxdy$, und wir erhalten

$$P(Z \le z) = \int\limits_{x=-\infty}^{\infty} \int\limits_{y=-\infty}^{z} f(x, y-x)dydx = \int\limits_{y=-\infty}^{z} \left(\int\limits_{-\infty}^{\infty} f(x, y-x)dx \right) dy\,,$$

also

$$f_Z(z) = \int\limits_{-\infty}^{\infty} f(t, z-t)dt\,. \qquad \square$$

Folgerung 2.21. *Sind X, Y unabhängig, so gilt*

$$f_{X+Y}(x) = \int\limits_{-\infty}^{\infty} f_X(t) f_Y(x-t)dt\,.$$

Wir nennen diese Operation die *Faltung* zweier Funktionen f_X, f_Y, in Zeichen $f_X * f_Y$. Somit ist für unabhängige X und Y: $f_{X+Y} = f_X * f_Y$.

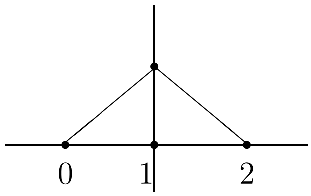
Beispiel. Es seien X und Y gleichverteilte unabhängige Variablen auf $[0, 1]$, also $f(x) = f_X(x) = 1$, $g(x) = f_Y(x) = 1$ für alle $x \in [0, 1]$ und 0 außerhalb. Für $X + Y$ erhalten wir die Dichte

$$h(x) = \int\limits_{0}^{x} f(t)g(x-t)dt\,.$$

Für $x \le 1$ ergibt dies $h(x) = \int\limits_0^x dt = x$. Sei $x \ge 1$. Mit $t \le 1$, $x - t \le 1$ haben wir $x - 1 \le t \le 1$

$$h(x) = \int\limits_{x-1}^{1} dt = 2 - x.$$

Die Dichte von $X + Y$ ist also die Funktion



Für den Erwartungswert und die Varianz ergibt sich daraus

$$\begin{aligned}
\mathrm{E}[X+Y] &= \int_0^2 xh(x)dx = \int\limits_0^1 x^2 dx + \int_1^2 (2x - x^2)dx \\
&= \frac{x^3}{3}\Big|_0^1 + x^2\Big|_1^2 - \frac{x^3}{3}\Big|_1^2 = \frac{1}{3} + 3 - \frac{8}{3} + \frac{1}{3} = 1, \\
\mathrm{E}[(X+Y)^2] &= \int\limits_0^2 x^2 h(x)dx = \frac{7}{6} \\
\mathrm{Var}[X+Y] &= \frac{7}{6} - 1 = \frac{1}{6}\,.
\end{aligned}$$

Wir sehen also mit $\mathrm{E}[X] = \mathrm{E}[Y] = \frac{1}{2}$, $\mathrm{Var}[X] = \mathrm{Var}[Y] = \frac{1}{12}$, dass tatsächlich $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$ und $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$ gilt. Dies wollen wir allgemein zeigen.

Folgerung 2.22. *Seien* $X, Y : \Omega \to \mathbb{R}$. *Dann gilt* $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$.

Beweis. Es seien $f_X(x)$, $f_Y(y)$ die Dichten und $f(x,y)$ die gemeinsame Dichte. Wir haben

$$\begin{aligned}
\mathrm{E}[X+Y] &= \int\limits_{z=-\infty}^{\infty} z \int\limits_{t=-\infty}^{\infty} f(t, z-t)dtdz \\
&= \int\limits_t \left(\int\limits_z zf(t, z-t)dz \right) dt \quad (z = y + t) \\
&= \int\limits_t \left(\int\limits_y (y+t)f(t,y)dy \right) dt
\end{aligned}$$

$$= \int_t \left(\int_y y f(t,y) dy + t \int_y f(t,y) dy \right) dt$$
$$= \int_y y \left(\int_t f(t,y) dt \right) dy + \int_t t \left(\int_y f(t,y) dy \right) dt$$
$$= \int_y y f_Y(y) dy + \int_t t f_X(t) dt = \mathrm{E}[Y] + \mathrm{E}[X]\,. \qquad \square$$

Daraus folgt nach Folgerung 2.18a) auch $\mathrm{E}[aX+bY] = a\mathrm{E}[X]+b\mathrm{E}[Y]$. Allgemein verwendet man Induktion, um die Linearität $\mathrm{E}[\sum_{i=1}^{n} \lambda_i X_i] = \sum_{i=1}^{n} \lambda_i \mathrm{E}[X_i]$ zu zeigen. Die Regel $\mathrm{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$ für unabhängige Variablen X_i wird ähnlich bewiesen.

Beispiel. Wir geben ohne Beweis die Resultate für die Normalverteilung und Γ-Verteilung an.

1. Ist X_1 $N(\mu_1, \sigma_1^2)$-verteilt, X_2 $N(\mu_2, \sigma_2^2)$-verteilt und X_1, X_2 unabhängig, so ist $X_1 + X_2$ $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$-verteilt.

2. Sind X_1 γ_{α,r_1}-verteilt, X_2 γ_{α,r_2}-verteilt, X_1, X_2 unabhängig, so ist $X_1 + X_2$ γ_{α,r_1+r_2}-verteilt.

3. Es seien $X_1, \ldots, X_n$ unabhängige Standard normalverteilte Zufallsvariablen. Dann ist $Y = X_1^2 + \cdots + X_n^2$ $\gamma_{\frac{1}{2},\frac{n}{2}}$-verteilt. Wir sagen, Y ist χ^2-verteilt mit n Freiheitsgraden. Die χ^2-Verteilung wird eine große Rolle in der Statistik spielen.

Beispiel. Seien $X, Y : \Omega \to [0, \infty)$ unabhängige Zufallsvariable, beide exponentiell verteilt zum selben $\lambda > 0$. Dann ist $f_X(x) = \lambda e^{-\lambda x}$, $f_Y(y) = \lambda e^{-\lambda y}$. Es gilt

$$f_{X+Y}(x) = \int_0^x \lambda e^{-\lambda t} \cdot \lambda e^{-\lambda(x-t)} dt$$
$$= \lambda^2 e^{-\lambda x} \int_0^x dt = \lambda^2 e^{-\lambda x} x\,.$$

Allgemein haben wir für unabhängige Variablen $X_1, \dots, X_n$ alle zum selben $\lambda > 0$

$$f_{X_1+\cdots+X_n}(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}\,.$$

Für $n = 1, 2$ wissen wir das. Nun verwenden wir Induktion.

$$f_{X_1+\cdots+X_n}(x) = f_{(X_1+\cdots+X_{n-1})+X_n}(x) = \int_0^x \frac{\lambda^{n-1} t^{n-2}}{(n-2)!} e^{-\lambda t} \cdot \lambda e^{-\lambda(x-t)} dt$$

$$= \frac{\lambda^n e^{-\lambda x}}{(n-2)!} \int_0^x t^{n-2} dt = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}\,.$$

2.6 Exponentialverteilung

Wartezeiten werden im einfachsten Fall folgendermaßen modelliert: Es sei X die Wartezeit mit Werten in $[0, \infty)$ mit Dichte $f(x)$.

Als Beispiel könnten wir X als Wartezeit auf den ersten freien Parkplatz (kontinuierlich) oder als Wartezeit auf das erste Mal Kopf beim Münzwurf (diskret) interpretieren.

Definition. Wir sprechen von einer *gedächtnislosen* Wartezeit, wenn

$$P(X \geq s + t \big| X \geq t) = P(X \geq s)$$

für alle s, t gilt.

Dies bedeutet mit

$$P(X \geq s + t \big| X \geq t) = \frac{P(X \geq s + t \wedge X \geq t)}{P(X \geq t)} = \frac{P(X \geq s + t)}{P(X \geq t)}\,,$$

dass die Definition äquivalent ist zu

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

Satz 2.23. a. *Sei X exponential verteilt, $f(x) = \lambda e^{-\lambda x}$, dann gilt $P(X \geq s + t) = P(X \geq s) P(X \geq t)$.*
b. *Sei umgekehrt $f : [0, \infty) \to [0, \infty)$ stetig mit Verteilung $F(x) = P(X \leq x)$ gegeben und $P(X \geq s + t) = P(X \geq s) P(X \geq t)$, dann ist $f(x) = \lambda e^{-\lambda x}$ für ein $\lambda > 0$.*

Beweis. a. $P(X \geq x) = 1 - F(x) = e^{-\lambda x}$, also

$$P(X \geq s+t) = e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t} = P(X \geq s)P(X \geq t)\,.$$

b. Sei $G(x) = \int\limits_x^\infty f(t)dt$, also

$$G(x) = P(X \geq x) = 1 - F(x)\,.$$

Wir stellen fest:

1. $G(0) = 1$,
2. $G(x) \to 0$ für $x \to \infty$,
3. $G(s+t) = G(s)G(t)$ für alle $s, t \geq 0$.

Angenommen es gibt $s \geq 0$ mit $G(s) = 0$. Dann ist wegen 3), $G(s) = G(\frac{s}{n})^n$, also $G(\frac{s}{n}) = 0$ für alle n, das heißt $G(0) = 0$ wegen der Stetigkeit von G, im Widerspruch zu 1).

Es sei $G(1) = e^\alpha$, also $\alpha = \log G(1)$.

Behauptung. $G(t) = e^{\alpha t}$ für alle t.
Zunächst haben wir $G(n) = G(1)^n = e^{\alpha n}$, dann $G(m) = G(\frac{m}{n})^n$ also $e^{\alpha m} = G(\frac{m}{n})^n$, das heißt $G(\frac{m}{n}) = e^{\alpha\frac{m}{n}}$ für alle $\frac{m}{n} \in \mathbb{Q}$. Die Funktionen $G(t)$ und $e^{\alpha t}$ stimmen also auf $\mathbb{Q}$ überein, und somit wegen der Stetigkeit auf $[0, \infty)$. Außerdem gilt $\alpha < 0$ wegen $G(x) = e^{\alpha x} \to 0$ für $x \to \infty$ nach 2).

Nun setzen wir $\lambda = -\alpha$ und erhalten

$$F(x) = 1 - G(x) = 1 - e^{-\lambda x}\,,$$

also $f(x) = F'(x) = \lambda e^{-\lambda x}\,.$ $\square$

Beispiel. Angenommen die Wartezeit im Sprechzimmer ist im Durchschnitt 20 Minuten. Was ist die W-keit, nach höchstens 10 Minuten dranzukommen? Wir setzen $\lambda = \frac{1}{20}$ ($\mathrm{E}[X] = \frac{1}{\lambda}$) und berechnen

$$P(X \leq 10) = 1 - e^{-\frac{10}{20}} = 1 - e^{-\frac{1}{2}} = 0,39.$$

Nach wievielen Minuten komme ich mit mindestens 90% W-keit an die Reihe? Sei T die Anzahl der Minuten, dann haben wir

$$\frac{1}{20}\int_0^T e^{-\frac{x}{20}}dx = 1 - e^{-\frac{T}{20}} \geq 0,9 \;\Rightarrow\; e^{-\frac{T}{20}} \leq 0,1$$

$$\Rightarrow -\frac{T}{20} \leq \log 0,1 \;\Rightarrow\; T \geq -20 \cdot \log 0,1 = 46,05.$$

Nun nehmen wir an, dass die Variablen $X_1, \ldots, X_n$ unabhängig sind, alle exponential verteilt zum selben $\lambda > 0$.

Satz 2.24. a. *Die Variable* $X_{\min} = \min(X_1, \ldots, X_n)$ *hat Dichtefunktion* $m_n(x) = n\lambda e^{-n\lambda x}$, *das heißt* X *ist exponential verteilt zu* $n\lambda$.

b. *Die Variable* $X_{\max} = \max(X_1, \ldots, X_n)$ *hat Dichtefunktion* $M_n(x) = n\lambda e^{-\lambda x} \cdot (1 - e^{-\lambda x})^{n-1}$.

Beweis. a. Gesucht ist $m_n(x)$ mit

$$P(X_{\min} \geq x) = \int_x^\infty m_n(t)dt\,.$$

Nun ist

$$\begin{aligned} P(X_{\min} \geq x) &= P(X_1 \geq x \wedge \ldots \wedge X_n \geq x) \\ &= \prod_{i=1}^n P(X_i \geq x) = \prod_{i=1}^n (1 - F(x)) \\ &= e^{-\lambda n x}\,, \end{aligned}$$

und somit

$$F_{X_{\min}}(x) = 1 - P(X_{\min} \geq x) = 1 - e^{-\lambda n x}\,,$$

und daher $m_n(x) = \lambda n e^{-\lambda n x}$.

b. Gesucht ist $M_n(x)$ mit $P(X_{\max} \leq x) = \int_0^x M_n(t)dt$. Nun ist

$$P(X_{\max} \leq x) = P(X_1 \leq x \wedge \ldots \wedge X_n \leq x) = \prod_{i=1}^n P(X_i \leq x)$$

$$= \prod_{i=1}^n (1 - e^{-\lambda x}) = (1 - e^{-\lambda x})^n\,,$$

und es folgt

$$M_n(x) = F_{X_{\max}}(x)' = n(1 - e^{-\lambda x})^{n-1} e^{-\lambda x} \lambda \,. \qquad \square$$

Folgerung 2.25. *Wir haben*

a. $\mathrm{E}[X_{\min}] = \frac{1}{n\lambda}$,

b. $\mathrm{E}[X_{\max}] = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n})$, $\mathrm{E}[X_{\max}] \to \infty$ *für* $n \to \infty$.

Beweis. Teil a) wissen wir schon, da $X_{\min}$ exponentialverteilt ist nach $n\lambda$.
b. Wir wenden Lemma 2.16 an und erhalten

$$\mathrm{E}[X_{\max}] = \int_0^\infty (1 - (1 - e^{-\lambda x})^n) dx \,.$$

Mit der Substitution $y = 1 - e^{-\lambda x}$ haben wir $e^{-\lambda x} = 1 - y$, $-\lambda x = \log(1-y)$, also $x = -\frac{1}{\lambda} \log(1 - y)$, $dx = \frac{1}{\lambda} \frac{1}{1-y} dy$. Dies ergibt

$$\begin{aligned} \mathrm{E}[X_{\max}] &= \frac{1}{\lambda} \int_0^1 (1 - y^n) \frac{1}{1-y} dy = \frac{1}{\lambda} \int_0^1 (1 + y + \cdots + y^{n-1}) dy \\ &= \frac{1}{\lambda}(y + \frac{y^2}{2} + \cdots + \frac{y^n}{n})\Big|_0^1 = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n}) \,. \end{aligned}$$

Insbesondere geht $\mathrm{E}[X_{\max}] \to \infty$, da die harmonische Reihe divergiert. $\square$

Beispiel. Wenn bei 5 Telefonzellen die mittlere Wartezeit 2 Minuten beträgt ($\lambda = \frac{1}{2}$), so ist

$$\begin{aligned} \mathrm{E}[X_{\min}] &= \frac{2}{5} = 0,4 \,, \\ \mathrm{E}[X_{\max}] &= 2(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}) = \frac{137}{30} = 4,57 \,. \end{aligned}$$

Bemerkung. Die diskrete Variante für gedächtnislose Wartezeiten wird genau von der geometrisch verteilten Zufallsvariablen beschrieben.

3 Grenzwertsätze

3.1 Konvergenz von Zufallsvariablen

Betrachten wir als Beispiel die Zahlen $y \in [0,1)$ in Dezimalform $y = 0, y_1 y_2 y_3 \cdots = \sum_{j=1}^{\infty} y_j 10^{-j}$. Es sei Y_j die diskrete Zufallsvariable $Y_j : [0,1) \to y_j \in \{0, 1, \ldots, 9\}$, mit $P(Y_j = y) = \frac{1}{10}$ für alle y. Nun setzen wir

$$X_n = \sum_{j=1}^{n} 10^{-j} Y_j$$

und lassen $X_n \to Y = \sum_{j=1}^{\infty} 10^{-j} Y_j$ gehen. Die Y_j sind unabhängig und wir vermuten, dass Y gleichverteilt auf $[0,1)$ ist.

Definition. Es seien $X_1, X_2, \ldots$ reelle Zufallsvariablen auf dem W-Raum $(\Omega, \mathcal{E}, P)$.

1. $X_n \to X$ *fast sicher*: $\Leftrightarrow P(\{\omega \in \Omega : X_n(\omega) \stackrel{n \longrightarrow \infty}{\longrightarrow} X(\omega)\}) = 1$, in Zeichen $X_n \xrightarrow{f.s.} X$.

2. $X_n \to X$ *in Wahrscheinlichkeit* (oder stochastisch): $\Leftrightarrow P(|X_n - X| > \varepsilon) \xrightarrow{n \to \infty} 0$ für alle $\varepsilon > 0$, in Zeichen $X_n \xrightarrow{i.W.} X$.

3. $X_n \to X$ *in Verteilung*: $\Leftrightarrow P(X_n \leq x) \xrightarrow{n \to \infty} P(X \leq x)$ für alle x, für die $F(x) = P(X \leq x)$ stetig ist, in Zeichen $X_n \xrightarrow{i.V.} X$.

Die Interpretation ist: $X_n \xrightarrow{f.s.} X$ falls für große n, $X_n(\omega) \sim X(\omega)$; $X_n \xrightarrow{i.W.} X$, falls für große n, X_n nahe bei X ist.

Satz 3.1. a. $X_n \xrightarrow{f.s.} X_n$ *impliziert* $X_n \xrightarrow{i.W.} X$. *Die Umkehrung gilt nicht allgemein.*
b. $X_n \xrightarrow{i.W.} X_n$ *impliziert* $X_n \xrightarrow{i.V.} X$. *Die Umkehrung gilt nicht allgemein.*

Beweis. a. Es sei o.B.d.A. $X = 0$, ansonsten betrachten wir $X_n - X$. Es sei $A_n = \{\omega \in \Omega : |X_n(\omega)|, |X_{n+1}(\omega)|, \ldots \leq \varepsilon\}$, also $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$. Gilt $X_n(\omega \to 0$, so haben wir $|X_n(\omega)| \leq \varepsilon$ für $n \geq N$, also $\omega \in A_N$. Es folgt

$$\{\omega \in \Omega : X_n(\omega) \to 0\} \subseteq \bigcup_{n \geq 1} A_n \,,$$

also aus Lemma 1.1a) und wegen $X_n \xrightarrow{f.s.} 0$

$$1 = P(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n)\,.$$

Nun haben wir $\{\omega \in \Omega : |X_n(\omega)| > \varepsilon\} \subseteq \Omega \setminus A_n$, also

$$P(|X_n| > \varepsilon) \leq 1 - P(A_n),$$

und es folgt $\lim\limits_{n\to\infty} P(|X_n| > \varepsilon) = 0$, das heißt $X_n \overset{i.W.}{\to} 0$.

Um zu zeigen, dass $X_n \overset{i.W.}{\to} X \not\Rightarrow X_n \overset{f.s.}{\to} X$ betrachten wir das folgende Gegenbeispiel. Es seien $X_1, X_2, \ldots$ unabhängig auf Ω mit Werten in $\{0, 1\}$,

$$P(X_n = 1) = \frac{1}{n},\; P(X_n = 0) = 1 - \frac{1}{n}\,.$$

Wir haben

$$\begin{aligned} P(X_n > \varepsilon) &= \frac{1}{n} \to 0 \quad \text{für } 0 < \varepsilon < 1 \\ P(X_n > \varepsilon) &= 0 \to 0 \quad \text{für } \varepsilon \geq 1\,, \end{aligned}$$

also $X_n \overset{i.W.}{\to} 0$.

Sei nun $A_n = \{\omega \in \Omega : X_k(\omega) = 0 \text{ für } k \geq n\}$. Dann ist wieder $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, also $P(\bigcup A_n) = \lim P(A_n)$. Nun gilt

$$\begin{aligned} \bigcup_{n\geq 1} A_n =&\; \{\omega \in \Omega : X_n(\omega) = 0 \text{ für } n \geq n_0\} = \{\omega \in \Omega : X_n(\omega) \to 0\}\,, \\ P(A_n) =&\; P(X_n = 0 \wedge X_{n+1} = 0 \wedge \ldots) = \prod_{i\geq 0} P(X_{n+i} = 0) \\ =&\; (1 - \frac{1}{n})(1 - \frac{1}{n+1}) \cdots\,, \end{aligned}$$

also

$$\begin{aligned} P(A_n) =&\; \lim_{M\to\infty} (1 - \frac{1}{n})(1 - \frac{1}{n+1}) \cdots (1 - \frac{1}{M}) \\ =&\; \lim_{M\to\infty} (\frac{n-1}{n} \cdot \frac{n}{n+1} \cdots \frac{M-1}{M}) = \lim_{M\to\infty} \frac{n-1}{M} = 0\,, \end{aligned}$$

und somit

$$P(\{\omega \in \Omega : X_n(\omega) \to 0\}) = P(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n) = 0\,.$$

Demnach gilt nicht $X_n \xrightarrow{f.s.} 0$.

b. Es gelte $X_n \xrightarrow{i.W.} X$, mit $F_n(x) = P(X_n \leq x)$, $F(x) = P(X \leq x)$. Für $\varepsilon > 0$ haben wir

$$\begin{aligned} F_n(x) = & \; P(X_n \leq x) = P(X_n \leq x \wedge X \leq x + \varepsilon) + P(X_n \leq x \wedge X > x + \varepsilon) \\ \leq & \; F(x+\varepsilon) + P(|X_n - X| > \varepsilon). \end{aligned}$$

Ferner gilt

$$\begin{aligned} F(x-\varepsilon) = & \; P(X \leq x - \varepsilon) = P(X \leq x - \varepsilon \wedge X_n \leq x) + P(X \leq x - \varepsilon \wedge X_n > x) \\ \leq & \; F_n(x) + P(|X_n - X| > \varepsilon). \end{aligned}$$

Es folgt

$$F(x-\varepsilon) - P(|X_n - x| > \varepsilon) \leq F_n(x) \leq F(x+\varepsilon) + P(|X_n - x| > \varepsilon).$$

Da $X_n \xrightarrow{i.W.} X$ vorausgesetzt ist, erhalten wir

$$F(x-\varepsilon) \leq \liminf F_n(x) \leq \limsup F_n(x) \leq F(x+\varepsilon).$$

Falls $F(x)$ bei x stetig ist, so gilt mit $\varepsilon \to 0$, $F(x) = \lim_{n\to\infty} F_n(x)$, und es folgt $X_n \xrightarrow{i.V.} X$.

Als Gegenbeispiel für $X_n \xrightarrow{i.V.} X \not\Longrightarrow X_n \xrightarrow{i.W.} X$ betrachten wir Bernoulli Variablen X_n mit Werten in $\{0,1\}$ und

$$P(X_n = 1) = P(X_n = 0) = \frac{1}{2} \;\text{ für alle } n,$$

also $X_1 = X_2 = \cdots = X$. Sei nun $Y = 1 - X$. Wir haben

$$F_n(x) = P(X_n \leq x) = \begin{cases} 1 & x \geq 1\,, \\ \frac{1}{2} & 0 \leq x < 1\,, \\ 0 & x < 0\,, \end{cases}$$

und ferner $P(Y \leq x) = P(X_n \leq x)$. Es gilt also $X_n \xrightarrow{i.V.} Y$, aber nicht $X_n \xrightarrow{i.W.} Y$, da $|X_n - Y| = 1$ ist für alle n. $\square$

3.2 Das schwache Gesetz der großen Zahlen

Es seien $X_1, X_2, \ldots$ unabhängige Kopien einer Zufallsvariablen X. Dann legt uns die Intuition nahe, dass der Erwartungswert $\mathrm{E}[X]$ ungefähr gleich $\frac{X_1+\cdots+X_n}{n}$ für große n sein sollte, und das ist auch so.

Für diskrete Zufallsvariable haben wir dies schon im Beispiel nach Satz 2.12 aus der Ungleichung von Tschebyschev begründet. Wir wollen also die Ungleichungen von Markov und Tschebyschev auch für stetige Zufallsvariablen verifizieren.

Satz 3.2. *Sei X stetige Zufallsvariable mit Dichte $f(x)$.*

a. *Ist $X \geq 0$, $a > 0$, so gilt $P(X \geq a) \leq \frac{\mathrm{E}[X]}{a}$. (Markov)*

b. *$P(|X - \mathrm{E}[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2}$, $a > 0$. (Tschebyschev)*

Beweis. a. Wir haben

$$\mathrm{E}[X] = \int\limits_0^\infty x f(x) dx = \int_a^\infty x f(x) dx + \underbrace{\int\limits_0^a x f(x) dx}_{\geq 0}$$

$$\geq a \int\limits_a^\infty f(x) dx = aP(X \geq a) .$$

b. Mit $Y = |X - \mathrm{E}[X]|$ erhalten wir aus a)

$$P(|X - \mathrm{E}[X]| \geq a) = P(|X - E[X]|^2 \geq a^2) \leq \frac{E[(X - \mathrm{E}[X])^2]}{a^2} = \frac{\mathrm{Var}[X]}{a^2} . \qquad \square$$

Satz 3.3 (Schwaches Gesetz der großen Zahlen)**.** *Die reelle Zufallsvariable habe Erwartungswert $\mathrm{E}[X]$ und Varianz $\mathrm{Var}[X]$. Sind $X_1, X_2, \ldots$ unabhängige Kopien von X, so gilt*

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mathrm{E}[X]\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}[X]}{n\varepsilon^2} \quad \text{für alle } \varepsilon > 0 .$$

Insbesondere gilt für die Folge $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$, dass $\bar{X}_n \xrightarrow{i.W.} \mathrm{E}[X]$ strebt.

Beweis. Wir haben $E[\bar{X}_n] = \frac{1}{n}\sum\limits_{i=1}^{n} E[X_i] = \mathrm{E}[X]$, $\mathrm{Var}[\bar{X}_n] = \frac{1}{n^2}\sum\limits_{i=1}^{n} \mathrm{Var}[X_i] = \frac{\mathrm{Var}[X]}{n}$, also gilt nach Tschebyschev

$$P(|\bar{X}_n - \mathrm{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}[\bar{X}_n]}{\varepsilon^2} = \frac{\mathrm{Var}[X]}{n\varepsilon^2}\,. \qquad \square$$

Folgerung 3.4. *Mit den Voraussetzungen des Satzes haben wir*

$$P(|\frac{X_1 + \cdots + X_n}{n} - \mathrm{E}[X]| < t) \geq 1 - \frac{\mathrm{Var}[X]}{nt^2} \textit{ für } t > 0.$$

Beispiel. Betrachten wir eine Bernoulli Variable X mit Werten in $\{0, 1\}$, $P(X = 1) = p$, $P(X = 0) = 1 - p$, also $\mathrm{E}[X] = p$, $\mathrm{Var}[X] = p(1 - p)$. Dann gilt

$$P(|\frac{X_1 + \cdots + X_n}{n} - p| \geq t) \leq \frac{p(1-p)}{nt^2}\,,$$
$$P(|\frac{X_1 + \cdots + X_n}{n} - p| < t) \geq 1 - \frac{p(1-p)}{nt^2}\,.$$

Die Formeln werden folgendermaßen angewandt.

Es sei p unbekannt. $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$ gibt die durchschnittliche Anzahl der Erfolge an. Aus $p(1-p) \leq \frac{1}{4}$ für $p \in [0, 1]$ folgt

$$P(|\bar{X}_n - p| < t) \geq 1 - \frac{1}{4nt^2}\,.$$

Die unbekannte Erfolgswahrscheinlichkeit p liegt daher im Intervall $[\bar{x}_n - t, \bar{x}_n + t)]$ mit W-keit $\geq 1 - \frac{1}{4nt^2}$, wobei $\bar{x}_n = \frac{x_1+\cdots+x_n}{n}$ der Durchschnitt der tatsächlich beobachteten Ausgänge ist.

Angenommen wir möchten p bis auf einen Fehler $t = 0,01$ mit W-keit 98% bestimmen. Wie oft müssen wir das Experiment durchführen? Aus der Formel sehen wir

$$1 - \frac{1}{4n \cdot 0,0001} \geq 0,98\,,$$

und dies ist für $n \geq 125.000$ erfüllt.

Man beachte, dass $1 - \frac{1}{4nt^2}$ stets < 1 ist. Mit Sicherheit kann p also nicht ermittelt werden.

Beispiel. Ein Würfel wird 20 Mal geworfen und zeigt immer 6. Ist der Würfel in Ordnung, das heißt, sind die Zahlen wirklich gleichverteilt?

Es sei X_i die Zufallsvariable mit $X_i = 1$, falls im i-ten Wurf 6 erscheint, $X_i = 0$, falls eine Zahl $\neq 6$ erscheint. Dann ist für $p = \frac{1}{6}$

$$P\left(|\underbrace{\frac{X_1 + \cdots + X_{20}}{20}}_{=1} - \frac{1}{6}| \geq \frac{5}{6}\right) \leq \frac{5}{36}\frac{1}{20(\frac{5}{6})^2} = \frac{1}{100},$$

also ist der Würfel mit W-keit$\geq 99\%$ nicht in Ordnung.

3.3 Lemma von Borel-Cantelli

Sei ein W-Raum $(\Omega, \mathcal{E}, P)$ gegeben und $A_1, A_2, \ldots$ eine Folge von Ereignissen. Wir definieren

$$A^\infty = \{\omega \in \Omega : \omega \in A_i \text{ für unendlich viele } i\}.$$

Man sieht sofort

$$A^\infty = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k = \bigcap_{n=1}^{\infty} B_n \text{ mit } B_n = \bigcup_{k \geq n} A_k.$$

A^∞ ist also in $\mathcal{E}$. Aus $B_1 \supset B_2 \supset B_3 \supset \ldots$ folgt

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n \to \infty} P(B_n) = P(A^\infty).$$

Beispiel. Wir werfen einen Würfel unendlich oft, und A_n sei das Ereignis, dass im n-ten Wurf 6 geworfen wird. Es sei $\Omega = \{(\omega_1, \omega_2, \ldots)\}$ mit $\omega_i \in \{1, 2, \ldots, 6\}$, also $A^\infty = \{(\omega_1, \omega_2, \ldots) : \omega_i = 6 \text{ für unendlich viele } i\}$.

Satz 3.5 (Lemma von Borel-Cantelli)**.** *Es sei $(\Omega, \mathcal{E}, P)$ gegeben und $A_1, A_2, \ldots$ eine Folge von Ereignissen.*

a. *Ist $\sum_{k \geq 1} P(A_k) < \infty$, so gilt $P(A^\infty) = 0$.*

b. *Sind $A_1, A_2, \ldots$ unabhängig und $\sum_{k \geq 1} P(A_k) = \infty$, so gilt $P(A^\infty) = 1$.*

Beweis. a. Wegen $\sum\limits_{k\geq 1} P(A_k) < \infty$ haben wir $\sum\limits_{k\geq n} P(A_k) \to 0$ für $n \to \infty$. Daraus folgt wegen $P(B_n) \leq \sum\limits_{k\geq n} P(A_k)$, dass $P(B_n) \overset{n\to\infty}{\longrightarrow} 0$ geht, also $P(A^\infty) = \lim P(B_n) = 0$.

b. Wir verwenden die Ungleichung $1 - x \leq e^{-x}$ für $0 \leq x \leq 1$. Da die Ereignisse A_n unabhängig sind, so auch die komplementären Ereignisse A_n^c (Lemma 1.5), und wir erhalten

$$P\left(\bigcap_{k=n}^{N} A_k^c\right) = \prod_{k=n}^{N}(1 - P(A_k)) \leq \prod_{k=n}^{N} e^{-P(A_k)} = e^{-\sum\limits_{k=n}^{N} P(A_k)} .$$

Mit $N \to \infty$ ergibt sich $P\left(\bigcap\limits_{k=n}^{N} A_k^c\right) \to 0$, da $\sum\limits_{k\geq 1} P(A_k) = \infty$ ist. Nun ist $\bigcap\limits_{k=n}^{N} A_k^c \supset \bigcap\limits_{k=n}^{N+1} A_k^c \supset \ldots$, also

$$P\left(\bigcap_{k=n}^{\infty} A^c\right) = \lim_{N\to\infty} P\left(\bigcap_{k=n}^{N} A_k^c\right) = 0$$

für alle n, und es folgt

$$P((A^\infty)^c) = P\left(\bigcup_{n=1}^{\infty}\bigcap_{k\geq n} A_k^c\right) \leq \sum_{n=1}^{\infty} P\left(\bigcap_{k\geq n} A_k^c\right) = 0\,,$$

und somit $P(A^\infty) = 1$. $\square$

Bemerkung. In Teil b) kann auf die Unabhängigkeit nicht verzichtet werden. Sei nämlich $A \in \mathcal{E}$, $0 < P(A) < 1$ und $A_n = A$ für alle n. Dann ist $A^\infty = A$, $\sum P(A_n) = \infty$, aber $P(A^\infty) < 1$.

Beispiele. Wir verwenden die folgende Sprechweise: Das Ereignis A tritt *fast sicher* ein, falls $P(A) = 1$ ist.

1. Betrachten wir das Würfelbeispiel von eben, $A_n = \{6 \text{ im } n\text{-ten Wurf}\}$. Die Würfe seien unabhängig, gleichverteilt, dann gilt $P(A_n) = \frac{1}{6}$, also $\sum\limits_{n\geq 1} P(A_n) = \infty$. Die 6 wird somit fast sicher unendlich oft geworfen.

2. Wir haben Urnen $U_1, U_2, \ldots$ gegeben, wobei sich in U_n 1 weiße und $n-1$ rote Kugeln befinden. Das Ereignis A_n sei, dass aus Urne U_n eine weiße Kugel gezogen wird. Wir haben $P(A_n) = \frac{1}{n}$, also $\sum\limits_{n\geq 1} P(A_n) = \infty$ und somit $P(A^\infty) = 1$. Das heißt, fast sicher wird weiß unendlich oft gezogen.

Angenommen in Urne U_n sind 1 weiße und n^2-1 rote Kugeln. Dann ist $P(A_n) = \frac{1}{n^2}$, $\sum\limits_{n\geq 1} P(A_n) = \sum\limits_{n\geq 1} \frac{1}{n^2} = \frac{\pi^2}{6} < \infty$, also wird weiß fast sicher nur endlich oft gezogen.

Das interessante ist, dass keine Verteilung weiß/rot auf die Urnen möglich ist, wo nicht einer dieser beiden Fälle eintritt.

3. Es seien unabhängige Ereignisse $A_1, A_2, \ldots, A_r$ gegeben mit $P(A_i) > 0$ für alle i. Wir betrachten $\Omega_\infty = (\omega_1, \omega_2, \ldots)$. Kommt es oft vor, dass es ein k gibt, mit $\omega_k \in A_1$, $\omega_{k+1} \in A_2, \ldots, \omega_{k+r-1} \in A_r$?

Nehmen wir eine Telefonnummer 8018418. Nun schreiben wir die Ziffern $0, 1, 2, \ldots, 9$ gleichverteilt in einer unendlichen Folge auf. Kommt die Telefonnummer oft in der Folge vor? Dazu definieren wir

$$\begin{aligned}
B_1 &= \{(\omega_n) : \omega_1 \in A_1, \ldots, \omega_r \in A_r\} \\
B_2 &= \{(\omega_n) : \omega_{r+1} \in A_1, \ldots, \omega_{2r} \in A_r\} \\
&\ \,\vdots \\
B_m &= \{(\omega_n) : \omega_{(m-1)r+1} \in A_1, \ldots, \omega_{mr} \in A_r\} \\
&\ \,\vdots
\end{aligned}$$

Die Ereignisse B_m sind unabhängig mit $P(B_m) = P(A_1)\cdots P(A_r) > 0$ für alle m. Aus $\sum_{m\geq 1} P(B_m) = \infty$ folgt $P(B^\infty) = 1$, und das bedeutet, dass es fast sicher unendlich viele B_m gibt, in denen r aufeinanderfolgende Ergebnisse in $A_1, A_2, \ldots, A_r$ sind. Die Telefonnummer kommt also fast sicher unendlich oft vor.

Eine amüsante Illustration ist der Affe am Computer. Er tippt Buchstaben nach Belieben ein. Nach Borel-Cantelli wird er fast sicher irgendwann Goethes Faust reproduzieren – und das nicht einmal, sondern unendlich oft.

3.4 Das starke Gesetz der großen Zahlen

Es seien wieder $X_1, X_2, \ldots$ unabhängige Kopien der reellen Zufallsvariablen X. Wir wollen nun zeigen, dass die Folge $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$ sogar fast sicher gegen den Erwartungswert konvergiert. In diesem Sinn können wir also sagen, dass der Zufall bei unendlicher Wiederholung verschwindet.

Satz 3.6 (Starkes Gesetz der großen Zahlen)**.** *Für die reelle Zufallsvariable X existieren $\mathrm{E}[X]$ und $\mathrm{Var}[X]$. Es seien $X_1, X_2, \ldots$ unabhängige Kopien von X. Dann gilt*

$$P(\{\omega \in \Omega : \bar{X}_n(\omega) \to \mathrm{E}[X]\}) = 1\,,$$

das heißt $\bar{X}_n \xrightarrow{f.s.} \mathrm{E}[X]$.

Beweis. Wir können wieder $\mathrm{E}[X] = 0$ annehmen, ansonsten verwenden wir die Folge $X_n - X$. Wir gliedern den Beweis in mehrere Schritte.

1. Sei $Y_1, Y_2, \ldots$ eine Folge von Zufallsvariablen auf Ω. Wir nennen $\omega \in \Omega$ *gut*, falls $Y_n(\omega) \to 0$. Sei $B_Y = \{\omega \in \Omega : \omega \text{ gut}\}$. Wann gilt $P(B_Y) = 1$? Dazu stellen wir fest:

$$\begin{aligned} Y_n(\omega) \to 0 \quad &\Longleftrightarrow \quad \forall \varepsilon > 0 \text{ gilt } |Y_n(\omega)| < \varepsilon \text{ für } n \geq n_0 \\ &\Longleftrightarrow \quad \forall \varepsilon \geq 0 \text{ gilt } |Y_n(\omega)| \geq \varepsilon \text{ für } \textit{endlich} \text{ viele } n \\ &\Longleftrightarrow \quad \forall k \in \mathbb{N} \text{ gilt } |Y_n(\omega)| \geq \frac{1}{k} \text{ für } \textit{endlich} \text{ viele } n\,. \end{aligned}$$

Sei $E_k = \{\omega \in \Omega : |Y_n(\omega)| \geq \frac{1}{k}$ für unendlich viele $n\}$, k fest. Dann ist $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots$, und ferner $B_Y = \bigcap_{k\geq 1} E_k^c$. Somit schließen wir

$$\begin{aligned} P(B_Y) = 1 \quad &\Longleftrightarrow \quad P\left(\bigcap_{k\geq 1} E_k^c\right) = 1 \\ &\Longleftrightarrow \quad P\left((\bigcup_{k\geq 1} E_k)^c\right) = 1 \\ &\Longleftrightarrow \quad P\left(\bigcup_{k\geq 1} E_k\right) = 0\,. \end{aligned}$$

Insbesondere gilt nach Lemma 1.1

$$P(E_k) = 0 \ (\forall k) \Longrightarrow P(B_Y) = 1\,.$$

2. Sei $k \geq 1$ fest, $A_{n,k} = \{\omega \in \Omega : |Y_n(\omega)| \geq \frac{1}{k}\}$. Wir erhalten also eine Folge von Ereignissen $A_{1,k}, A_{2,k}, \ldots$ und setzen

$$A_k^\infty = \{\omega \in \Omega : \omega \text{ ist in unendlich vielen } A_{n,k}\} = E_k$$

aus 1). Aus $P(A_k^\infty) = 0$ $(\forall k)$ folgt also $P(B_Y) = 1$. Laut dem Lemma von Borel-Cantelli können wir festhalten

$$\sum_{n\geq 1} P(A_{n,k}) < \infty \ (\forall k) \Longrightarrow P(B_Y) = 1\,. \tag{1}$$

3. Sei wieder $k \geq 1$ fest. Wir setzen $Y_n = \bar{X}_n$, dann ist das starke Gesetz genau die Aussage $P(B_Y) = 1$. Wenn wir also zeigen können, dass $\sum_{n\geq 1} P(A_{n,k}) < \infty$ ist für alle k, so sind wir fertig. Nun verwenden wir die Ungleichung von Tschebyschev. Wir haben $A_{n,k} = \{\omega \in \Omega : |\bar{X}_n(\omega)| \geq \frac{1}{k}\}$, und nach Tschebyschev gilt mit $\mathrm{E}[\bar{X}_n] = 0$, $\mathrm{Var}[\bar{X}_n] = \frac{\mathrm{Var}[X]}{n}$

$$P(A_{n.k}) = P(|\bar{X}_n(\omega)| \geq \frac{1}{k}) \leq \frac{\mathrm{Var}[X]}{n} k^2\,,$$

und somit

$$\sum_{n\geq 1} P(A_{n,k}) \leq \mathrm{Var}[X] k^2 \sum_{n\geq 1} \frac{1}{n}\,.$$

Leider divergiert die harmonische Reihe, also müssen wir die Summe $\sum\limits_{n\geq 1} P(A_{n,k})$ besser abschätzen.

Wir betrachten die Teilfolge $\bar{X}_{n^2}$, $n = 1, 2, \ldots$. Für diese gilt nach Tschebyschev

$$\sum_{n=1}^\infty P(A_{n^2,k}) = \sum_{n=1}^\infty P(|\bar{X}_{n^2}| \geq \frac{1}{k}) \leq \mathrm{Var}[X] k^2 \sum \frac{1}{n^2} < \infty\,,$$

und somit nach (1)

$$P(\{\omega \in \Omega : \bar{X}_{n^2}(\omega) \to 0\}) = 1\,. \tag{2}$$

4. Sei $S_\ell = X_1 + \cdots + X_\ell$, und für $m \in \mathbb{N}$ sei $n = n(m)$ mit

$$n^2 \le m < (n+1)^2 \,.$$

Wir haben

$$E[S_m - S_{n^2}] = 0,\ \mathrm{Var}[S_m - S_{n^2}] = (m - n^2)\mathrm{Var}[X]\,.$$

Sei k fest, dann gilt wieder mit Tschebyschev

$$\begin{aligned} &P(\{\omega \in \Omega : |S_m(\omega) - S_{n^2}(\omega)| \ge \frac{n^2}{k}\}) = \\ &P(|S_m - S_{n^2}| \ge \frac{n^2}{k}) \le \frac{(m-n^2)k^2}{n^4}\mathrm{Var}[X]\,. \end{aligned}$$

Nun definieren wir die Ereignisse

$$B_{m,k} = \{\omega \in \Omega : \frac{|S_m(\omega) - S_{n^2}(\omega)|}{n^2} \ge \frac{1}{k}\}\,,\ n = n(m)\,,$$

und haben

$$P(B_{m,k}) \le \frac{m-n^2}{n^4}k^2\mathrm{Var}[X]\,,\ m \ge 1\,,\ n = n(m)\,.$$

Summation über m ergibt

$$\begin{aligned} \sum_{m \ge 1} P(B_{m,k}) &\le k^2\mathrm{Var}[X]\sum_{m=1}^{\infty}\frac{m - n(m)^2}{n(m)^4} \\ &= k^2\mathrm{Var}[X]\sum_{n=1}^{\infty}\sum_{m=n^2}^{(n+1)^2-1}\frac{m-n^2}{n^4} \\ &= k^2\mathrm{Var}[X]\sum_{n=1}^{\infty}\frac{1+2+\cdots+2n}{n^4} \\ &= k^2\mathrm{Var}[X]\sum_{n=1}^{\infty}\frac{2n(2n+1)}{2n^4} \\ &= k^2\mathrm{Var}[X]\sum_{n=1}^{\infty}\frac{2n+1}{n^3} < \infty\,. \end{aligned}$$

Mit $Y_m = \frac{S_m - S_{n(m)^2}}{n(m)^2}$ haben wir also nach (1)

$$P(\{\omega \in \Omega : \frac{S_m(\omega) - S_{n(m)^2}(\omega)}{n(m)^2} \stackrel{m\to\infty}{\longrightarrow} 0\}) = 1\,. \tag{3}$$

5. Nun setzen wir die Ergebnisse zusammen. Es sei

$$A = \{\omega : \bar{X}_{n^2}(\omega) \to 0\} \text{ mit } P(A) = 1 \text{ nach } (2)$$
$$B = \{\omega : \frac{S_m(\omega) - S_{n(m)^2}(\omega)}{n(m)^2} \to 0\} \text{ mit } P(B) = 1 \text{ nach } (3)\,,$$

also gilt $P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq 2 - 1 = 1$, somit $P(A \cap B) = 1$. Für ein $\omega \in A \cap B$ gilt $\bar{X}_{n^2}(\omega) \stackrel{n\to\infty}{\longrightarrow} 0$, also auch für die Teilfolge $\bar{X}_{n(m)^2}(\omega) \stackrel{m\to\infty}{\longrightarrow} 0$. Ferner ist

$$\frac{S_m}{n(m)^2} = \frac{S_m - S_{n(m)^2}}{n(m)^2} + \frac{S_{n(m)^2}}{n(m)^2} = \frac{S_m - S_{n(m)^2}}{n(m)^2} + \bar{X}_{n(m)^2}\,.$$

Da $\omega \in B$ ist, gilt $\frac{S_m - S_{n(m)^2}}{n(m)^2}(\omega) \longrightarrow 0$ mit $m \to \infty$, und wir erhalten als Ergebnis:

$$\omega \in A \cap B \Longrightarrow \frac{S_m(\omega)}{n(m)^2} \stackrel{m\to\infty}{\longrightarrow} 0\,.$$

Wegen $m \geq n(m)^2$ gilt ferner für $\omega \in A \cap B$

$$|\bar{X}_m(\omega)| = \left|\frac{S_m(\omega)}{m}\right| \leq \left|\frac{S_m(\omega)}{n(m)^2}\right| \stackrel{m\to\infty}{\longrightarrow} 0\,,$$

also

$$\omega \in A \cap B \Longrightarrow \bar{X}_m(\omega) \stackrel{m\to\infty}{\longrightarrow} 0$$

oder

$$A \cap B \subseteq \{\omega : \bar{X}_m(\omega) \stackrel{m\to\infty}{\longrightarrow} 0\}\,.$$

Da wie gesehen $P(A \cap B) = 1$ ist, erhalten wir daraus

$$P(\{\omega \in \Omega : \bar{X}_m(\omega) \stackrel{m\to\infty}{\longrightarrow} 0\}) = 1\,,$$

und der Beweis ist fertig. □

Beispiel. Betrachten wir das Intervall $\Omega = (0, 1] \subseteq \mathbb{R}$. Eine Zahl $\omega = 0, y_1 y_2 y_3 \ldots$ heißt *normal*, falls jeder mögliche Block $a_1 a_2 \ldots a_k \in \{0, 1, \ldots, 9\}^k$

der Länge k in der Dezimalentwicklung im Limes mit relativer Häufigkeit 10^{-k} vorkommt, und zwar für alle $k \geq 1$. Also, wenn

$$\lim_{n\to\infty} \frac{\#\{a_1 \dots a_k \text{ in } y_1 \dots y_n\}}{n} = \frac{1}{10^k}\,,$$

gilt. Gibt es überhaupt normale Zahlen? Eine besonders schöne Anwendung des starken Gesetzes der großen Zahlen zeigt, dass bei Gleichverteilung auf Ω gilt: $P(\{\omega \in \Omega : \omega \text{ normal}\}) = 1$. Es gibt also überabzählbar viele normale Zahlen.

Wir beweisen die Aussage für $k = 1$, der allgemeine Fall ist nur wenig schwieriger.

Es sei $\omega = 0, y_1(\omega)y_2(\omega)\dots$ die Dezimalentwicklung und $Y_i(\omega) = y_i(\omega)$ die Zufallsvariable (unter Gleichverteilung), $i \geq 1$. wir haben

$$P(Y_1 = b_1 \wedge \dots \wedge Y_m = b_m) = \frac{1}{10^m}\,,$$

da $\{\omega : y_1(\omega) = b_1 \wedge \dots \wedge y_m(\omega) = b_m\}$ ein Intervall der Länge $\frac{1}{10^m}$ ist. Ebenso gilt $P(Y_i = b_i) = \frac{1}{10}$ für alle i, also sind die Variablen Y_i unabhängig.

Nun sei $a \in \{0, 1, \dots, 9\}$ fest vorgegeben, und $X_1, X_2, \dots$ die Folge von Bernoulli Variablen mit

$$X_i(\omega) = \begin{cases} 1 & \text{falls } y_i(\omega) = a \\ 0 & \text{sonst}\,. \end{cases}$$

Wie eben gesehen ist $P(X_i = 1) = \frac{1}{10}$ für alle i, die Variablen X_i sind somit unabhängige Kopien der Variablen X mit $E[X] = \frac{1}{10}$. Für $\omega \in \Omega$ ist

$$\overline{X}_n(\omega) = \frac{\#\{i : y_i = a, 1 \leq i \leq n\}}{n} = \text{relative Häufigkeit von } a\,,$$

und das starke Gesetz besagt

$$P(\{\omega : \text{ relative Hüfigkeit von } a \to \frac{1}{10}\}) = 1$$

für alle $a \in \{0, 1, \dots, 9\}$.

3.5 Der zentrale Grenzwertsatz

Der folgende Satz ist der berühmteste Satz der W-Theorie und zeigt die in der Praxis beobachtete zentrale Stellung der Gaußschen Glockenkurve.

Satz 3.7 (Zentraler Grenzwertsatz). *Seien $X_1, X_2, \ldots$ unabhängige Kopien von X mit Erwartungswert $E[X]$ und Varianz $\mathrm{Var}[X]$. Dann gilt*

$$\frac{X_1 + \cdots + X_n - n\mathrm{E}[X]}{\sqrt{n}\sqrt{\mathrm{Var}[X]}} \xrightarrow{i.V.} Y \sim N(0,1),$$

das heißt Y ist nach der Standardnormalverteilung verteilt.

Beispiele. 1. Sei X gleichverteilt auf $[0,1]$, $\mathrm{E}[X] = \frac{1}{2}$, $\mathrm{Var}[X] = \frac{1}{12}$. Dann gilt

$$\frac{X_1 + \cdots + X_n - \frac{n}{2}}{\sqrt{n}\sqrt{\frac{1}{12}}} \xrightarrow{i.V.} N(0,1)\,.$$

2. Sei X Bernoulli Variable mit $P(X=1) = p$, $P(X=0) = q = 1-p$, $0 < p < 1$. Dann ist $E[X] = p$, $\mathrm{Var}[X] = pq$, und es folgt

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{npq}} \xrightarrow{i.V.} N(0,1)\,. \tag{1}$$

Damit können wir die Standardnormalverteilung simulieren. Wir werfen eine Münze n Mal, $p = \frac{1}{2}$, dann ist

$$\frac{\text{Anzahl Kopf} - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \xrightarrow{i.V.} N(0,1)\,.$$

Wir werden den zentralen Grenzwertsatz für Bernoulli Variablen beweisen, also die Aussage (1). Dieser Spezialfall heißt Satz von de Moivre-Laplace. Dazu benötigen wir einige Vorbereitungen.

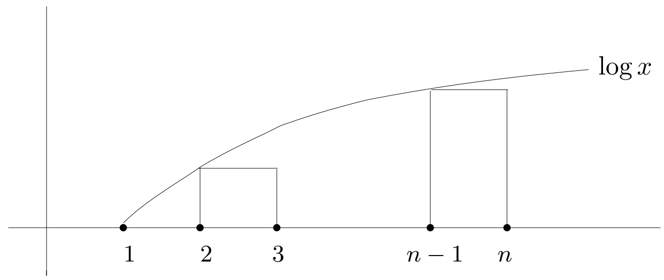
Lemma 3.8 (Stirling Formel). *Wir haben*

$$n! \sim \sqrt{2\pi n}(\frac{n}{e})^n\,,$$

das heißt

$$\lim_{n\to\infty} \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{n!} = 1\,.$$

Wir können uns die Formel plausibel machen. Betrachten wir die Logarithmuskurve:



Durch Ober- und Untersummenbildung erhalten wir

$$\log(n-1)! < \int\limits_1^n \log x dx < \log(n!)$$

und wegen $\int\limits_1^n \log x dx = x \log x - x$

$$\begin{aligned} \log(n-1)! <\ & n \log n - n + 1 < \log(n!) \\ (n-1)! <\ & e(\frac{n}{e})^n < n!\,. \end{aligned}$$

Für $n, k,\ n-k$ groß können wir also abschätzen:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \sim \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} (\frac{n}{k})^k (\frac{n}{n-k})^{n-k}\,. \tag{2}$$

Als nächsten wollen wir den Koeffizienten $b(k, n; p)$ abschätzen. Wir gehen in mehreren Schritten vor.

1. Für $b(k, n; p)$ schreiben wir $b_{n,p}(k) = \binom{n}{k} p^k q^{n-k}$, $q = 1-p$. Nun betrachten wir Folgen (k_n) mit $\frac{k_n}{n} \to p$. Es muß also auch $k_n \to \infty$ gelten, $\frac{n-k_n}{n} = 1 - \frac{k_n}{n} \to 1 - p = q$, also auch $n - k_n \to \infty$.

Sei $k = k_n$, dann haben wir mit (2)

$$b_{n,p}(k) \sim \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} (\frac{pn}{k})^k (\frac{qn}{n-k})^{n-k}\,, \tag{3}$$

und mit $\sigma_n = \sqrt{npq}$, $\frac{k}{n} \sim p$, $\frac{n-k}{n} \sim q$

$$\sqrt{\frac{n}{k(n-k)}} \sim \sqrt{\frac{1}{npq}} = \frac{1}{\sigma_n}\,. \tag{4}$$

2. Als nächstes untersuchen wir das Grenzwertverhalten von

$$f(n,k) = (\frac{pn}{k})^k (\frac{qn}{n-k})^{n-k}\,.$$

Sei $t_n = \frac{k_n}{n}$, also $t_n \to p$, $\frac{n-k_n}{n} = 1 - \frac{k_n}{n} = 1 - t_n \to q$. Wir setzen kurz $t = \frac{k}{n}$. Dann gilt

$$\begin{aligned} \log f(n,k) &= k \log \frac{p}{t} + (n-k) \log \frac{q}{1-t}\,, \\ -\log f(n,k) &= k \log \frac{t}{p} + (n-k) \log \frac{1-t}{q} \\ &= n \underbrace{(t \log \frac{t}{p} + (1-t) \log \frac{1-t}{q})}_{g(t)}\,. \end{aligned}$$

Für die Funktion $g(t)$ sehen wir:
a. $g(p) = 0$,

b. $\begin{aligned} g'(t) &= \log \tfrac{t}{p} + 1 - \log \tfrac{1-t}{q} + (1-t)\tfrac{q}{1-t}(-\tfrac{1}{q}) \\ &= \log \tfrac{t}{p} - \log \tfrac{1-t}{q}\,, \end{aligned}$

also insbesondere $g'(p) = 0$.

c. $g''(t) = \frac{1}{t} + \frac{1}{1-t}$, $g''(p) = \frac{1}{p} + \frac{1}{q} = \frac{1}{pq}$.

Taylorentwicklung in p ergibt

$$g(t) = \frac{1}{2pq}(t-p)^2 + h(t-p) \text{ mit } |h(t-p)| \leq c|t-p|^3\,.$$

3. Angenommen, es gilt sogar $n(t-p)^3 \to 0$. Dann ist $nh(t-p) \to 0$, und somit

$$|-\log f(n,k) - \frac{n(t-p)^2}{2pq}| = |ng(t) - n\frac{(t-p)^2}{2pq}| = |nh(t-p)| \to 0\,.$$

Wir setzen nun

$$x(n,k) = \frac{k-np}{\sigma_n}\,, \tag{5}$$

dann haben wir

$$\frac{n(t-p)^2}{2pq} = \frac{n(\frac{k}{n}-p)^2}{2pq} = \frac{(k-np)^2}{2npq} = \frac{x(n,k)^2}{2}\,,$$

und es folgt

$$|-\log f(n,k) - \frac{x(n,k)^2}{2}| \longrightarrow 0\,.$$

Daraus folgt

$$e^{-\log f(n,k) - \frac{x(n,k)^2}{2}} \to 1\,,$$

$$\frac{e^{-\frac{x(n,k)^2}{2}}}{f(n,k)} \to 1\,,$$

also

$$f(n,k) \sim e^{-\frac{x(n,k)^2}{2}}\,. \tag{6}$$

4. Was bedeutet $n(t-p)^3 \to 0$ für die Zahlen $x(n,k)$? Wir haben

$$n(t-p)^3 = n(\frac{k}{n}-p)^3 = \frac{(k-np)^3}{n^2} = \frac{x(n,k)^3\sigma_n^3}{n^2}$$
$$= \frac{x(n,k)^3(pq)^{\frac{3}{2}}}{\sqrt{n}}\,,$$

und sehen

$$n(t-p)^3 \to 0 \Longleftrightarrow \frac{x(n,k_n)^3}{\sqrt{n}} \to 0\,.$$

5. Fassen wir zusammen: Sei $x(n,k) = \frac{k_n - np}{\sigma_n}$. Falls $\frac{x(n,k_n)^3}{\sqrt{n}} \to 0$ geht, so ist nach (3), (4) und (6)

$$b_{n,p}(k_n) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_n}e^{-\frac{x(n,k)^2}{2}}\,,$$

und wir erhalten das folgende Resultat.

Lemma 3.9. *Sind* (α_n), (β_n) *Folgen mit* $\frac{x(n,\alpha_n)^3}{\sqrt{n}}, \frac{x(n,\beta_n)^3}{\sqrt{n}} \to 0$, *so ist*

$$b_{n,p}(k_n) \sim \frac{e^{-\frac{x(n,k_n)^2}{2}}}{\sqrt{2\pi}\sigma_n} \quad \textit{für } \alpha_n \le k_n \le \beta_n.$$

Satz 3.10 (Satz von de Moivre Laplace)**.** *Es sei X eine Bernoulli Variable mit $0 < p < 1$, und $X_1, X_2, \dots,$ unabhängige Kopien von X. Dann gilt für die zentrierte Variable*

$$S_n^* = \frac{X_1 + \cdots + X_n - np}{\sqrt{npq}}, \; q = 1 - p\,,$$

$$S_n^* \xrightarrow{i.V.} N(0,1)\,,$$

das heißt

$$\lim_{n\to\infty} P(a \le S_n^* \le b) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{x^2}{2}} dx \quad (a, b \in \mathbb{R})\,.$$

Beweis. Wir teilen den Beweis wieder in mehrere Schritte ein.
1. Sei $S_n = X_1 + \cdots + X_n$. Dann gilt mit $\sigma_n = \sqrt{npq}$

$$a \le S_n^* \le b \Longleftrightarrow a \le \frac{S_n - np}{\sigma_n} \le b \Longleftrightarrow a\sigma_n + np \le S_n \le b\sigma_n + np\,.$$

Sei $\alpha_n \in \mathbb{N}$ minimal mit $\alpha_n \ge a\sigma_n + np$ und $\beta_n \in \mathbb{N}$ maximal mit $\beta_n \le b\sigma_n + np$. Dann gilt

$$a \le S_n^* \le b \Longleftrightarrow \alpha_n \le S_n \le \beta_n\,,$$

somit

$$P(a \le S_n^* \le b) = P(\alpha_n \le S_n \le \beta_n) = \sum_{k=\alpha_n}^{\beta_n} b_{n,p}(k)\,.$$

Wir müssen also zeigen:

$$\lim_{n\to\infty} \sum_{k=\alpha_n}^{\beta_n} b_{n,p}(k) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{x^2}{2}} dx\,. \tag{7}$$

2. Aus $\beta_n \le b\sigma_n + np < \beta_n + 1$ folgt

$$b\sigma_n + np - 1 < \beta_n \le b\sigma_n + np\,,$$

also gilt $\frac{\beta_n}{n} \to p$ und analog $\frac{\alpha_n}{n} \to p$. Mit $x(n,\beta_n) = \frac{\beta_n - np}{\sigma_n}$ haben wir

$$b - \frac{1}{\sigma_n} < x(n,\beta_n) \le b$$

und analog

$$a \le x(n,\alpha_n) < a + \frac{1}{\sigma_n}\,.$$

Die Folgen $(x(n,\beta_n))$ und $(x(n,\alpha_n))$ sind somit beschränkt und es folgt

$$\frac{x(n,\beta_n)^3}{\sqrt{n}} \to 0,\ \ \frac{x(n,\alpha_n)^3}{\sqrt{n}} \to 0\,.$$

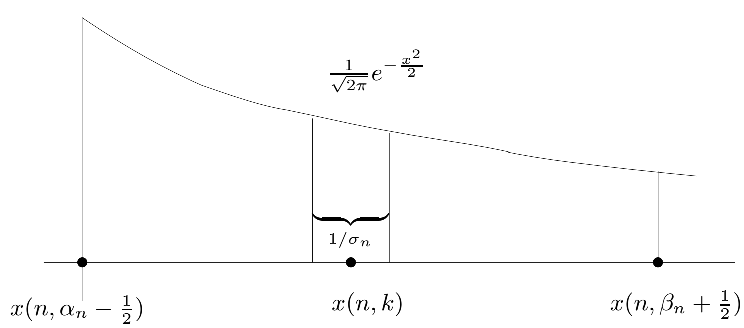
Aus Lemma 3.9 folgt daher

$$b_{n,p}(k) \sim \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{x(n,k)^2}{2}} \text{ für alle } \alpha_n \le k \le \beta_n\,.$$

und somit

$$\sum_{k=\alpha_n}^{\beta_n} b_{n,p}(k) \sim \frac{1}{\sqrt{2\pi}\sigma_n} \sum_{k=\alpha_n}^{\beta_n} e^{-\frac{x(n,k)^2}{2}}\,. \tag{8}$$

3. Nun sehen wir uns die rechte Summe in (8) näher an.



Die rechte Summe ist eine Riemannsche Summe mit Intervallen der Länge $\frac{1}{\sigma_n}$, und wir haben

$$x(n,\alpha_n - \frac{1}{2}) = \frac{\alpha_n - \frac{1}{2} - np}{\sigma_n} \to a,\ \ x(n,\beta_n + \frac{1}{2}) = \frac{\beta_n + \frac{1}{2} - np}{\sigma_n} \to b\,.$$

Es folgt

$$\sum_{k=\alpha_n}^{\beta_n} b_{n,p}(k) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx\,,$$

und der Satz ist bewiesen. $\square$

Folgerung 3.11. *Seien $X_1, X_2, \ldots$ unabhängige Kopien der Bernoulli Variablen X mit $0 < p < 1$, $\sigma_n = \sqrt{npq}$. Dann gilt für $S_n = X_1 + \cdots + X_n$ (Anzahl der Erfolge)*

$$\lim_{n\to\infty} P(\alpha \leq S_n \leq \beta) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-np}{\sigma_n}}^{\frac{\beta-np}{\sigma_n}} e^{-\frac{x^2}{2}} dx = \phi(\frac{\beta - np}{\sigma_n}) - \phi(\frac{\alpha - np}{\sigma_n})\,,$$

mit $\phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.$

Das letzte Resultat wird in der Praxis zur Abschätzung der Erfolgswahrscheinlichkeit verwendet.

Beispiel. Ein Würfel mit Gleichverteilung wird 600 Mal geworfen. Wie groß ist die W-keit, zwischen 90 und 100 6en zu erhalten? Hier ist $n = 600$, $p = \frac{1}{6}$, $np = 100$, $\sigma_n = \sqrt{600 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 9,13$, $\alpha - np = -10$, $\beta - np = 0$, also

$$P(90 \leq S_n \leq 100) \sim \phi(0) - \phi(-\frac{10}{9,13}) \sim 0,36\,.$$

4 Parametrische statistische Methoden

4.1 Datenniveau

Jede statistische Arbeit beginnt mit der Erhebung der Daten. Es ist heute üblich, die Daten in vier Gruppen einzuteilen.

1. Nominale Skala: Zum Beispiel Blutgruppe, Geschlecht, Partei, Postleitzahl.
2. Ordinale Skala: Noten, Sportranglisten, Windstärke.
3. Intervallskala: Temperatur, Zeitdauer.
4. Verhältnisskala: Länge, Gewicht, Fläche, Kosten.

Nominale und ordinale Daten gehören zum Bereich der *nichtparametrischen* Statistik, die wir im nächsten Kapitel behandeln, Intervall- und Verhältnisskala zur *parametrischen* Statistik.

Die parametrische Statistik stellt sich die Aufgabe, aus einer vorgegebenen Familie von möglichen Verteilungen Parameter zu bestimmen oder zumindest zu schätzen.

Die nichtparametrische Statistik versucht, aus den Daten allgemeine Strukturaussagen zu erzielen.

Von Skala 1 zu 4 erhalten wir einen Informationsgewinn, andererseits werden die Meßungenauigkeiten größer. Zum Beispiel ist die richtige Blutgruppe leicht anzugeben, während die exakte Flächenmessung deutlich schwieriger ist.

Die Datenerhebung erfolgt üblicherweise durch

- Befragung
- Beobachtung
- Experiment.

4.2 Ansatz der Statistik

Wir beginnen mit einem typischen Beispiel, der Qualitätskontrolle.

Ein Importeur erhält eine Lieferung von $N = 10.000$ Orangen. Er möchte schätzen, wieviele, sagen wir w, davon faul sind. Er macht eine Stichprobe von $n = 50$ Orangen und stellt fest, dass x davon faul sind.

1. Idee: Er sagt sich, $\frac{x}{n} \sim \frac{w}{N}$, also wird er als *Schätzer* von w den Ausdruck $T(x) = \frac{N}{n}x$ verwenden. $T(x)$ ist vom Zufall (der ausgesuchten Orangen) abhängig, also eine *Zufallsvariable* $T : \mathcal{X} \to \mathbb{R}$, wobei $\mathcal{X} = \{0, 1, 2, \ldots, n\}$ der *Stichprobenraum* ist.

2. Idee: Der Importeur sucht ein Intervall $C(x)$, in dem die richtige Anzahl w an faulen Orangen mit genügend hoher W-keit liegt. $C(x)$ heißt ein *Konfidenzintervall* und hängt von der beobachteten Größe x ab. $C(x)$ darf *nicht* von w abhängen, da w ja nicht bekannt ist. Die Forderung ist also

$$P_w(\{x \in \mathcal{X} : w \in C(x)\}) \geq 1 - \alpha \text{ für } \textit{alle } w \in \{0, 1, \ldots, N\}\,,$$

wobei P_w das Wahrscheinlichkeitsmaß ist, falls w die richtige Anzahl ist. Die Zahl α heißt *Konfidenzniveau* und wird üblicherweise als $\alpha = 0,05$ oder $0,025$ angenommen.

3. Idee: Angenommen, der Importeur muß den Preis nur zahlen, wenn höchstens 5% der Orangen faul sind. Er stellt folgende Hypothesen auf:

$$\begin{aligned} H_0 &: \; w \in \{0, 1, \ldots, 500\} \\ H_1 &: \; w \in \{501, \ldots, 10.000\}\,. \end{aligned}$$

H_0 heißt die *Nullhypothese*, H_1 die *Alternativhypothese*. Nun entwirft er ein *Entscheidungsverfahren*, um festzustellen, welche Hypothese zutrifft. Zum Beispiel bestimmt er eine Zahl c mit

$$\begin{aligned} x \leq c &\implies H_0 \;\; \text{(Importeur akzeptiert)}, \\ x > c &\implies H_1 \;\; \text{(Importeur lehnt ab)}. \end{aligned}$$

Die kritische Zahl c soll so bestimmt werden, dass

(I) $P_w(x > c)$ klein ist für $w \leq 500$

(II) $P_w(x > c)$ groß ist für $w > 500$.

Bei (I) sprechen wir von einem *Irrtum* 1.*Art*. Es soll vermieden werden, dass der Importeur die Sendung ablehnt, obwohl sie in Ordnung ist.

Bei (II) sprechen wir von einem *Irrtum* 2.*Art*. Falls die Sendung nicht in Ordnung ist, so soll er sie nur mit kleiner W-keit akzeptieren.

In der Praxis ist H_0 oft die konservative Hypothese und der Irrtum 1.Art wichtiger. Soll zum Beispiel ein neues Medikament gegen ein am Markt bewährtes getestet werden, so wird als Nullhypothese H_0 angenommen, dass kein signifikanter Unterschied besteht, während die Alternativhypothese H_1 sagt, dass ein Unterschied besteht (zweiseitiger Test) oder dass das neue Medikament signifikant besser ist (einseitiger Test). Der Irrtum 1.Art besagt: Das neue Medikament wird eingeführt, obwohl es nichts bringt.

Wir haben also bei parametrischen Verfahren die folgende Struktur:

1. Die Beobachtungsergebnisse x bilden den Stichprobenraum $\mathcal{X}$, der mit einer σ-algebra $\mathcal{E}$ versehen wird.

2. Die Menge der möglichen Verteilungen ist $\{P_\vartheta : \vartheta \in \theta\}$, wobei θ der Parameterraum ist. Ein *parametrisches Modell* ist dann

$$\mathcal{M} = \{\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta\},\ \theta \subseteq \mathbb{R} \text{ oder } \theta \subseteq \mathbb{R}^n\,.$$

Beispiel. Unser Orangenbeispiel entspricht offenbar den hypergeometrischen Verteilungen

$$\mathcal{X} = \{0, 1, \ldots, n\},\ \mathcal{E} = 2^{\mathcal{X}},\ \theta = \{0, 1, \ldots, N\}, \vartheta = \#\text{ faule Orangen;}$$

$$P_\vartheta(X = x) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}}\,.$$

Machen wir n unabhängige Stichproben, so haben wir das Modell

$$(\mathcal{X}^n, \mathcal{E}^{\otimes n}, P_\vartheta^n : \vartheta \in \theta)\,.$$

Wir halten fest:

$$\begin{aligned} X &= \text{ Zufallsvariable der Stichprobe,} \\ x &= \text{ beobachteter Wert.} \end{aligned}$$

4.3 Parameterschätzung

Es sei das Modell $\mathcal{M} = (\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ gegeben, und $(W, \mathcal{F})$ sei eine σ-Algebra auf W. Sei $\tau : \theta \to W$ eine Abbildung, die jedem $\vartheta \in \theta$ einen gewissen Parameter $\tau(\vartheta) \in W$ zuordnet.

Zu einer Beobachtung $x \in \mathcal{X}$ wollen wir $T(x) \in W$ angeben, das den Parameter $\tau(\vartheta)$ schätzt. $T(x)$ heißt *Schätzer* von $\tau(\vartheta)$, und die Zufallsvariable $T : \mathcal{X} \to W$ heißt *Statistik*. Meist wird $\tau(\vartheta) = \vartheta$ sein.

Es seien $X_1, \ldots, X_n$ unabhängige identisch verteilte Stichprobenvariablen, $T_n = T(X_1, \ldots, X_n)$. T_n ist also eine Zufallsvariable

$$T_n : \mathcal{X}^n \to W\,.$$

Wir stellen folgende plausible Forderungen, wobei E_ϑ, Var_ϑ Erwartungswert und Varianz bedeuten, falls ϑ der richtige Parameter ist.

1. T_n heißt *erwartungstreu*, falls

$$E_\vartheta[T_n] = \tau(\vartheta) \text{ für alle } \vartheta \in \theta\,.$$

2. Etwas schwächer ist: T_n ist *asymptotisch erwartungstreu*, falls

$$\lim_{n\to\infty} E_\vartheta[T_n] = \tau(\vartheta) \text{ für alle } \vartheta \in \theta\,.$$

3. T_n heißt *konsistent*, wenn $T_n \xrightarrow{i.W.} \tau(\vartheta)$ für alle ϑ gilt, das heißt

$$P_\vartheta(|T_n - \tau(\vartheta)| \geq \varepsilon) \overset{n\to\infty}{\longrightarrow} 0 \text{ für alle } \varepsilon > 0\,.$$

Definition. T_n heißt *bester Schätzer*, falls gilt:

a. $E_\vartheta[T_n] = \tau(\vartheta)$ (T_n ist erwartungstreu),

b. für alle erwartungstreuen Schätzer U_n gilt

$$\mathrm{Var}_\vartheta[T_n] \leq \mathrm{Var}_\vartheta[U_n]\,.$$

Beispiel. Das Intervall $[0, \vartheta]$ sei gegeben, wobei $\vartheta > 0$ unbekannt ist. Es werden n unabhängige gleichverteilte Zufallszahlen aus $[0, \vartheta]$ gezogen. Es soll ϑ geschätzt werden. Hier ist

1. $\mathcal{X} = [0, \infty)$, $\mathcal{X}^n = [0, \infty)^n$, $\tau(\vartheta) = \vartheta$.

2. Gleichverteilung bedeutet für die Dichte $f_\vartheta(x) = \frac{1}{\vartheta}$ für alle $0 \le x \le \vartheta$, also $f_\vartheta(x_1, \ldots, x_n) = \frac{1}{\vartheta^n}$ für alle $(x_1, \ldots, x_n)$. Wir haben

$$E_\vartheta[X] = \frac{\vartheta}{2}, \mathrm{Var}_\vartheta[X] = \frac{\vartheta^2}{12}\,.$$

Idee I. Zu $X_1, \ldots, X_n$ nehmen wir als Schätzer

$$T_n(X_1, \ldots, X_n) = \frac{2}{n}(X_1 + \cdots + X_n)\,,$$

da $2\frac{X_1+\cdots+X_n}{n} \approx \vartheta$ zu erwarten ist.

Wir haben

a. $E_\vartheta(T_n) = \frac{2}{n} n E_\vartheta[X] = 2\frac{\vartheta}{2} = \vartheta$, also ist T_n erwartungstreu.

b. Nach dem schwachen Gesetz für große Zahlen gilt $T_n = 2\frac{X_1+\cdots+X_n}{n} \xrightarrow{i.W.} 2E_\vartheta[X] = \vartheta$, also ist T_n konsistent.

c. $\mathrm{Var}_\vartheta[T_n] = (\frac{2}{n})^2 n \mathrm{Var}_\vartheta[X] = \frac{4}{n^2} n \frac{\vartheta^2}{12} = \frac{\vartheta^2}{3n}$.

Idee II. Wir nehmen als Schätzer

$$\widetilde{T}_n(X_1, \ldots, X_n) = \max(X_1, \ldots, X_n)\,.$$

a. Wegen $\widetilde{T}_n \le \vartheta$ ist $\widetilde{T}_n$ sicher nicht erwartungstreu. Es ist aber

$$P_\vartheta(\widetilde{T}_n \le x) = P_\vartheta(X_1 \le x \wedge \ldots \wedge X_n \le x) = \prod_{i=1}^{n} P_\vartheta(X_i \le x) = (\frac{x}{\vartheta})^n \ (x \in [0, \vartheta])\,,$$

also ist die Dichte $\widetilde{f}(x) = \frac{nx^{n-1}}{\vartheta^n}$, und wir erhalten

$$E_\vartheta[\widetilde{T}_n] = \int_0^\vartheta x \frac{nx^{n-1}}{\vartheta^n} dx = \frac{nx^{n+1}}{(n+1)\vartheta^n}\Big|_0^\vartheta = \frac{n}{n+1}\vartheta \stackrel{n\to\infty}{\longrightarrow} \vartheta\,,$$

das heißt, $\widetilde{T}_n$ ist asymptotisch erwartungstreu.

b. Für die Varianz haben wir

$$E_\theta[\widetilde{T}_n^2] = \int_0^\vartheta x^2 \frac{nx^{n-1}}{\vartheta^n} dx = \frac{nx^{n+2}}{(n+2)\vartheta^n}\Big|_0^\vartheta = \frac{n}{n+2}\vartheta^2\,,$$

also
$$\mathrm{Var}_\vartheta[\widetilde{T}_n] = \frac{n}{n+2}\vartheta^2 - (\frac{n}{n+1}\vartheta)^2 = \frac{n}{(n+1)^2(n+2)}\vartheta^2 .$$

c. Nun nehmen wir die Modifikation $T_n^* = \frac{n+1}{n}\widetilde{T}_n$. Dann ist $E_\vartheta[T_n^*] = \frac{n+1}{n}E_\vartheta[\widetilde{T}_n]$ $= \vartheta$, also ist T_n^* erwartungstreu. Für die Varianz gilt

$$\mathrm{Var}[T_n^*] = \frac{(n+1)^2}{n^2}\mathrm{Var}[\widetilde{T}_n] = \frac{\vartheta^2}{n(n+2)} ,$$

also $\mathrm{Var}_\vartheta[T_n^*] < \mathrm{Var}_\vartheta[T_n]$ für $n \geq 2$ und alle $\vartheta \in \theta$.
T_n^* ist also ein besserer Schätzer als T_n.

Die nächste Idee führt zum wichtigsten allgemeinen Schätzer, dem Maximum Likelihood Schätzer. Wenn wir x beobachten, so ist im diskreten Fall $P_\vartheta(X = x)$ die Wahrscheinlichkeit, dass x eintritt, falls ϑ der richtige Parameter ist. Ein ϑ mit kleiner W-keit $P_\vartheta(X = x)$ wird also nicht der richtige Parameter sein.

Schätzregel. Man bestimme $T(x)$ zu x so, dass

$$P_{T(x)}(X = x) = \max_{\vartheta\in\theta} P_\vartheta(X = x) .$$

$T(x)$ heißt *Maximum Likelihood Schätzer*, kurz ML-Schätzer.

Beispiel. Analysieren wir das Beispiel mit der Orangenlieferung. Hier ist

$$P_\vartheta(X = x) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}} = \max$$

zu bilden.

Wir haben

$$\frac{P_\vartheta(X = x)}{P_{\vartheta-1}(X = x)} = \frac{\binom{\vartheta}{x}}{\binom{\vartheta-1}{x}}\frac{\binom{N-\vartheta}{n-x}}{\binom{N-\vartheta+1}{n-x}} = \frac{\vartheta}{\vartheta - x}\,\frac{N-\vartheta-n+x+1}{N-\vartheta+1} \geq 1$$

für
$$\vartheta N - \vartheta^2 - \vartheta n + \vartheta x + \vartheta \geq \vartheta N - \vartheta^2 + \vartheta - Nx + \vartheta x - x$$

also für
$$-\vartheta n \geq -Nx - x$$

das heißt für

$$\vartheta \le \frac{x(N+1)}{n} \, .$$

Der ML-Schätzer ist daher

$$T(x) = \lfloor \frac{x(N+1)}{n} \rfloor \sim \frac{xN}{n} \, ,$$

er entspricht also bis auf Rundung dem naiven Schätzer $\frac{xN}{n}$.

Allgemein gehen wir folgendermaßen vor. Für diskrete Zufallsvariablen wollen wir bei gegebenem x die Funktion $P_\vartheta(X = x)$ in ϑ maximieren, und für stetige Variable die Dichte $f_\vartheta(x)$. Die *Likelihood Funktion* ist

$$\rho : \mathcal{X} \times \theta \to [0, \infty) \text{ mit } \rho(x, \vartheta) = \begin{cases} P_\vartheta(X = x) & X \text{ diskret} \\ f_\vartheta(x) & X \text{ stetig.} \end{cases}$$

Definition. $T(x)$ heißt *ML-Schätzer*, falls

$$\rho(x, T(x)) = \max_{\vartheta \in \theta} \rho(x, \vartheta).$$

Beispiele. 1. Ein Bernoulli Experiment mit $P(X = 1) = \vartheta$, $P(X = 0) = 1 - \vartheta$, $0 < \vartheta < 1$, wird n Mal wiederholt. Gesucht ist ein Schätzer für ϑ. Sei x die Anzahl der Erfolge. Wir haben

$$\mathcal{X} = \{0, 1, \dots, n\}, \ \theta = (0, 1), \ P_\vartheta(X = x) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}.$$

Um $\max\limits_{\vartheta \in \theta} P_\vartheta(X = x)$ zu berechnen, maximieren wir den Logarithmus und berechnen

$$\begin{aligned} \frac{d}{d\vartheta} \log P_\vartheta(X = x) &= \frac{d}{d\vartheta} \left[\log \binom{n}{x} + x \log \vartheta + (n - x) \log(1 - \vartheta) \right] \\ &= \frac{x}{\vartheta} - \frac{n - x}{1 - \vartheta} \, , \end{aligned}$$

und diese Funktion ist monoton fallend in ϑ. Das Maximum ergibt sich also für $\frac{d}{d\vartheta} \log P_\vartheta(X = x) = 0$, und wir erhalten

$$\frac{x}{\vartheta} = \frac{n - x}{1 - \vartheta} \Leftrightarrow x - x\vartheta = n\vartheta - x\vartheta \Leftrightarrow \vartheta = \frac{x}{n} \, .$$

Der naive Schätzer $T_n(x) = \frac{x}{n}$ ist also auch der ML-Schätzer. Wir haben ferner

$$\begin{aligned} E_\vartheta(T_n) &= E_\vartheta(\frac{X_1 + \cdots + X_n)}{n}) = \vartheta \\ \mathrm{Var}_\vartheta(T_n) &= \mathrm{Var}_\vartheta(\frac{X_1 + \cdots + X_n}{n}) = \frac{\vartheta(1-\vartheta)}{n}\,. \end{aligned}$$

2. Betrachten wir nochmals das Beispiel der Ziehung von n Zahlen aus $[0, \vartheta]$, $\vartheta > 0$ mit Gleichverteilung. Hier ist

$$f_\vartheta(x_1, \ldots, x_n) = \begin{cases} \frac{1}{\vartheta^n} & \text{falls } x_1, \ldots, x_n \leq \vartheta \\ 0 & \text{sonst}\,. \end{cases}$$

Der ML-Schätzer ist daher $\widetilde{T}_n(x) = \max(x_1, \ldots, x_n)$, da $\vartheta \geq x_1, \ldots, x_n$ möglichst klein sein soll, um $f_\vartheta(x_1, \ldots, x_n)$ zu maximieren.

3. Ausfallswahrscheinlichkeit von Geräten
Es wird angenommen, dass die Lebensdauer von Glühbirnen exponential verteilt ist mit $f_\vartheta(x) = \vartheta e^{-\vartheta x}$, ϑ unbekannt. Es werden n Stichproben gezogen. Wir haben also

$$\mathcal{X} = [0, \infty),\ f_\vartheta(x) = \vartheta e^{-\vartheta x}, f_\vartheta(x_1, \ldots, x_n) = \vartheta^n e^{-\vartheta(x_1 + \cdots + x_n)}\,.$$

Um das Maximum zu berechnen, logarithmieren wir wie eben und erhalten

$$\begin{aligned} \frac{d}{d\vartheta} \log f_\vartheta(x_1, \ldots, x_n) &= \frac{d}{d\vartheta}(n \log \vartheta - \vartheta(x_1 + \cdots + x_n)) \\ &= \frac{n}{\vartheta} - (x_1 + \cdots + x_n)\,. \end{aligned}$$

Der ML-Schätzer ist daher

$$T_n(x_1, \ldots, x_n) = \frac{n}{x_1 + \cdots + x_n} = \frac{1}{\bar{x}}\,,$$

wobei $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$ der Durchschnittswert ist.
Ist $\tau(\vartheta)$ die W-keit für den Ausfall der Glühbirne bis zur Zeit t, so haben wir

$$P_\vartheta(X \leq t) = \int_0^t \vartheta e^{-\vartheta} dx = 1 - e^{-\vartheta t}\,.$$

Der ML-Schätzer dafür ist also $T_n(x_1, \ldots, x_n) = 1 - e^{-\frac{t}{\bar{x}}}$.

4. Sei $X \sim N(\mu, \sigma^2)$ normalverteilt. Wir wollen μ, σ^2 schätzen, und führen dazu n Messungen durch. Also

$$\mathcal{X} = \mathbb{R},\ f_{\mu,\sigma^2}(x_1, \ldots, x_n) = (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}\,.$$

Logarithmieren ergibt

$$\log f_{\mu,\sigma^2}(x_1, \ldots, x_n) = -n\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\,.$$

Fall 1. σ^2 bekannt, μ unbekannt.
Mit

$$\frac{d}{d\mu}\log f_{\mu,\sigma^2}(x_1, \ldots, x_n) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu)\ (= 0)$$

erhalten wir den ML-Schätzer $\widehat{\mu} = \overline{x} = \frac{x_1+\cdots+x_n}{n}$.
Fall 2. μ bekannt, σ^2 unbekannt.
Mit

$$\frac{d}{d\sigma}\log f_{\mu,\sigma^2} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (x_i - \mu)^2\ (= 0)$$

erhalten wir $n = \frac{1}{\sigma^2}\sum\limits_{i=1}^n (x_i - \mu)^2$, also den ML-Schätzer

$$\widehat{\sigma^2} = \frac{\sum\limits_{i=1}^n (x_i - \mu)^2}{n}\,.$$

Fall 3. μ, σ unbekannt.
Dann nehmen wir $\widehat{\mu} = \overline{x}$, $\widehat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n}$ und haben

$$E[\widehat{\sigma}^2] = \frac{E[\sum_{i=1}^n (X_i - \overline{X})^2]}{n} = \frac{\sum_{i=1}^n E[(X_i - \overline{X})^2]}{n}$$

und berechnet leicht (beachte $\sigma^2 = \mathrm{Var}[X]$)

$$E[\widehat{\sigma}^2] = \frac{n-1}{n}\sigma^2\,.$$

Der normierte Schätzer $\frac{\sum (x_i - \overline{x})^2}{n-1}$ ist also erwartungstreuer Schätzer für σ^2.

Schließlich wollen wir noch *beste* Schätzer analysieren. Wir machen die folgenden Annahmen:

1. Es sei $\theta \subseteq \mathbb{R}$ ein offenes Intervall.

2. Die Likelihood Funktion $\rho(x, \vartheta)$ ist auf $\mathcal{X} \times \theta$ positiv und nach ϑ stetig differenzierbar.

3. Es gilt die Vertauschungsrelation

$$\int\limits_{\mathcal{X}} \frac{d}{d\vartheta}\rho(x,\vartheta)dx = \frac{d}{d\vartheta}\int\limits_{\mathcal{X}} \rho(x,\vartheta)dx\,.$$

   Wenn $\mathcal{X}$ diskret ist, so wird das Integral durch die Summe $\sum\limits_{\mathcal{X}}$ ersetzt.

4. Sei $U_\vartheta(x) = \frac{d}{d\vartheta}\log\rho(x,\vartheta) = \frac{\frac{d}{d\vartheta}\rho(x,\vartheta)}{\rho(x,\vartheta)}$. Für jedes $\vartheta \in \theta$ existiert $\mathrm{Var}_\vartheta[U_\vartheta(X)]$ und ist $\neq 0$.

Ein Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$, das diese Bedingungen erfüllt, heißt *regulär*. Es soll $\tau(\vartheta)$ geschätzt werden. Ein Schätzer $T(x)$ heißt *regulär*, falls für alle ϑ

$$\int\limits_{\mathcal{X}} T(x)\frac{d}{d\vartheta}\rho(x,\vartheta)dx = \frac{d}{d\vartheta}\int\limits_{\mathcal{X}} T(x)\rho(x,\vartheta)dx$$

gilt.

Satz 4.1. *Gegeben ein reguläres Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ und ein erwartungstreuer regulärer Schätzer T für $\tau(\vartheta)$. Dann gilt*

$$\mathrm{Var}_\vartheta[T] \geq \frac{\tau'(\vartheta)^2}{\mathrm{Var}_\vartheta[U_\vartheta]} \quad \textit{für alle } \vartheta \in \theta\,.$$

Gleichheit für alle ϑ gilt genau dann, wenn

$$T = \tau(\vartheta) + \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}[U_\vartheta]} \quad \textit{für alle } \vartheta$$

ist.

Beweis. Sei T regulärer erwartungstreuer Schätzer. Wir haben

$$\begin{aligned} E_\vartheta[U_\vartheta] &= \int \frac{\frac{d}{d\vartheta}\rho(x,\vartheta)}{\rho(x,\vartheta)}\rho(x,\vartheta)dx = \int \frac{d}{d\vartheta}\rho(x,\vartheta)dx \\ &= \frac{d}{d\vartheta}\int \rho(x,\vartheta)dx = \frac{d}{d\vartheta}1 = 0\,. \end{aligned}$$

Für die Covarianz $\mathrm{cov}[T,U_\vartheta]$ erhalten wir mit $E_\vartheta[T]=\tau(\vartheta)$

$$\begin{aligned} \mathrm{cov}_\vartheta[T,U_\vartheta] &= E_\vartheta[TU_\vartheta] - E_\vartheta[T]E_\vartheta[U_\vartheta] \\ &= \int T(x)\frac{\frac{d}{d\vartheta}\rho(x,\vartheta)}{\rho(x,\vartheta)}\rho(x,\vartheta)dx \\ &= \frac{d}{d\vartheta}\int T(x)\rho(x,\vartheta)dx = \frac{d}{d\vartheta}E_\vartheta[T] \\ &= \frac{d}{d\vartheta}\tau(\vartheta) = \tau'(\vartheta)\,. \end{aligned}$$

Mit Lemma 2.8 folgt daraus

$$\begin{aligned} 0 \le \mathrm{Var}_\vartheta\left[T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}\right] &= \mathrm{Var}_\vartheta[T] + \frac{\tau'(\vartheta)^2\mathrm{Var}[U_\vartheta]}{(\mathrm{Var}_\vartheta[U_\vartheta])^2} - \frac{2\tau'(\vartheta)}{\mathrm{Var}_\vartheta[U_\vartheta]}\underbrace{\mathrm{cov}_\vartheta[T,U_\vartheta]}_{\tau'(\vartheta)} \\ &= \mathrm{Var}_\vartheta[T] - \frac{\tau'(\vartheta)^2}{\mathrm{Var}_\vartheta[U_\vartheta]}\,, \end{aligned}$$

also

$$\mathrm{Var}_\vartheta[T] \ge \frac{\tau'(\vartheta)^2}{\mathrm{Var}_\vartheta[U_\vartheta]}\,.$$

Gleichheit gilt genau dann, wenn $\mathrm{Var}_\theta[T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}[U_\vartheta]}] = 0$ ist. Nach Lemma 2.9 bedeutet dies

$$P_\vartheta\left(T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]} = E_\vartheta\left[T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}\right]\right) = 1\,.$$

Nun ist $E_\vartheta[T]=\tau(\vartheta)$, $E[\frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}] = 0$. Gleichheit gilt also genau dann, wenn

$$P_\vartheta\left(T = \tau(\vartheta) + \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}\right) = 1$$

ist. Da $\rho(x,\vartheta) > 0$ ist, folgt daraus sofort $T = \tau(\vartheta) + \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}$ im diskreten Fall. Im stetigen Fall kommt man aus Stetigkeitsgründen auf dasselbe Ergebnis. $\square$

Beispiele. 1. Eine Münze mit $P(K) = \vartheta$, $P(Z) = 1-\vartheta$, $0 < \vartheta < 1$, wird n Mal geworfen, x sei die Anzahl von Kopf, ϑ soll geschätzt werden. Hier haben wir $\mathcal{X} = \{0, 1, \ldots, n\}$, $\vartheta = (0,1)$,

$$\begin{aligned} \rho(x,\vartheta) &= \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} > 0\,, \\ U_\vartheta(X) &= \frac{X}{\vartheta} - \frac{n-X}{1-\vartheta} = \frac{X}{\vartheta(1-\vartheta)} - \frac{n}{1-\vartheta}\,, \end{aligned}$$

$$\mathrm{Var}_\vartheta[U_\vartheta(X)] = \frac{1}{\vartheta^2(1-\vartheta)^2}\mathrm{Var}[X] = \frac{n\vartheta(1-\vartheta)}{\vartheta^2(1-\vartheta)^2} = \frac{n}{\vartheta(1-\vartheta)} \neq 0\,.$$

Da $\mathcal{X}$ endlich ist, ist die Vertauschungsrelation trivialerweise erfüllt, $(\mathcal{X}, 2^{\mathcal{X}}, P_\vartheta : \vartheta \in \theta)$ ist also ein reguläres Modell. Es folgt für jeden erwartungstreuen Schätzer T (die Regularität ist wiederum erfüllt)

$$\mathrm{Var}_\vartheta(T) \geq \frac{1}{\mathrm{Var}_\vartheta[U_\vartheta]} = \frac{\vartheta(1-\vartheta)}{n}\,.$$

Der ML-Schätzer $\frac{x}{n}$ ist also bester Schätzer.

2. Es sei $(\mathcal{X}, 2^{\mathcal{X}}, P_\vartheta : \vartheta \in \theta)$, wobei $\mathcal{X} = \{0, 1, 2, \ldots\}$ und P_ϑ die Familie der Poissonverteilungen mit $\rho(x,\vartheta) = e^{-\vartheta}\frac{\vartheta^x}{x!}$ ist. Es soll ϑ geschätzt werden. Wir haben $\theta = (0,\infty)$, $\rho(x,\vartheta) > 0$. Wir verwenden den Schätzer $\widetilde{T}(x) = x$. Da $E_\vartheta[X] = \vartheta$ ist, so ist $\widetilde{T}$ erwartungstreu, außerdem ist $\mathrm{Var}_\vartheta[\widetilde{T}] = \vartheta$.

Für U_ϑ berechnen wir $U_\vartheta(X) = -\vartheta + \frac{X}{\vartheta} \neq 0$, $\mathrm{Var}_\vartheta[U_\vartheta(X)] = \frac{\mathrm{Var}[X]}{\vartheta^2} = \frac{1}{\vartheta} \neq 0$, und die Vertauschungsrelation ist leicht verifiziert. Nach dem Satz gilt für jeden erwartungstreuen Schätzer T

$$\mathrm{Var}_\vartheta[T] \geq \frac{1}{1/\vartheta} = \vartheta\,,$$

$\widetilde{T}(x) = x$ ist also ein bester Schätzer.

4.4 Konfidenzintervalle

Sei das Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ gegeben, ϑ oder allgemein $\tau(\vartheta)$ soll geschätzt werden.

Definition. Die Abbildung $C : \mathcal{X} \to 2^{\mathbb{R}}$, $x \mapsto C(x)$ Intervall $\subseteq \mathbb{R}$, heißt *Konfidenzabbildung zum Irrtumsniveau* α, $0 \leq \alpha \leq 1$, falls

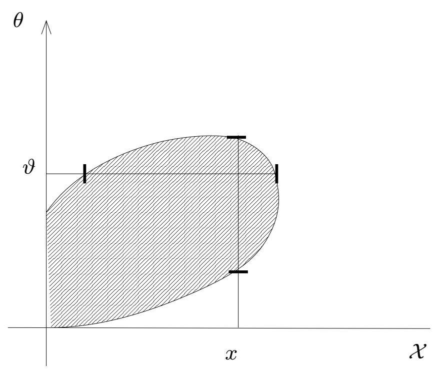
$$P_\vartheta(\{x \in \mathcal{X} : \vartheta \in C(x)\}) \geq 1 - \alpha \text{ für alle } \vartheta$$

gilt, also

$$\inf_\vartheta P_\vartheta(\{x \in \mathcal{X} : \vartheta \in C(x)\}) \geq 1 - \alpha \,.$$

Wir wollen nun eine Konstruktion solcher Konfidenzintervalle angeben. Natürlich soll $C(x)$ möglichst klein sein.

Sei $C = \{(x, \vartheta) \in \mathcal{X} \times \theta : \vartheta \in C(x)\}$, etwa



Für $x \in \mathcal{X}$ ist $C(x)$ der vertikale Schnitt. Für $\vartheta \in \theta$ ist

$$C_\vartheta = \{x \in \mathcal{X} : (x, \vartheta) \in C\}$$

der horizontale Schnitt, mit

$$P_\vartheta(C_\vartheta) = P_\vartheta(\{x \in \mathcal{X} : \vartheta \in C(x)\}) \,.$$

Wir verlangen also

$$\inf_\vartheta P_\vartheta(C_\vartheta) \geq 1 - \alpha \,.$$

Die Konstruktion erfolgt in zwei Schritten:

A. Zu $\vartheta \in \theta$ bestimme ein möglichst kleines C_ϑ mit $P_\vartheta(C_\vartheta) \geq 1-\alpha$, z.B. $C_\vartheta = \{x \in \mathcal{X} : \rho(x,\vartheta) \geq c_\vartheta\}$, wobei c_ϑ so bestimmt ist, dass $P_\vartheta(C_\vartheta) \geq 1-\alpha$ erfüllt ist.

B. Setze $C = \{(x,\vartheta) \in \mathcal{X} \times \theta : x \in C_\vartheta\}$, und dann

$$C(x) = \{\vartheta \in \theta : x \in C_\vartheta\}.$$

Dann gilt $\inf_\vartheta P_\vartheta(C_\vartheta) \geq 1-\alpha$.

Beispiele. 1. Wir werfen eine Münze n Mal mit $P(K) = \vartheta$, $P(Z) = 1-\vartheta$, $0 < \vartheta < 1$, $x =$Anzahl Kopf. Wir wissen, dass $\frac{x}{n}$ ein bester Schätzer für ϑ ist. Wir wählen $C(x) = (\frac{x}{n} - \varepsilon, \frac{x}{n} + \varepsilon)$ mit $\varepsilon > 0$, so dass

$$P_\vartheta(\{x : |\frac{x}{n} - \vartheta| < \varepsilon\}) \geq 1-\alpha \Leftrightarrow P_\vartheta(\{x : |\frac{x}{n} - \vartheta| \geq \varepsilon) \leq \alpha\,.$$

Wie erreichen wir das? Nach der Ungleichung von Tschebyschev haben wir

$$P_\vartheta(\{x : |\frac{x}{n} - \vartheta| \geq \varepsilon) \leq \frac{\vartheta(1-\vartheta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}\,.$$

Ein $\varepsilon > 0$ funktioniert also, sobald

$$\frac{1}{4n\varepsilon^2} \leq \alpha \Longleftrightarrow \varepsilon \geq \frac{1}{2\sqrt{n\alpha}}$$

gilt.

Ist zum Beispiel $n = 1000$, $\alpha = 0.025$, so genügt $\varepsilon = 0,1$. Wollen wir ein Konfidenzintervall mit Radius $\varepsilon = 0,01$ erhalten, so muß $\sqrt{n} \geq \frac{1}{2\varepsilon\sqrt{\alpha}}$ sein, also $n \geq \frac{1}{4\varepsilon^2\alpha} = 100.000$.

Mit de Moivre-Laplace haben wir

$$\begin{aligned} P_\vartheta(\{x : |\frac{x}{n} - \vartheta| < \varepsilon\}) &= P_\vartheta\left(\left\{x : |\frac{x-n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}| < \varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right\}\right) \\ &\approx \phi\left(\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - \phi\left(-\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \\ &= 2\phi\left(\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - 1\,, \end{aligned}$$

da $\phi(-x) = 1 - \phi(x)$ ist.

Für $\alpha = 0,025$ ergibt dies $\varepsilon \geq 0,0446$ bei $n = 1000$.

2. Wir wollen den Prozentsatz der Wähler einer Partei A schätzen. Werden n Wähler befragt und x sind Wähler von A, so nehmen wir als Schätzer $\frac{x}{n}$ für die W-keit ϑ, dass ein zufällig ausgewählter Wähler für A stimmt. Wie groß muß n sein, dass die W-keit eines Irrtums von höchstens 1% nicht mehr als $0,05$ beträgt?
Wir verlangen also

$$P_\vartheta(|\frac{x}{n} - \vartheta| \leq 0,01) \geq 0,95.$$

Durch Zentrieren ist dies äquivalent zu

$$P_\vartheta\left(|\frac{x - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}| \leq 0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \geq 0,95\,,$$

also

$$\phi\left(0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - \phi\left(-0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) = 2\phi\left(0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - 1 \geq 0,95\,,$$

somit

$$\phi\left(0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \geq 0,975\,.$$

Aus der Tabelle ergibt sich

$$0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}} \geq 1,96\,,$$

das heißt

$$n \geq 38416 \cdot \vartheta(1-\vartheta)\,.$$

Da $\vartheta(1-\vartheta) \leq \frac{1}{4}$ ist, genügt jedenfalls $n \geq 9604$. Falls wir von vorneherein wissen, dass $\vartheta \leq 0,1$ ist, so genügt $n \geq 38416 \cdot 0,09 \approx 3458$.

3. Die Familie der Verteilungen sei $N(\mu, \sigma^2)$, wobei μ der unbekannte Parameter ist, $\sigma^2 > 0$ ist bekannt. Wir machen n Beobachtungen $x_1, \ldots, x_n$, dann ist $X_1 + \cdots + X_n$ nach $N(n\mu, n\sigma^2)$ verteilt, also

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \sim N(\mu, \frac{\sigma^2}{n}),\ Z = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0,1)\,.$$

Sei t_α so gewählt, dass gilt

$$P_\mu(|Z| \leq t_\alpha) \geq 1 - \alpha \Longleftrightarrow P_\mu(|Z| \geq t_\alpha) \leq \alpha\,.$$

Wir nehmen den Schätzer $\overline{x} = \frac{x_1+\cdots+x_n}{n}$ und wählen das Intervall

$$C(x) = (\overline{x} - \frac{t_\alpha \sigma}{\sqrt{n}},\ \overline{x} + \frac{t_\alpha \sigma}{\sqrt{n}})\,,$$

das heißt

$$\mu \in C(x) \Longleftrightarrow |\frac{\sqrt{n}}{\sigma}(\overline{x} - \mu)| \leq t_\alpha\,.$$

Es folgt $P_\mu(\{x : \mu \in C(x)\}) \geq 1 - \alpha$, also ist $C(x)$ Konfidenzintervall zum Niveau α.

Wählen wir zum Beispiel $\alpha = 0,05$, dann berechnet man $t_\alpha = 1,96$, also ist $C(x) = (\overline{x} - \frac{2\sigma}{\sqrt{n}}, \overline{x} + \frac{2\sigma}{\sqrt{n}})$ Konfidenzintervall für den Mittelwert μ.

Bei unbekannten μ und σ^2 nimmt man als Konfidenzintervall

$$C(x) = (\overline{x} - \frac{2s}{\sqrt{n}},\ \overline{x} + \frac{2s}{\sqrt{n}})$$

wobei $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$ ist.

4.5 Testen von Hypothesen

Wir wiederholen nochmals die Situation in den folgenden fünf Schritten:

1. Das Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ wird festgelegt.
2. Die Hypothesen werden formuliert:

   $\theta = \theta_0 \,\dot{\cup}\, \theta_1$,
   $H_0 : \vartheta \in \theta_0$ Nullhypothese,
   $H_1 : \vartheta \in \theta_1$ Alternativhypothese.
3. Das Irrtumsniveau α wird gewählt, $0 < \alpha < 1$ meist $\alpha = 0,1$ oder $0,05$ oder $0,025$.

   Irrtum 1. Art: $\vartheta \in \theta_0$, aber H_1 wird angenommen

   Irrtum 2. Art: $\vartheta \in \theta_1$, aber H_0 wird angenommen.

   Ein Irrtum 1. Art soll höchstens mit W-keit α vorkommen.

4. Eine Entscheidungsregel wird festgelegt. Man wählt einen Test $\varphi : \mathcal{X} \to [0,1]$, φ ist Zufallsvariable.

   Deterministischer Test:

$$\varphi(x) = \begin{cases} 0 & H_0 \text{ wird angenommen} \\ 1 & H_1 \text{ wird angenommen.} \end{cases}$$

   Randomisierter Test: $\varphi(x) \in [0,1]$ ist die W-keit, mit der man sich für H_1 entscheidet.

5. Jetzt erst wird das Experiment durchgeführt!

Definition. Sei $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$, $\theta = \theta_0 \mathbin{\dot\cup} \theta_1$, $\varphi : \mathcal{X} \to [0,1]$ gegeben. $A = \{x \in \mathcal{X} : \varphi(x) = 0\}$ heißt *Annahmebereich*, $R = \{x \in \mathcal{X} : \varphi(x) = 1\}$ *Ablehnungsbereich* (von H_0). Falls φ deterministischer Test ist, so heißt

$$G_\varphi : \theta \to [0,1],\ G_\varphi(\vartheta) = P_\vartheta(\varphi \in R)\ \textit{Gütefunktion.}$$

Falls φ randomisierter Test ist, so ist die Gütefunktion

$$G_\varphi(\vartheta) = E_\vartheta[\varphi]\,.$$

Natürlich ist $G_\varphi(\vartheta) = E_\vartheta[\varphi]$ auch für deterministische Tests.

Die Forderungen an die Gütefunktion sind demnach:

$$\begin{aligned} \vartheta \in \theta_0 \implies & G_\varphi(\vartheta) \le \alpha \quad \text{Irrtum 1. Art,} \\ \vartheta \in \theta_1 \implies & G_\varphi(\vartheta) \text{ möglichst groß } (1 - G_\varphi(\vartheta) \text{ Irrtum 2. Art}), \end{aligned}$$

und wir nennen dann φ einen Test zum *Niveau* α.

Definition. $\varphi : \mathcal{X} \to [0,1]$ heißt *bester Test* zum Niveau α, wenn für alle Tests ψ zum Niveau α gilt

$$G_\varphi(\vartheta) \ge G_\psi(\vartheta) \text{ für alle } \vartheta \in \theta_1\,.$$

Beispiel. Tea tasting lady. Die Dame behauptet, sie könne feststellen, ob zuerst Tee und dann Milch in die Tasse gegeben wurde, oder umgekehrt. Es seien $n = 8$ Tassen und 4 von jedem Typ, x sei die Anzahl der Treffer. Wir haben also folgende Situation:

1. $\mathcal{X} = \{0, 1, \ldots, 8\}$, $P_\vartheta(X = x) = \binom{8}{x}\vartheta^x(1-\vartheta)^{8-x}$,

2. $\theta = [\frac{1}{2}, 1]$,

   $H_0 : \theta_0 = \{\frac{1}{2}\}$, sie hat nicht die Fähigkeit

   $H_1 : \theta_1 = (\frac{1}{2}, 1]$, sie kann tatsächlich die Zusammensetzung richtig bestimmen.

3. $\alpha = 0,05$.

4. $\varphi(x) = \begin{cases} 0 & x < c \Rightarrow H_0 \\ 1 & x \geq c \Rightarrow H_1 \end{cases}$ deterministisch.

Zur richtigen Wahl von c berechnen wir

$$G_\varphi(\frac{1}{2}) = P_{\frac{1}{2}}(X \geq c) = \sum_{k=c}^{8} \binom{8}{k} \frac{1}{2^8} \leq 0,05$$

und das gilt für $c \geq 7$. Also ist die Entscheidungsregel

$$\begin{aligned} x < 7 &\Rightarrow H_0 \\ x \geq 7 &\Rightarrow H_1 \,. \end{aligned}$$

Für $\vartheta \in \theta_1$, das heißt $\vartheta > \frac{1}{2}$, haben wir

$$G_\varphi(\vartheta) = \sum_{k=7}^{8} \binom{8}{k} \vartheta^k (1-\vartheta)^{8-k} = 8\vartheta^7(1-\vartheta) + \vartheta^8 = \vartheta^7(8 - 7\vartheta)\,,$$

$G_\varphi(\vartheta)$ ist monoton steigend. Der Irrtum 2. Art beträgt $1 - \vartheta^7(8 - 7\vartheta)$. Zum Beispiel erhalten wir für $\vartheta = \frac{3}{4}$ (die Lady bestimmt im Mittel 6 von 8 Tassen richtig), $G_\varphi(\frac{3}{4}) \sim 0,367$, also ist der Irrtum 2. Art $0,633$. Die Lady hat keine Chance, weil die Stichprobenzahl n zu klein ist.

Definition. Ein Test φ heißt *unverfälscht* (unbiased) zum Niveau α, falls

$$G_\varphi(\vartheta_0) \leq \alpha \leq G_\varphi(\vartheta_1) \text{ für alle } \vartheta_0 \in \theta_0, \vartheta_1 \in \theta_1$$

gilt.

Das Verfahren bei der Tea tasting lady ist biased, da

$$G_\varphi(\frac{1}{2}) = \frac{9}{2^8} < 0,05$$

ist, also $G_\varphi(\frac{1}{2} + \varepsilon) < \alpha = 0,05$ für kleine $\varepsilon > 0$.

Wir studieren nun im Detail sogenannte *Alternativtests*. Es gibt nur zwei Verteilungen P_0, P_1,

$$(\mathcal{X}, \mathcal{E}, P_0 \cup P_1),\ \theta = \{0, 1\},$$

$$H_0 : \theta_0 = \{0\},\ H_1 : \theta_1 = \{1\}\,.$$

Wir setzen für die Likelihood Funktion $\rho(x, \vartheta)$

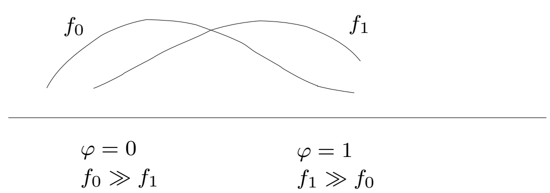
$$\rho_0(x) = \rho(x, 0), \rho_1(x) = \rho(x, 1)$$

und setzen voraus $\rho_0(x) + \rho_1(x) > 0$ für alle $x \in \mathcal{X}$, das heißt:

$P_0(X = x) + P_1(X = x) > 0$	falls X diskrete Variable ist
$f_0(x) + f_1(x) > 0$	falls X stetige Variable ist, mit Dichten $f_0(x)$, $f_1(x)$.

Dies ist keine Einschränkung, da wir ansonsten $\mathcal{X}$ kleiner machen können.

Die Idee ist, dass wir H_0 beibehalten, falls f_0 größer als f_1 ist bzw. H_1 nehmen, falls f_1 größer als f_0 ist:



Definition. Der *Likelihood Quotient* ist

$$R(x) = \begin{cases} \frac{\rho_1(x)}{\rho_0(x)} & (\rho_0(x) > 0) \\ \infty & (\rho_0(x) = 0) \end{cases}.$$

Wenn also $R(x) > c$ für geeignetes c ist, so ist die Tendenz zu H_1 "hinreichend stark".

Definition. Sei $(\mathcal{X}, \mathcal{E}, P_0 \cup P_1)$, $\theta = \{0, 1\}$ gegeben. $\varphi^* : \mathcal{X} \to [0, 1]$ heißt *Neyman-Pearson Test*, wenn es eine Konstante $c^* > 0$ gibt mit

$$\varphi^*(x) = \begin{cases} 0 & R(x) < c^* \\ 1 & R(x) > c^* \\ \gamma(x) & R(x) = c^*,\ \gamma(x) \text{ beliebig in } [0, 1]\,. \end{cases}$$

Satz 4.2 (Neyman-Pearson)**.** *Gegeben* $(\mathcal{X}, \mathcal{E}, P_0 \cup P_1)$, $\theta = \{0,1\}$, $0 < \alpha < 1$, $\rho_0(x) + \rho_1(x) > 0$ *für alle* $x \in \mathcal{X}$. *Dann gilt:*

a. *Jeder Neyman-Pearson Test* φ^* *mit* $E_0[\varphi^*] = \alpha$ *ist bester Test.*

b. *Es gibt einen Neyman-Pearson Test* φ^* *mit* $E_0[\varphi^*] = \alpha$.

c. *Jeder beste Test zum Niveau* α *ist Neyman-Pearson Test, bis auf eine Menge vom Maß 0.*

Beweis. a. Sei $\varphi*$ Neyman-Pearson Test mit $c^* > 0$ und φ ein beliebiger Test mit $E_0[\varphi^*] = \alpha$, $E_0[\varphi] \leq \alpha$. Zu zeigen ist $E_1[\varphi] \leq E_1[\varphi^*]$.

Sei $g = (\rho_1 - c^*\rho_0)(\varphi^* - \varphi)$. Für $R(x) = \frac{\rho_1(x)}{\rho_0(x)} < c^*$ haben wir $\rho_1(x) - c^*\rho_0(x) < 0$, $\varphi^*(x) = 0$, also $\varphi^*(x) - \varphi(x) \leq 0$, das heißt $g(x) \geq 0$.
Für $R(x) > c^*$ ist $\rho_1(x) - c^*\rho_0(x) > 0$ (wegen $\rho_1(x) + \rho_0(x) > 0$) und $\varphi^*(x) = 1$, $\varphi^*(x) - \varphi(x) \geq 0$, also wiederum $g(x) \geq 0$. Für $R(x) = c^*$ ist $\rho_1(x) - c^*\rho_0(x) = 0$, also $g(x) = 0$. Daraus folgt (für diskrete Variablen nehme man die Summe)

$$
\begin{aligned}
0 \leq \int g(x)dx &= \int (\varphi^*(x) - \varphi(x))(\rho_1(x) - c^*\rho_0(x))dx \\
&= E_1[\varphi^* - \varphi] - c^* E_0[\varphi^* - \varphi] \\
&= E_1[\varphi*] - E_1[\varphi] - c^*(E_0 \underbrace{[\varphi^*] - E_0[}_{\geq 0}\varphi]) \\
&\leq E_1[\varphi^*] - E_1[\varphi]\,,
\end{aligned}
$$

also $E_1[\varphi^*] \geq E_1[\varphi]$.

b. Für $c \geq 0$ setzen wir

$$\alpha(c) = P_0(R(X) > c),\ \overline{\alpha}(c) = P_0(R(X) \geq c)\,.$$

Dann ist $\alpha(c) \leq \overline{\alpha}(c)$, $\alpha(c)$ ist monoton fallend für steigendes c, und $\overline{\alpha}(0) = 1$. Ferner ist

$$\overline{\alpha}(c) - \alpha(c) = P_0(R(X) = c).$$

Sei (c_n) eine strikt monoton steigende Folge gegen ∞, und $A_n = \{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) > c_n\}$. Wir haben

$$A_1 \supseteq A_2 \supseteq \dots\,,$$

$$P_0(\bigcap A_n) = P_0(\{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) = \infty\}) = P_0(\varnothing) = 0\,,$$

somit $\lim P(A_n) = 0$, das heißt $\lim\limits_{c_n \to \infty} \alpha(c_n) = 0$.

Sei (c_n) eine strikt monoton steigende Folge gegen $c > 0$. Mit der Definition der A_n wie oben haben wir

$$P_0(\bigcap A_n) = P_0(\{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) \geq c\}) = \overline{\alpha}(c)\,,$$

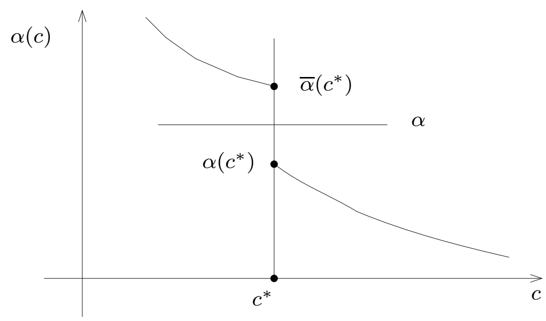
also $\lim\limits_{c_n \to c} \alpha(c_n) = \overline{\alpha}(c)$.
Schließlich sei (b_n) eine strikt fallende Folge gegen b,

$$B_n = \{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) > b_n\}\,.$$

Hier ist

$$\begin{aligned} &B_1 \subseteq B_2 \subseteq \dots\,, \ \bigcup B_n = \{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) > b\}\,, \\ &P_0(\bigcup B_n) = \alpha(b)\,, \end{aligned}$$

also $\lim\limits_{b_n \to b} \alpha(b_n) = \alpha(b)\,.$
Wir sehen, dass $\alpha(c)$ eine rechtsstetige Funktion in c ist. Sei $c^* = \inf\{c : \alpha(c) \leq \alpha\}$, dann ist $\overline{\alpha}(c^*) \geq \alpha \geq \alpha(c^*)$.



Falls $\overline{\alpha}(c^*) = \alpha(c^*)$ ist, so setzen wir $\gamma^* = 0$. Falls $\overline{\alpha}(c^*) > \alpha(c^*)$ ist, so setzen wir

$$\gamma^* = \frac{\alpha - \alpha(c^*)}{\overline{\alpha}(c^*) - \alpha(c^*)}\,,$$

und erklären in beiden Fällen den Test φ^* durch

$$\varphi^*(x) = \begin{cases} 0 & R(x) < c^* \\ 1 & R(x) > c^* \\ \gamma^* & R(x) = c^*\,. \end{cases}$$

Daraus ergibt sich (im Fall einer diskreten Variablen nehmen wir wieder die Summe)

$$\begin{aligned}
E_0[\varphi^*] &= \int \varphi^*(x)\rho_0(x)dx = \int\limits_{R(x)>c^*} \rho_0(x)dx + \int\limits_{R(x)=c^*} \gamma^*\rho_0(x)dx \\
&= P_0(R(X) > c^*) + \gamma^* P_0(R(X) = c^*) \\
&= \alpha(c^*) + \begin{cases} 0 \quad \text{falls} \quad \overline{\alpha}(c^*) = \alpha(c^*) = \alpha \\ \frac{\alpha-\alpha(c^*)}{\overline{\alpha}(c^*)-\alpha(c^*)}(\overline{\alpha}(c^*) - \alpha(c^*)) \text{ falls } \overline{\alpha}(c^*) > \alpha(c^*). \end{cases}
\end{aligned}$$

In beiden Fällen ergibt sich $E_0[\varphi^*] = \alpha$, also ist das Niveau α ausgeschöpft.

c. Es sei φ ein beliebiger bester Test mit $E_0[\varphi] = \alpha$, und φ^* mit $c^* > 0$ ein Neyman-Pearson Test mit $E_1[\varphi] = E_1[\varphi^*]$. Mit der Funktion $g(x)$ wie in Teil a) haben wir $0 = \int g(x)dx$, also ist wegen $g(x) \geq 0$, $g(x) = 0$ bis auf eine Menge mit Maß 0. Da $\{x : \rho_1(x) - c^*\rho_0 = 0\}$ Maß 0 hat, muß also $\varphi(x) = \varphi^*(x)$ sein bis auf eine Menge vom Maß 0. $\square$

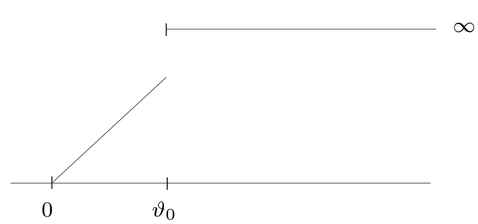
Beispiel. Betrachten wir noch einmal das Orangen Beispiel mit Parametern N und n. Wir wollen

$$\theta_0 = \{0, 1, \ldots, \vartheta_0\} \text{ gegen } \theta_1 = \{\vartheta_0 + 1, \ldots, N\}$$

testen. Für irgendein $\vartheta_1 \in \theta_1$ ist (wir setzen $P_\vartheta(x) = P_\vartheta(X = x)$)

$$\begin{aligned}
R(x) &= \frac{P_{\vartheta_1}(x)}{P_{\vartheta_0}(x)} = \frac{P_{\vartheta_0+1}(x)P_{\vartheta_0+2}(x)\cdots P_{\vartheta_1}(x)}{P_{\vartheta_0}(x)P_{\vartheta_0+1}(x)\cdots P_{\vartheta_1-1(x)}} = \prod_{k=\vartheta_0}^{\vartheta_1-1} \frac{P_{k+1}(x)}{P_k(x)} \\
&= \prod_{k=\vartheta_0}^{\vartheta_1-1} \frac{\binom{k+1}{x}\binom{N-k-1}{n-x}}{\binom{k}{x}\binom{N-k}{n-x}} = \prod_{k=\vartheta_0}^{\vartheta_1-1} \frac{k+1}{k+1-x} \cdot \frac{N-k-n+x}{N-k},
\end{aligned}$$

und diese Funktion ist für $x \leq \vartheta_0$ monoton steigend, und $R(x) = \infty$ für $x > \vartheta_0$:



Wir setzen den Neyman-Pearson an:

$$\varphi^*(x) = \begin{cases} 0 & x < c^* \\ 1 & x > c^* \\ \gamma^* & x = c^*\,. \end{cases}$$

Dabei werden die Konstanten c^*, γ^* aus

$$E_{\vartheta_0}[\varphi^*] = P_{\vartheta_0}(X > c^*) + \gamma^* P_{\vartheta_0}(X = c) = \alpha$$

ermittelt. Der Test φ^* hängt nicht von ϑ_1 ab, und daher ist ϑ_0 gegen jedes ϑ_1 testbar.

Nehmen wir $\vartheta < \vartheta_0$, so sagt der Satz, dass φ_{ϑ_0} besser als φ_ϑ ist, also $\alpha = E_0[\varphi^*_{\vartheta_0}] \geq E_0[\varphi^*_\vartheta]$. Unsere Intuition ist also richtig. Man suche ϑ_0 mit $E_{\vartheta_0}[\varphi^*] = \alpha$.

Unser Eingangsbeispiel mit $N = 10.000$, $n = 50$ führt bei $\alpha = 0,025$ zu $c^* = 6$, $\gamma^* = 0,52$, $\vartheta_0 = 500$ $(= 5\%)$, und daher zum Test

$$\begin{aligned} x < 6 &\;\Rightarrow\; H_0 \\ x > 6 &\;\Rightarrow\; H_1 \\ x = 6 &\;\Rightarrow\; H_1 \text{ mit } W\text{-keit } 0,52\,. \end{aligned}$$

5 Nichtparametrische Methoden

Bisher haben wir angenommen, dass eine Familie von Verteilungen gegeben ist. Wenn die möglichen Verteilungen unbekannt sind, so versucht man, aus den Daten strukturelle Zusammenhänge zu finden.

5.1 Empirische Verteilungsfunktion

Wir nehmen an, dass die Daten mindestens ordinales Niveau haben. Also zum Beispiel Klassennoten, Ranglisten oder Körpergrößen.

Wir ziehen eine Stichprobe $x_1, x_2, \ldots, x_n$. Die Voraussetzungen sind:

1. Die Zufallsvariablen $X_1, \ldots, X_n$ sind identisch unabhängig
2. Die unbekannte Verteilungsfunktion $F(x)$ ist stetig.

Aus der Stichprobe ermittelt man zunächst folgende Größen:

1. Die Reihenfolge $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$, wobei Bindungen $x_i = x_j$ ausgeschlossen werden oder durch Werfen einer Münze aufgelöst werden.
2. $x_{(1)}$ =Minimum, $x_{(n)}$ = Maximum, Median m

$$m = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ ungerade} \\ \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right) & n \text{ gerade,} \end{cases}$$

   $d = x_{(n)} - x_{(1)} =$ Spanne.

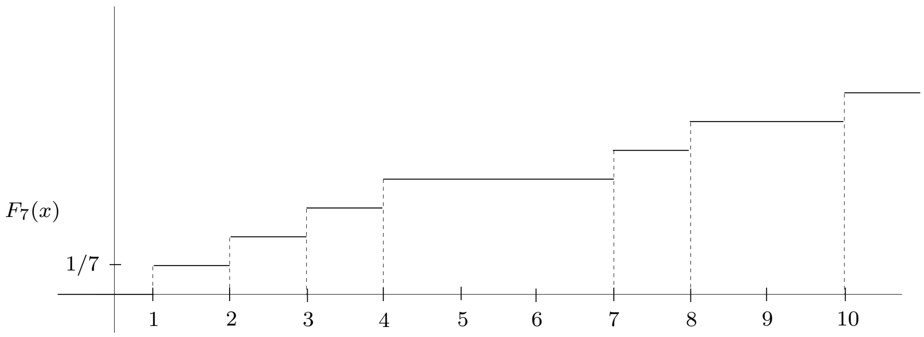
Definition. $F_n(x) = \frac{1}{n}\#\{i : x_i \leq x\}$ heißt die *empirische Verteilungsfunktion.*

$F_n(x)$ ist also eine Treppenfunktion.

Beispiel.

x_1	x_2	x_3	x_4	x_5	x_6	x_7
3	7	1	8	10	2	4

Dann ist $x_{(1)} = 1$, $x_{(7)} = 10$, , $m = 4$, $d = 9$



Die empirische Verteilungsfunktion hat offenbar folgende Eigenschaften:

1. $F_n(x)$ ist monoton steigend,
2. $\lim\limits_{x\to-\infty} F_n(x) = 0$, $\lim\limits_{x\to\infty} F_n(x) = 1$.

Ferner ist $F_n(x)$ diskrete Zufallsvariable mit Werten in $\{\frac{0}{n}, \frac{1}{n}, \cdots, \frac{n}{n}\}$ für jedes x.

Satz 5.1. *Sei $F(x)$ die unbekannte Verteilungsfunktion. Dann gilt*

$$P(F_n(x) = \frac{k}{n}) = \binom{n}{k} F(x)^k (1 - F(x))^{n-k} \ (k = 0, \ldots, n)\,,$$

das heißt die Zufallsvariable $nF_n(x)$ ist binomialverteilt $b(k, n; F(x))$ für jedes x.

Beweis. Wir haben $P(X_i \leq x) = F(x)$ für $i = 1, \ldots, n$. Sei x fest und $Y_i(x)$ erklärt durch

$$Y_i(x) = \begin{cases} 1 & X_i \leq x \\ 0 & X_i > x\,, \end{cases}$$

dann ist $P(Y_i(x) = 1) = F(x)$. Die Variable $Y_i(x)$ ist also Bernoulli verteilt mit Erfolgswahrscheinlichkeit $p = F(x)$. Aus $nF_n = Y_1(x) + \cdots + Y_n(x)$ folgt daher, dass $nF_n(x)$ binomialverteilt ist mit $b(k, n; F(x))$, und somit

$$P(F_n(x) = \frac{k}{n}) = P(nF_n(x) = k) = \binom{n}{k} F(x)^k (1 - F(x))^{n-k}\,. \qquad \square$$

Folgerung 5.2. *Wir haben für jedes x*

a. $E[F_n(x)] = F(x)$

b. $\mathrm{Var}[F_n(x)] = \frac{F(x)(1-F(x))}{n}$.

Beweis. Aus $E[nF_n(x)] = nF(x)$ folgt $E[F_n(x)] = F(x)$, und aus $\mathrm{Var}[nF_n(x)] = nF(x)(1-F(x))$ folgt $\mathrm{Var}[F_n(x)] = \frac{F(x)(1-F(x))}{n}$. □

Definition. Bei kardinalem Niveau sind der *empirische Erwartungswert* und die *empirische Varianz* gegeben durch

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n}, \; s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 .$$

Es seien $\overline{X} = \frac{X_1+\cdots+X_n}{n}$, $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{2}\sum\limits_{i=1}^{n} X_i^2 - \overline{X}^2$ die entsprechenden Zufallsvariablen.

Satz 5.3. *Es gilt*

a. $E[\overline{X}] = E[X]$, $\mathrm{Var}[\overline{X}] = \frac{1}{n}\mathrm{Var}[X]$,

b. $E[S^2] = \frac{n-1}{n}\mathrm{Var}[X]$.

Beweis. a. Wir haben

$$E[\overline{X}] = \frac{1}{n}E[X_1 + \cdots + X_n] = E[X] ,$$

$$\mathrm{Var}[\overline{X}] = \frac{1}{n^2}\mathrm{Var}[X_1 + \cdots + X_n] = \frac{1}{n}\mathrm{Var}[X] .$$

b. Für S^2 erhalten wir

$$E[S^2] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - E[\overline{X}^2] = E[X^2] - E[\overline{X}^2] .$$

Ferner ist

$$\begin{aligned} E[\overline{X}^2] = & \frac{1}{n^2}E[(X_1 + \cdots + X_n)^2] = \frac{1}{n^2}(\sum_{i=1}^{n} E[X_i^2] + 2\sum_{i<j} E[X_iX_j]) \\ = & \frac{1}{n}E[X^2] + \frac{2}{n^2}\sum_{i<j} E[X_i]E[X_j] \\ = & \frac{1}{n}E[X^2] + \frac{n-1}{n}E[X]^2 , \end{aligned}$$

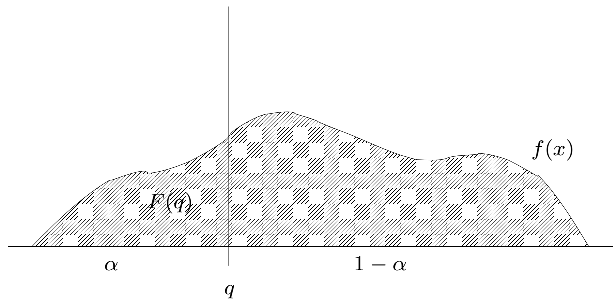
also

$$E[S^2] = \frac{n-1}{n}(E[X^2] - E[X]^2) = \frac{n-1}{n}\mathrm{Var}[X]\,. \qquad \square$$

Definition. Sei $F(x)$ stetige Verteilungsfunktion, $0 < \alpha < 1$. Die Zahl q heißt α-*Quartil*, falls

$$P(X < q) \leq \alpha,\ P(X > q) \leq 1 - \alpha\,,$$

das heißt $F(q) = \alpha$.

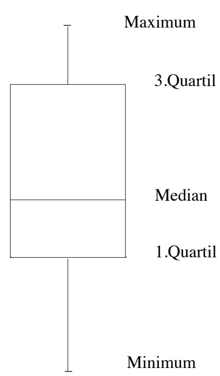


Für $\alpha = \frac{1}{4}$ sprechen wir vom 1. *Quartil*, und für $\alpha = \frac{3}{4}$ vom 3. *Quartil.*

Für die empirische Verteilungsfunktion $F_n(x)$ bedeutet dies

$$\alpha = \frac{1}{4} : F_n(q) \leq \frac{1}{4} \Longrightarrow x_{\lfloor(\frac{n}{4})\rfloor}$$
$$\alpha = \frac{3}{4} : F_n(q) \leq \frac{3}{4} \Longrightarrow x_{\lfloor(\frac{3n}{4})\rfloor}\,.$$

Die übliche graphische Darstellung einer Stichprobe $x_1, \ldots, x_n$ ist das sogenannte Boxplot:



5.2 Verteilung der Ränge

Die Daten seien mindestens auf ordinalem Niveau, $x_1, \ldots, x_n$ die Stichprobe. Die entsprechenden Zufallsvariablen sind identisch unabhängig verteilt.

Definition. Die Variable $R_i = R(X_i)$ ist die Zufallsvariable, die X_i den Rang in der Stichprobe zuteilt, mit Wertebereich $\{1, 2, \ldots, n\}$.

Satz 5.4. *Es gilt*

a. $P(X_1 = r_1 \wedge X_2 = r_2 \wedge \ldots \wedge X_n = r_n) = \frac{1}{n!}$ *für alle Permutationen* $r_1 r_2 \ldots r_n$ *von* $\{1, 2, \ldots, n\}$,

b. $P(R_i = j) = \frac{1}{n}$ *für alle* i *und* j,

c. $P(R_i = k \wedge R_j = \ell) = \frac{1}{n(n-1)}$ *für alle* $i \neq j$, $k \neq \ell$,

d. $E[R_i] = \frac{n+1}{2}$, $\mathrm{Var}[R_i] = \frac{n^2-1}{12}$ *für alle* i,

e. $\mathrm{cov}(R_i, R_j) = -\frac{n+1}{12}$ *für alle* $i \neq j$,

f. $\rho(R_i, R_j) = -\frac{1}{n-1}$ *für alle* $i \neq j$.

Beweis. a. Dies ist wegen der Unabhängigkeit klar.
b. $P(R_i = j) = P(X_i = x_{(j)}) = \frac{(n-1)!}{n!} = \frac{1}{n}$.
c. $P(R_i = k \wedge R_j = \ell) = P(X_i = x_{(k)} \wedge X_j = x_{(\ell)}) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$.
d. $E[R_i] = \sum_{j=1}^{n} j P(R_i = j) = \frac{1}{n} \sum_{j=1}^{n} j = \frac{n+1}{2}$. Ferner ist

$$E[R_i^2] = \sum_{j=1}^{n} j^2 P(R_i = j) = \frac{1}{n} \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6},$$

also

$$\mathrm{Var}[R_i] = E[R_i^2] - E[R_i]^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.$$

e. Wir haben $\operatorname{cov}(R_i, R_j) = E[R_iR_j] - E[R_i]E[R_j]$, also

$$\begin{aligned}
\operatorname{cov}(R_i, R_j) &= \sum_{k=1}^{n} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} k\ell P(R_i = k \wedge R_j = \ell) - (\frac{n+1}{2})^2 \\
&= \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} k\ell - (\frac{n+1}{2})^2 \\
&= \frac{1}{n(n-1)} \sum_{k=1}^{n} k(\frac{n(n+1)}{2} - k) - (\frac{n+1}{2})^2 \\
&= \frac{1}{n(n-1)} \frac{n^2(n+1)^2}{4} - \frac{1}{n(n-1)} \frac{n(n+1)(2n+1)}{6} - (\frac{n+1}{2})^2 \\
&= -\frac{n+1}{12} .
\end{aligned}$$

f. Wir haben

$$\rho(R_i, R_j) = \frac{\operatorname{cov}(R_i, R_j)}{\sqrt{\operatorname{Var}[R_i]\operatorname{Var}[R_j]}} = -\frac{n+1}{12} \cdot \frac{12}{n^2-1} = -\frac{1}{n-1} . \qquad \square$$

5.3 Tests mittels geordneter Statistiken

Typische Beispiele aus der Praxis sind:

A. Männer haben mehr Autounfälle als Frauen.

B. Medikament A ist wirksamer als Medikament B.

C. Kindergartenkinder haben in der Schule bessere Noten als Nicht-Kindergartenkinder.

Unsere Vorgangsweise orientiert sich am Hypothesen Testen nach den üblichen 5 Schritten:

1. Wir treffen Annahmen über das Modell und die Daten.
2. Wir stellen die Hypothesen auf.
3. Wir wählen das Irrtumsniveau α.

4. Wir formulieren die Teststatistik und die Entscheidungsregel.

5. Jetzt erst wird das Experiment durchgeführt.

A. Vorzeichentest

Die Daten sind mindestens auf ordinalem Niveau. Sie werden in Paaren (x_i, y_i) erhoben, $i = 1, \ldots, n$. Die Zufallsvariabeln (X_i, Y_i) sind unabhängig und intern unabhängig.

Beispiel. Die Daten (X_i, Y_i) entsprechen Mann/Frau, und es werden die Körpergröße oder Unfallhäufigkeit erhoben.

Die Ereignisse werden folgendermaßen bezeichnet:

$$\begin{array}{lr} X_i < Y_i : & \text{Ereignis } + \\ X_i = Y_i : & 0 \\ X_i > Y_i : & - \end{array}$$

Der zweiseitige Test betrachtet die Hypothesen

$$\begin{array}{ll} H_0 : & P(+) = P(-) \\ H_1 : & P(+) \neq P(-)\,. \end{array}$$

Der einseitige Test betrachtet

$$\begin{array}{ll} H_0 : & P(+) \leq P(-) \\ H_1 : & P(+) > P(-) \end{array}\,.$$

Unentschieden (Ereignisse 0) werden entfernt. Ist α Irrtumsniveau im einseitigen Test, so wird $\frac{\alpha}{2}$ im zweiseitigen Test verwendet.

Zweiseitiger Test. Wir nehmen als Teststatistik

$$T = \#(+)\,.$$

Unter H_0 haben wir $P(+) = P(-) = \frac{1}{2}$, und T ist binomialverteilt $b(k, n; \frac{1}{2})$. Die Entscheidungsregel ist demnach

$$t < T < n - t \;\Rightarrow\; H_0$$

$$T \leq t \text{ oder } T \geq n - t \;\Rightarrow\; H_1\,,$$

für geeignestes t.

Analyse: Unter H_0 ist der Irrtum 1. Art

$$\sum_{k=0}^{t} \binom{n}{k} \frac{1}{2^n} \leq \frac{\alpha}{2}\,.$$

Für große n und $\alpha = 0,05$ ergibt dies $t \sim \frac{n}{2} - \sqrt{n}\,.$

Einseitiger Test. Unter H_0 ist $X_i > Y_i$ wahrscheinlicher. Wir verwenden als Teststatistik

$$T = \#(-)$$

mit der Entscheidungsregel

$$\begin{aligned} T > t \;&\Rightarrow\; H_0 \\ T \leq t \;&\Rightarrow\; H_1\,. \end{aligned}$$

Für große n und $\alpha = 0,025$ ist $t \sim \frac{n}{2} - \sqrt{n}\,.$

B. Wilcoxon Rangsummentest

Die Daten sind mindestens auf ordinalem Niveau. Gegeben sind $m + n$ unabhängige Variablen $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ mit den Rängen $1, \ldots, N = m+n$. $F(x)$ sei die (unbekannte) Verteilungsfunktion der identisch verteilten X_i und $G(x)$ jene der identisch verteilten Y_j.

Beispiel. Es sollen 4 Kindergartenkinder $X_1, \ldots, X_4$ gegen 8 Nicht-Kindergartenkinder $Y_1, \ldots, Y_8$ nach ihren schulischen Leistungen verglichen werden. Die X_i haben die Ränge 4,7,9,12 (1 = schlechtester bis 12 = bester Rang).

Die Hypothesen sind

$$\begin{aligned} &H_0 : F(x) = G(x) \quad \text{(kein Unterschied)} \\ &H_1 : F(x) \neq G(x)\,. \end{aligned}$$

Sei $Z_{(1)} < Z_{(2)} < \cdots < Z_{(N)}$ die geordnete Stichprobe, und

$$V_i = \begin{cases} 1 & Z_{(i)} \text{ ist } X\text{-Variable} \\ 0 & Z_{(i)} \text{ ist } Y\text{-Variable.} \end{cases}$$

Die Variablen $V_1, \ldots, V_N$ sind nicht unabhängig, aber unter der Hypothese H_0 haben alle $0,1$-Vektoren mit m 1en und n 0en dieselbe W-keit $\frac{1}{\binom{N}{m}}\,.$

Lemma 5.5. *Unter H_0 gilt*

a. $E[V_i] = \frac{m}{N}$ *für alle* i,

b. $\mathrm{Var}[V_i] = \frac{mn}{N^2}$ *für alle* i,

c. $\mathrm{cov}(V_i, V_j) = -\frac{mn}{N^2(N-1)}$ *für alle* $i \neq j$.

Beweis. a. $P(V_i = 1) = P(Z_{(i)} = X \text{ Variable}) = \frac{m}{N}$. V_i ist Bernoulli Variable mit $p = \frac{m}{N}$, somit gilt $E[V_i] = \frac{m}{N}$.

b. $\mathrm{Var}[V_i] = \frac{m}{N} \cdot \frac{N-m}{N} = \frac{mn}{N^2}$.

c. Wir haben für $i \neq j$

$$E[V_iV_j] = P(V_i = 1 \wedge V_j = 1) = \frac{\binom{m}{2}}{\binom{N}{2}} = \frac{m(m-1)}{N(N-1)},$$

und somit

$$\begin{aligned} \mathrm{cov}(V_i, V_j) &= E[V_iV_j] - E[V_i]E[V_j] \\ &= \frac{m(m-1)}{N(N-1)} - \frac{m^2}{N^2} = -\frac{mn}{N^2(N-1)}. \qquad \square \end{aligned}$$

Als Teststatistik verwenden wir

$$W_N = \sum_{i=1}^{N} iV_i = \text{Summe der Ränge der } X_j\,.$$

Unter H_0 ist

$$E[W_N] = \sum_{i=1}^{N} iE[V_i] = \frac{m}{N} \sum_{i=1}^{N} i = \frac{m(N+1)}{2}\,.$$

Ausrechnen ergibt

$$\mathrm{Var}[W_N] = \frac{mn(N+1)}{12}\,.$$

Sei $W_{\min} = \sum_{i=1}^{m} i$, $W_{\max} = \sum_{i=N-m+1}^{N} i$. Die Entscheidungsregel ist demnach aus Symmetriegründen

$$W_N \leq t \quad \text{oder} \quad W_N \geq W_{\max} - (t - W_{\min}) \Longrightarrow H_1$$

mit

$$P(W_N \le t) \le \frac{\alpha}{2} \quad \text{unter } H_0\,.$$

Beispiel. Analysieren wir das Kindergartenbeispiel. Hier ist $m = 4$, $n = 8$, $N = 12$, $\binom{N}{m} = \binom{12}{4} = 495$, $\alpha = 0,05$. Wir haben $W_{\max} = 9+10+11+12 = 42$, $W_{\min} = 1+2+3+4 = 10$. Wir stellen eine Tabelle auf:

$$\begin{array}{rl} W = 10 & 1,2,3,4 \\ 11 & 1,2,3,5 \\ 12 & 1,2,3,6; 1,2,4,5 \\ 13 & 1,2,3,7; 1,2,4,6; 1,3,4,5 \\ 14 & 1,2,3,8; 1,2,4,7; 1,2,5,6; 1,3,4,6; 2,3,4,5\,. \end{array}$$

Also ist $P(T \le 14) = \frac{12}{495} \sim 0,024 \le 0,025$

Die Entscheidungsregel lautet demnach

$$W \le 14 \;\text{ oder }\; W \ge 38 \Longrightarrow H_1\,.$$

In unserem Beispiel ist $W = 4+7+9+12 = 32$. Die Nullhypothese H_0 kann also nicht verworfen werden.

C. Mediantest

Beispiel. Eine Reifenfirma entwickelt einen Reifentyp und behauptet, dass die Reifen im Median ≥ 33.000 km halten. Wie sollen wir das testen?
Wir nehmen kardinales Niveau an. Die Variablen $X_1, \ldots, X_n$ sind unabhängig identisch verteilt mit stetiger Verteilungsfunktion $F(x)$, die symmetrisch um den Median liegt. M_0 ist vorgegeben.

Zweiseitiger Test:

$$\begin{array}{ll} H_0: & M = M_0 \\ H_1: & M \ne M_0\,. \end{array}$$

Einseitiger Test:

$$\begin{array}{ll} H_0: & M \ge M_0 \\ H_1: & M < M_1\,. \end{array}$$

Es sei $Y_i = X_i - M_0$ und $r(|Y_i|)$ der Rang von $|Y_i|$. Nun setzen wir

$$Z_i = \begin{cases} 1 & \text{falls} \quad Y_j > 0 \;\text{ wobei }\; r(|Y_j|) = i \\ 0 & \text{falls} \quad Y_j < 0 \;\text{ wobei }\; r(|Y_j|) = i\,. \end{cases}$$

Wir können $Y_j = 0$ wegen $P(Y_j = 0) = 0$ vernachlässigen.

Nun betrachten wir die Teststatistik

$$W^+ = \sum_{i=1}^{n} iZ_i = \text{Summe der Ränge der } \textit{positiven} \text{ Differenzen.}$$

Zweiseitiger Test: Unter H_0 ist $E[Z_i] = \frac{1}{2}$ wegen der symmetrischen Verteilung um M_0, also

$$E[W_i^+] = \sum_{i=1}^{n} iE[Z_i] = \frac{n(n+1)}{4}.$$

Alle $0,1$-Vektoren $(z_1, \ldots, z_n)$ haben W-keit $\frac{1}{2^n}$, somit ist

$$P(W^+ = w) = \frac{a(w)}{2^n},$$

wobei $a(w)$ die Anzahl der n-Tupel $(y_1, \ldots, y_n)$ ist, so dass die Summe der Ränge der positiven y_i gleich w ist.

Die Entscheidungsregel ist demnach

$$W^+ \leq t \;\text{ oder }\; W^+ \text{ groß } \Longrightarrow H_1,$$

wobei

$$P(W^+ \leq t) \leq \frac{\alpha}{2}.$$

Im einseitigen Test wird $P(W^+ \leq t) \leq \alpha$ verwendet.

5.4 Tests auf Korrelation

Wir untersuchen zwei Merkmale X, Y, zum Beispiel

X:	Körpergröße Vater	Y:	Körpergröße Sohn
	Geschlecht		Wahlverhalten
	Gesundheit		Schulerfolg

und stellen uns die Fragen

A. Sind X und Y unabhängig?

B. Sind X und Y positiv (negativ) korreliert?

Die Daten sind mindestens auf ordinalem Niveau.

Beispiel. Sieben Bewerber stellen sich vor. Zwei Personalvertreter stellen jeweils eine Rangliste auf (1 =bester bis 7 = schlechtester) .

Bewerber	1	2	3	4	5	6	7
A	5	7	1	3	4	6	2
B	3	6	1	2	4	7	5

Sind die Ranglisten korreliert?

Es seien $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ die Stichprobenwerte,

$$\begin{array}{lll} r_1, \ldots, r_n & \text{die Ränge von} & x_1, \ldots, x_n \\ s_1, \ldots, s_n & \text{die Ränge von} & y_1, \ldots, y_n, \end{array}$$

$\{r_1, \ldots, r_n\} = \{s_1, \ldots, s_n\} = \{1, \ldots, n\}$.

Definition. Der *Korrelationskoeffizient von Spearman* ist

$$r = \frac{\sum_{i=1}^n (r_i - \overline{r})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^n (r_i - \overline{r})^2}\sqrt{\sum_{i=1}^n (s_i - \overline{s})^2}}\,,$$

wobei $\overline{r} = \overline{s} = \frac{n+1}{2}$.

Satz 5.6. *Wir haben*

$$r = 1 - \frac{6\sum_{i=1}^n d_i^2}{(n-1)n(n+1)}\,, d_i = r_i - s_i \ (i = 1, \ldots, n)\,.$$

Beweis. Es ist

$$\begin{aligned} \sum_{i=1}^n (r_i - \overline{r})^2 &= \sum_{i=1}^n (s_i - \overline{s})^2 = \sum_{i=1}^n (i - \frac{n+1}{2})^2 \\ &= \frac{n(n+1)(2n+1)}{6} - (n+1)\frac{n(n+1)}{2} + \frac{n(n+1)^2}{4} \\ &= \frac{n(n+1)}{12}(4n + 2 - 6n - 6 + 3n + 3) \\ &= \frac{(n-1)n(n+1)}{12}\,. \end{aligned}$$

Daraus folgt

$$r = \frac{12}{(n-1)n(n+1)} \sum_{i=1}^{n} (r_i - \frac{n+1}{2})(s_i - \frac{n+1}{2}) .$$

Setzen wir $d_i = r_i - s_i = (r_i - \frac{n+1}{2}) - (s_i - \frac{n+1}{2})$, so ergibt sich

$$\begin{aligned}
\sum_{i=1}^{n} d_i^2 &= \sum_{i=1}^{n} (r_i - \frac{n+1}{2})^2 - 2 \sum_{i=1}^{n} (r_i - \frac{n+1}{2})(s_i - \frac{n+1}{2}) + \sum_{i=1}^{n} (s_i - \frac{n+1}{2})^2 \\
&= \frac{(n-1)n(n+1)}{6} - 2 \sum_{i=1}^{n} (r_i - \overline{r})(s_i - \overline{s}) ,
\end{aligned}$$

also

$$\begin{aligned}
r &= \frac{12}{(n-1)n(n+1)} \frac{(n-1)n(n+1) - 6 \sum_{i=1}^{n} d_i^2}{12} \\
&= 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{(n-1)n(n+1)} . \qquad \square
\end{aligned}$$

Beispiel. In unserem Beispiel ist

$$d_1^2 = 4, d_2^2 = 1, d_3^2 = 0, d_4^2 = 1, d_5^2 = 0, d_6^2 = 1, d_7^2 = 9 ,$$

also

$$r = 1 - \frac{6 \cdot 16}{6 \cdot 7 \cdot 8} = \frac{5}{7} = 0,714 .$$

Folgerung 5.7. *Für den Korrelationskoeffizienten von Spearman gilt*

a. $-1 \leq r \leq 1$,

b. $r = 1 \Longrightarrow r_i = s_i \ (i = 1, \ldots, n)$,

c. $r = -1 \Longrightarrow r_i = n + 1 - s_i \ (i = 1, \ldots, n)$.

Beweis. Aus Satz 5.6 folgt, dass $-1 \leq r \leq 1$ äquivalent ist zu

$$-1 \leq 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{(n-1)n(n+1)} \leq 1$$

also zu

$$0 \leq \sum_{i=1}^{n} d_i^2 \leq \frac{(n-1)n(n+1)}{3},$$

wobei die linke Seite genau $r = 1$ entspricht und die rechte Seite $r = -1$. Die linke Seite $0 \leq \sum\limits_{i=1}^{n} d_i^2$ ist offensichtlich gültig mit $\sum\limits_{i=1}^{n} d_i^2 = \sum\limits_{i=1}^{n} (r_i - s_i)^2 = 0 \Longrightarrow r_i = s_i$ für alle i. Um die rechte Seite zu verifizieren, können wir o.B.d.A. $s_i = i$ setzen, also müssen wir

$$\sum_{i=1}^{n} (r_i - i)^2 \leq \frac{(n-1)n(n+1)}{3}$$

zeigen, mit Gleichheit genau für $r_i = n+1-i$. Betrachten wir eine Permutation $(r_1, \ldots, r_n)$. Falls $r_k < r_{k+1}$ gilt, so gilt für die Permutation $(r'_1, \ldots, r'_k, r'_{k+1}, \ldots, r'_n)$ mit $r'_k = r_{k+1}$, $r'_{k+1} = r_k$, $r'_i = r_i$ $(i \neq k, k+1)$

$$\begin{aligned}
&\sum_{i=1}^{n} (r'_i - i)^2 - \sum_{i=1}^{n} (r_i - i)^2 = (r'_k - k)^2 + (r'_{k+1} - (k+1))^2 \\
&- (r_k - k)^2 - (r_{k+1} - (k+1))^2 \\
&= (r_{k+1} - k)^2 + (r_k - (k+1))^2 - (r_k - k)^2 - (r_{k+1} - (k+1))^2 \\
&= -2kr_{k+1} - 2(k+1)r_k + 2kr_k + 2(k+1)r_{k+1} \\
&= 2r_{k+1} - 2r_k > 0\,.
\end{aligned}$$

Der größte Ausdruck ist daher genau für $r_1 = n, r_2 = n-1, \ldots, r_n = 1$ gegeben und für diesen berechnen wir

$$\begin{aligned}
\sum_{i=1}^{n} (n+1-2i)^2 &= n(n+1)^2 - 4\frac{(n+1)^2 n}{2} + 4\frac{n(n+1)(2n+1)}{6} \\
&= -n(n+1)^2 + \frac{2}{3}n(n+1)(2n+1) \\
&= \frac{n(n+1)}{3}(4n+2-3n-3) = \frac{(n-1)n(n+1)}{3},
\end{aligned}$$

wie gewünscht. $\square$

Damit können wir sagen:

Falls $r \sim 1$ ist,	dann sind X, Y positiv korreliert, große Ränge der x_i entsprechen großen Rängen der y_i, und umgekehrt.
Falls $r \sim -1$ ist,	dann sind X, Y negativ korreliert, kleine Ränge der x_i entsprechen großen Rängen der y_i und umgekehrt.
Falls $r \sim 0$ ist,	dann sind X, Y unkorreliert.

5.5 Korrelationstest von Spearman

Die Variablen $X_1, \ldots, X_n$ sind unabhängig identisch verteilt, ebenso die Variablen $Y_1, \ldots, Y_n$, die Paare $(X_1, Y_1), \ldots, (X_n, Y_n)$ sind unabhängig.

Zweiseitiger Test.

$$\begin{array}{ll} H_0: & X, Y \ \text{unabhängig} \\ H_1: & X, Y \ \text{korreliert}\,. \end{array}$$

Einseitiger Test.

$$\begin{array}{ll} H_0: & X, Y \ \text{unabhängig} \\ H_1: & X, Y \ \text{positiv korreliert}\,. \end{array}$$

Wir definieren die Rang-Zufallsvariablen

$$R_i = R(X_i),\ S_i = R(Y_i),\ D_i = R_i - S_i\,,$$

$$D = \sum_{i=1}^{n} D_i^2 = \sum_{i=1}^{n} (R_i - S_i)^2\,.$$

O.B.d.A. sei $s_i = i$, also $R(S_i) = i$, und somit

$$D = \sum_{i=1}^{n} (R_i - i)^2 = \sum_{i=1}^{n} R_i^2 - 2\sum_{i=1}^{n} iR_i + \sum_{i=1}^{n} i^2\,.$$

Wir wissen

$$E[R_i] = \frac{n+1}{2},\ E[R_i^2] = \frac{(n+1)(2n+1)}{6},\ \mathrm{Var}[R_i] = \frac{n^2-1}{12}\,.$$

Daraus folgt

$$\begin{aligned}
E[D] &= \frac{n(n+1)(2n+1)}{3} - 2\sum_{i=1}^{n} iE[R_i] \\
&= \frac{n(n+1)(2n+1)}{3} - \frac{n(n+1)^2}{2} = \frac{(n-1)n(n+1)}{6}\,.
\end{aligned}$$

Ist R die Zufallsvariable für den Spearman Koeffizienten, so folgt aus Satz 5.6

$$E[R] = 1 - \frac{6E[D]}{(n-1)n(n+1)} = 0\,.$$

Wir nehmen als Teststatistik

$$d = \sum_{i=1}^{n}(r_i - s_i)^2\,.$$

Niedrige Werte von d weisen auf positive Korrelation hin, hohe Werte auf negative Korrelation.

Einseitiger Test. Die Entscheidungsregel ist

$$\begin{array}{lcl}
d > t_\alpha & \Rightarrow & H_0 \\
d \le t_\alpha & \Rightarrow & H_1\,,
\end{array}$$

wobei t_α das α-Quantil der Gleichverteilung über alle Permutationen ist.

Beispiel. In unserem Beispiel haben wir $n = 7$, $d = 16$. Alle Permutationen $(r_1, \ldots, r_7)$ haben dieselbe W-keit $\frac{1}{7!}$. Gesucht ist das α-Quantil $t = t_\alpha$ mit

$$\sum_{w=0}^{t} \frac{a(w)}{7!} \le \alpha, \quad t = \text{maximal groß}\,,$$

wobei $a(w) = \#\{(r_1, \ldots, r_7) \colon d = \sum_{i=1}^{7}(r_i - i)^2 = w\}$. Wir haben

$$\begin{array}{ll}
a(0) = 1 & 1\,2\,3\,4\,5\,6\,7 \\
a(1) = 0 & \\
a(2) = 6 & 2\,1\,3\,4\,5\,6\,7, \ldots, 1\,2\,3\,4\,5\,7\,6 \\
a(3) = 0 & \\
\quad\vdots &
\end{array}$$

Es ergibt sich bei $\alpha = 0,05$ das α-Quantil $t_\alpha = 18$. Die Entscheidungsregel $d \le 18$ besagt also, dass H_0 verworfen wird, die Ranglisten sind positiv korreliert.

Für große Stichproben nimmt man in der Praxis

$$Z_n = \frac{R_n}{\sqrt{n-1}} \xrightarrow{\text{i.V.}} N(0,1) \text{ (wenn } X, Y \text{ unabhängig sind)}\,.$$

Die Nullhypothese wird abgelehnt, wenn $Z \geq z_{1-\alpha}$ mit $\phi(z_{1-\alpha}) = 1-\alpha$.

5.6 Tests für nominale Skalen

Ein typisches Beispiel aus der Wahlforschung vergleicht die Variable X =Einkommen, Y =Parteipräferenz, und stellt die Frage, ob X und Y unabhängig sind oder nicht.

Wir haben zwei Merkmale A und B mit den disjunkten Klassen

$$\begin{array}{ll} A: & A_1, A_2, \ldots, A_k \\ B: & B_1, B_2, \ldots, B_\ell \end{array}$$

Die Daten fasst man in einer sogenannten *Kontingenztafel* zusammen, das heißt in einer Matrix (n_{ij}), $i = 1, \ldots, k$, $j = 1, \ldots, \ell$, wobei n_{ij} = #Merkmale in $A_i \cap B_j$.

Beispiel. A =Einkommen, B =Parteipräferenz

A \ B	CDU	SPD	FDP	Andere	$\sum$
hoch	35 / 22,5	7 / 21,5	5 / 4,5	3 / 1,5	50
mittel	250 / 270	250 / 258	80 / 54	20 / 18	600
niedrig	165 / 157,5	173 / 150,5	5 / 31,5	7 / 10,5	350
	450	430	90	30	1000

In den Kästchen wird n_{ij}, $\widetilde{n}_{ij}$ notiert, wobei

$$
\begin{array}{ll}
n_{ij} = \# \text{ in } A_i \cap B_j & \text{tatsächlich erhoben} \\
\widetilde{n}_{ij} = \# \text{ in } A_i \cap B_j & \text{wenn die Merkmale unabhängig sind.}
\end{array}
$$

Wir schreiben für die *Randhäufigkeiten* kurz

$$
n_{i\cdot} = \# A_i = \sum_{j=1}^{\ell} n_{ij} \,,
$$

$$
n_{\cdot j} = \# B_j = \sum_{i=1}^{k} n_{ij} \,,
$$

alsos

$$
n = \sum_{i=1}^{k} n_{i\cdot} = \sum_{j=1}^{\ell} n_{\cdot j} = \sum_{i,j} n_{ij} \,.
$$

Falls die Merkmale unabhängig sind, so haben wir

$$
\frac{\widetilde{n}_{ij}}{n} = P(X \in A_i \wedge Y \in B_j) = P(X \in A_i) P(Y \in B_j) = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} \,,
$$

also

$$
\widetilde{n}_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{n} \quad (i = 1, \dots, k, j = 1, \dots, \ell) \,.
$$

Es gilt

$$
\sum_{i=1}^{k} \widetilde{n}_{ij} = \frac{n_{\cdot j}}{n} \sum_{i=1}^{k} n_{i\cdot} = n_{\cdot j} \,, \sum_{j=1}^{\ell} \widetilde{n}_{ij} = n_{i\cdot} \,.
$$

A. Chi-Quadrat Test auf Unabhängigkeit

Dies ist wohl das bekannteste aller Testverfahren. Wir haben die Merkmale $A_1, \dots, A_k$ und $B_1, \dots, B_\ell$ mit den Zufallsvariablen X und Y.

Die zu testenden Hypothesen sind

H_0: Merkmale A und B sind unabhängig
H_1: Merkmale sind abhängig.

Unter H_0 bietet sich als Teststatistik

$$X^2 = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$$

an. Der Nenner erklärt sich aus folgender Überlegung. Wenn n_{ij} und $\widetilde{n}_{ij}$ klein sind, so fällt $n_{ij} - \widetilde{n}_{ij}$ mehr ins Gewicht, daher die Gewichtung mit $\frac{1}{\widetilde{n}_{ij}}$.

Sei

$$P(X \in A_i) = p_{i\cdot}\,, \; P(Y \in B_j) = p_{\cdot j}\,, \; P(X \in A_i \wedge Y \in B_j) = p_{ij}\,.$$

Unter H_0 gilt dann

$$p_{ij} = p_{i\cdot} p_{\cdot j} \quad \text{für alle } i, j\,.$$

Somit haben wir

$$\begin{array}{lll} H_0\text{:} & p_{ij} = p_{i\cdot} p_{\cdot j} & \text{für alle } i, j \\ H_1\text{:} & p_{ij} \neq p_{i\cdot} p_{\cdot j} & \text{für ein Paar } (i, j)\,. \end{array}$$

Unter H_0 sei N_{ij} die Zufallsvariable

$$N_{ij} = \# \text{ in } A_i \cap B_j\,,$$

somit

$$\begin{aligned} P(N_{11} = n_{11} \wedge \ldots \wedge N_{k\ell} = n_{k\ell}) &= \frac{n!}{n_{11}! \cdots n_{k\ell}!}\, p_{11}^{n_{11}} \ldots p_{k\ell}^{n_{k\ell}} \\ &= \frac{n!}{n_{11}! \ldots n_{k\ell}!}\, (p_{1\cdot} p_{\cdot 1})^{n_{11}} \cdots (p_{k\cdot} p_{\cdot \ell})^{n_{k\ell}}\,. \end{aligned}$$

Die Teststatistik X^2 hängt also von den unbekannten Parametern $p_{i\cdot}$, $p_{\cdot j}$ ab. Wir verwenden nun die Maximum Likelihood Methode zur Schätzung dieser Parameter. Wir haben

$$P(N_{11} = n_{11} \wedge \ldots \wedge N_{k\ell} = n_{k\ell}) = \max\,,$$

das heißt

$$\prod_{i=1}^{k} \prod_{j=1}^{\ell} (p_{i\cdot} p_{\cdot j})^{n_{ij}} = \max$$

unter den Nebenbedingungen $\sum\limits_{i=1}^{k} p_{i\cdot} = 1\,, \; \sum\limits_{j=1}^{\ell} p_{\cdot j} = 1\,.$

Mit der Methode von Lagrange aus der Analysis ergeben sich die Schätzer

$$\widehat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}\,,\ \widehat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}$$

und daraus die Teststatistik

$$X^2 = \sum_{i,j} \frac{(n_{ij} - n\widehat{p}_{i\cdot}\widehat{p}_{\cdot j})^2}{n\widehat{p}_{i\cdot}\widehat{p}_{\cdot j}}\,.$$

Diese Statistik ist χ^2-verteilt mit $k\ell - (k-1) - (\ell-1) - 1 = (k-1)(\ell-1)$ Freiheitsgraden, also

$$X^2 \sim \gamma_{\frac{1}{2},\frac{(k-1)(\ell-1)}{2}} \text{ gamma-verteilt.}$$

Die Entscheidungsregel besagt:

$$X^2 \geq \chi^2_{1-\alpha,(k-1)(\ell-1)} \Longrightarrow H_1\,,$$

wobei $\chi^2_{1-\alpha,(k-1)(\ell-1)}$ das $(1-\alpha)$-Quantil der χ^2-Verteilung ist.

Beispiel. In userem Ausgangsbeispiel berechnet man $X^2 = 59,63$. Beim Niveau $\alpha = 0,05$ ist $\chi^2_{0,95;6} = 12,59$. Die Nullhypothese der Unabhängigkeit wird abgelehnt.

B. Fisher Test bei 2×2-Tafeln

Bei Merkmalen mit jeweils zwei Ausprägungen gibt es einen weiteren sehr bekannten Test auf Unabhängigkeit.

Beispiel. Gegeben sei folgende Kontinggenztafel

$A \backslash B$	B_1	B_2	Σ
A_1	2	8	10
A_2	3	7	10
Σ	5	15	

Wiederum soll gestestet werden:

$$\begin{array}{ll} H_0: & X, Y \text{ unabhängig} \\ H_1: & X, Y \text{ abhängig}. \end{array}$$

Die Idee besteht darin, alle 2×2-Tafeln mit den gleichen Randhäufigkeiten zu betrachten. In unserem Beispiel sind dies

$$\begin{array}{c|c} 0 & 10 \\ \hline 5 & 5 \end{array} \quad \begin{array}{c|c} 1 & 9 \\ \hline 4 & 6 \end{array} \quad \begin{array}{c|c} 2 & 8 \\ \hline 3 & 7 \end{array} \quad \begin{array}{c|c} 3 & 7 \\ \hline 2 & 8 \end{array} \quad \begin{array}{c|c} 4 & 6 \\ \hline 1 & 9 \end{array} \quad \begin{array}{c|c} 5 & 5 \\ \hline 0 & 10 \end{array}.$$

Allgemein sei die beobachtete Tafel

	B_1	B_2	$\sum$
A_1	a	b	$a+b$
A_2	c	d	$c+d$
$\sum$	$a+c$	$b+d$	

dann sind alle Tafeln mit gleicher Randhäufigkeit von der Form mit $0 \leq x \leq \min(a+b, a+c)$.

x	$a+b-x$	$a+b$	Ziehung
$a+c-x$	$d-a+x$	$c+d$	
$a+c$	$b+d$		
rot	weiß		

Als Modell stellen wir uns eine Urne vor mit $a+c$ roten Kugeln und $b+d$ weißen Kugeln. Man ziehe $a+b$ Kugeln ohne Zurücklegen. Die Merkmale entsprechen also folgenden Ereignissen.

A: Ziehung der Kugel

A_1: Kugel in der Ziehung
A_2: Kugel nicht in der Ziehung

B: Kugelfarbe

B_1: rot
B_2: weiß.

Als Teststatistik wählen wir

$$T = \#\text{rote Kugeln in Stichprobe.}$$

Unter H_0 ist T hypergeometrisch verteilt mit

$$P(T = x) = \frac{\binom{a+c}{x}\binom{b+d}{a+b-x}}{\binom{n}{a+b}}\,.$$

Die Entscheidungsregel ist demnach

$$T \leq c_{\alpha/2} \;\text{ oder }\; T \geq C_{1-\alpha/2} \Longrightarrow H_1\,,$$

wobei $c_{\alpha/2}$ das $\alpha/2$-Quantil der hypergeometrischen Verteilung ist.

Beispiel. In unserem Beispiel ist $T = 2$. Für $\alpha = 0,05$ erhalten wir $c_{\alpha/2} = 0$, $c_{1-\alpha/2} = 5$. H_0 kann also nicht abgelehnt werden.

Literatur

H. Büning, G. Trenkler: Nichtparametrische statistische Methoden 1978, de Gruyter.

W.J. Conover: Practical Nonparametric Statistics, 2nd edition 1971, John Wiley.

W. Feller: Probability Theory and its Applications, vol. I, 1950, John Wiley.

H.-O. Georgii: Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik, 2. Auflage 2007, de Gruyter.

G.R. Grimmett, D.R. Stirzaker: Probability and Random Processes, 2nd edition, 1992, Clarendon Press.

A. Klenke: Wahrscheinlichkeitstheorie, 2. Auflage 2008, Springer.

U. Krengel: Einführung in die Wahrscheinlichkeitstheorie und Statistik, 8. Auflage 2008, Vieweg-Verlag.

H. Toutenburg, C. Heumann: Deskriptive Statistik, 5. Auflage 2006, Springer.