Freie Universität Berlin

Fachbereich Mathematik/Informatik

Stochastik I

WS 2011/12

Mitschrift von: Moritz Hoffmann (Email: clonker@googlemail.com)

Vorlesung von: Prof. Dr. Carsten Hartmann

Inhaltsverzeichnis

1 Grundbegriffe

Stochastik nach Georgii: "Stochastik ist die Lehre von den Gesetzmäßigkeiten des Zufalls". Was ist Zufall? Nach Aristoteles: "Wenn im Bereich der Geschehnisse, die im strengen Sinne wegen etwas eintreten, und deren Ursache außer ihnen liegt, etwas geschieht, das mit dem Ergebnis nicht in eine Deswegen-Beziehung zu bringen ist, dann nennen wir das zufällig."

1.0.1 Beispiel[Zufallsexperimente]**.**

- Würfel, Münzwurf, Ching-Chang-Chong $\to$ zufällige Augenzahl /... / ...
- Schlange an der Supermarktkasse $\to$ zufällige Länge der Schlange / Verteilung, Wartezeit
- Zufallsprozesse, Radioaktivität $\to$ zufällige Anzahl der Klicks (im z.B. Geigerzähler) pro Minute

Wie modelliert / formalisiert man das?

(i) Modellierungsannahme
  - Das Zufallsexperiment kann beliebig oft wiederholt werden
  - Das Wissen über zurückliegende Ausgänge des Zufallsexperiments nützt nichts (Unabhängigkeit)

(ii) Formalisierung
  - Stichprobenraum (Menge aller möglichen Ereignisse), z.B. $\{1, \dots, 6\}$ beim Würfel
  - "Häufigkeit" bestimmter Ereignisse, z.B. $P(\text{Wurf} = \{1\}) = \frac{1}{6}$
  - Wiederholbarkeit: Grenzwert $n \to \infty$

Zunächst brauchen wir drei Grundbegriffe:

(i) Stichprobenraum (Raum der Elementarereignisse) Ω, Elementarereignis $\omega \in \Omega$ (mögliches Ergebnis). Bsp:
  - 1x Würfeln: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Nx Würfeln: $\Omega = \{1, 2, 3, 4, 5, 6\}^N$

(ii) Menge von Ereignissen $E \subset \Omega$, bzw "System" von Ereignissen $\mathcal{E}$, so dass jedem $E \in \mathcal{E}$ "sinnvoll" eine Wahrscheinlichkeit $P(E)$ zugeordnet werden kann.

  Beispiel[Würfeln]**.** $\Omega = \{1, 2, 3, 4, 5, 6\} \Rightarrow \mathcal{E} = 2^\Omega = \mathcal{P}(\Omega)$.
  - $E = \{2, 4, 6\} \in 2^\Omega$. $P(E) = \frac{1}{2}$
  - $E = \{1, 3, 5\} \in 2^\Omega$. $P(E) = \frac{1}{2}$

(iii) Wahrscheinlichkeit bzw. Wahrscheinlichkeitsmaß P. P ist eine Abbildung, die jedem $E \in \mathcal{E}$ eine Zahl zwischen 0 und 1 zuordnet.
($P : \mathcal{E} \to [0, 1]$, $E \mapsto P(E)$). Beispiel siehe Punkt 2.

1.1 Systeme von Ereignissen

Frage: Welche Teilmengen von Ω sind zulässig, um sinnvoll über Wahrscheinlichkeiten sprechen zu können?
Antwort: σ-Algebra auf Ω.

1.1.1 Definition[σ-Algebra]**.** Es sei $\mathcal{E} \subset 2^{\Omega}$. Dann heißt $\mathcal{E}$ σ-Algebra, wenn

- $\emptyset, \Omega$ in ihr enthalten sind
- $E \in \mathcal{E} \Rightarrow E^c \in \mathcal{E}$ (Abgeschlossenheit bzgl. Komplement)
- $E_1, E_2, \ldots \in \mathcal{E} \Rightarrow \bigcup\limits_{i \geq 1} E_i \in \mathcal{E}$

1.1.2 Beispiel. $\{\emptyset, \Omega\}$ und 2^{Ω} sind die kleinst- bzw. größtmöglichen σ-Algebren.

1.2 Wahrscheinlichkeitsräume

1.2.1 Definition[Wahrscheinlichkeitsraum]**.** Ein Tripel $(\Omega, \mathcal{E}, P)$ wird Wahrscheinlichkeitsraum genannt, wenn

(i) $\mathcal{E}$ eine σ-Algebra ist, d.h.
   - $\emptyset \in \mathcal{E}, \Omega \in \mathcal{E}$
   - $E \in \mathcal{E} \Rightarrow E^c \in \mathcal{E}$
   - $E_1, E_2, \ldots \in \mathcal{E} \Rightarrow \bigcup\limits_{i \geq 1} E_i \in \mathcal{E}$

(ii) $P : \mathcal{E} \to [0, 1]$ ist ein Wahrscheinlichkeits-Maß, d.h.
   a) $P(\Omega) = 1$ (Normierung)
   b) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, wenn $E_1 \cap E_2 = \emptyset$ (σ-Additivität)

   (Axiome der Wahrscheinlichkeitstheorie nach Kolmogorov, 1933)

1.2.2 Bemerkung[zur σ-Algebra]**.**

- Ist $\mathcal{E}_0 \subset 2^{\Omega}$, dann gibt es eine kleinste σ-Algebra $[\mathcal{E}_0]$, die $\mathcal{E}_0$ enthält, d.h. wenn $\mathcal{E}' \subset 2^{\Omega}$ eine σ-Algebra mit $\mathcal{E}_0 \subset \mathcal{E}'$, dann ist $[\mathcal{E}_0] \subset \mathcal{E}'$. $[\mathcal{E}_0]$ wird die von $\mathcal{E}_0$ erzeugte σ-Algebra genannt. Bsp:

  $$\Omega = \{1, 2, 3, 4, 5, 6\}, E_0 = \{1, 3, 5\} \Rightarrow [E_0] = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$$

  Allgemeiner: $[E_0] = \{\emptyset, E_0, E_0^c, \Omega\}$.

- Ein wichtiges Beispiel ist die Borel-σ-Algebra auf $\mathbb{R}$, die die Borelmengen enthält, und die von den offenen Intervallen auf $\mathbb{R}$ erzeugt wird. Praktisch alle Mengen, die man hinschreiben kann, sind Borelmengen, sie sind Lebesgue-Meßbar, deshalb sind sie wichtig.

- Zwar hat es den Anschein, als enthielte die σ-Algebra nicht mehr als Ω, $\emptyset$, Komplemente und abzählbare Vereinigungen, doch bei näherer Betrachtung enthält sie mehr:

  (i) endliche Vereinigungen $E \cup F = E \cup F \cup \emptyset \cup \emptyset \cup \ldots$

(ii) endliche / abzählbare Durchschnitte: $E_1 \cap E_2 \cap E_3 \cap \ldots = (E_1^c \cup E_2^c \cup E_3^c \cup \ldots)^c$

(iii) Relativkomplemente: $E, F \in \mathcal{E}, E \subset F \Rightarrow F \setminus E = F \cap E^c$

1.2.3 Notiz[Rechenregeln für Wahrscheinlichkeiten]**.** Unmittelbare Folgerungen aus den Kolmogorov-Axiomen:

(i) $P(\emptyset) = 0$

(ii) $P(E) + P(E^c) = 1$

(iii) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

(iv) Monotonie: $E \subset F \Rightarrow P(E) \leq P(F)$, da $P(F) = P(E) + P(F \setminus E) \geq P(E)$

1.2.4 Beispiel.

- Würfeln: $\Omega = \{1, 2, 3, 4, 5, 6\}, \mathcal{E} = 2^\Omega$, P als Gleichverteilung auf Ω.
- Münzwurf: $\Omega = \{0, 1\}, P(\{0\} = p, P(\{1\} = 1 - p))$.
- $\Omega = [0, 1] \subset \mathbb{R}, \mathcal{E} = \mathcal{B}(\Omega)$ Borel-σ-Algebra, $a, b \in \Omega, P([a, b]) = b - a$. Wahrscheinlichkeit ist Länge des Intervalls, also Gleichverteilung.

1.2.5 Bemerkung. Ist Ω eine überabzählbare Menge, dann ist 2^Ω "zu groß", d.h. es kann i.A. kein Wahlscheinlichkeitsmaß definiert werden, das die Kolmogorov-Axiome erfüllt.

1.2.6 Beobachtung. Praktisch alle elementaren Rechenregeln für Wahrscheinlichkeiten folgen aus den Axiomen.

1.2.7 Lemma. Es sei $E_1 \subset E_2 \subset \ldots$ eine aufsteigende Kette. Dann gilt

$$P(\bigcup_{n=1}^{\infty} E_n) = \lim_{n\to\infty} P(E_n)$$

Andersrum: $E_1 \supset E_2 \supset \ldots$, so gilt

$$P(\bigcap_{n=1}^{\infty} E_n) = \lim_{n\to\infty} P(E_n)$$

1.2.8 Bemerkung[Mathematisches Zählen]**.** Wir kennen bereits die Regel $P(E \cup F) = P(E) + P(F) - P(E \cap F)$. Wie sieht das Ganze für n Mengen aus?

1.2.9 Lemma[Einschluss-Ausschluss-Prinzip]**.** Es gilt

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 \leq \ldots \leq i_k \leq n} P(E_{i_1} \cap \ldots \cap E_{i_k})$$

1.2.10 Beispiel[Beispiel mit E_1, E_2, E_3]**.**

$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$

1.3 diskrete Wahrscheinlichkeitsräume

1.3.1 Definition[diskreter Wahrscheinlichkeitsraum]**.** Ist Ω endlich oder abzählbar, so wird $(\Omega, \mathcal{E}, P)$ diskreter Wahrscheinlichkeitsraum genannt.

1.3.2 Bemerkung[Vorbemerkungen]**.**

- Im diskreten Fall wählen wir $\mathcal{E} = 2^\Omega$, d.h. jeder Teilmenge $E \in 2^\Omega$ ist ein zulässiges Ereignis.
- Nun $P : 2^\Omega \to [0,1]$, aber schon im endlichen Fall ist 2^Ω recht groß ($|2^\Omega| = 2^{|\Omega|}$)
- Die gute Nachricht ist: Nur $P(\{\omega\})$ müssen gegeben sein aufgrund der σ-Additivität.

1.3.3 Beispiel[Beispiele für diskrete Wahrscheinlichkeitsräume]**.**

(i) Gleichverteilung (Laplaceraum):
Ω endlich mit $|\Omega| = n$ und alle $\omega \in \Omega$ sind gleich wahrscheinlich $\Rightarrow P(\{\omega\}) = \frac{1}{n}$.
Wahrscheinlichkeit eines Ereignisses $E \in 2^\Omega$: $P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{n}$
Beispiele: Faire Münze ($n = 2$), Würfel ($n = 6$), Lotto ($n = 49$)

(ii) Binomialverteilung (Bernoulliraum):

- Experiment mit 2 möglichen Ausgängen $\{0,1\}$ (Misserfolg / Erfolg) mit $P(\{1\}) = p$ und $P(\{0\}) = 1 - p$.
- Wahrscheinlichkeit für k Erfolge bei n Wiederholungen (Bernoulliprozess):

$$P_n(\{k\}) = B(k,n) = \binom{n}{k} p^k (1-p)^{n-k} \text{ mit } \binom{n}{k} = \frac{n!}{(n-k)!k!} \text{ (Binomialverteilung)}$$

- Passende Menge von Elementarereignissen: $\Omega = \{0,1\}^n$, $\Omega = \{1, \ldots, n\}$ ("sparsamer")
- Beispiel Lotto: $\binom{49}{6} \approx 14$ Millionen. Spezialfall: $n = 1$, $k = 0, 1$

$$\Rightarrow P_1(\{k\}) = \begin{cases} p & \text{, falls } k = 1 \\ 1 - p & \text{, falls } k = 0 \end{cases}$$

  Verschiedene Grenzfälle für $n \to \infty$.
- Beispiel Geburtstagsparadoxon 1: Wie groß ist die Wahrscheinlichkeit, dass jemand der Anwesenden am 30.5. Geburtstag hat? $n = 150$ (# Leute), $p = \frac{1}{365}$.

$$P_{150}(\{0\}) \approx 66.50\%, P_{150}(\{1\}) \approx 27.25\%, P_{150}(\{2\}) \approx 6\%$$

$$P(\{\text{mindestens ein Teilnehmer}\}) = 1 - P_{150}(\{0\}) \approx 33.7\%$$

(iii) Poissonverteilung (Grenzfall der Binomialverteilung für $n \to \infty$ und $p \to 0$):
Modellierung von "seltenen Ereignissen". $\Omega = \mathbb{N}_0 = \{0,1,2,\ldots\}$, $P_\lambda(\{k\}) = P(k,\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$ mit $\lambda = n \cdot p$ (Rate).
Beispiele: Radioaktive Zerfälle, Fehler pro Vorlesung, ...

(iv) Geometrische Verteilung (Wartezeiten):
Bernoulli-prozess mit der Frage: Wie groß ist die Wahrscheinlichkeit, dass beim n-ten Versuch Erfolg eintritt?

$\Omega = \mathbb{N}, P_{\text{geo}}(\{n\}) = G(n) = (1-q)q^{n-1} = p(1-p)^{n-1}$ mit $q = 1-p$ (Misserfolgswahrscheinlichkeit)

Beispiel: Wahrscheinlichkeit für gerades n?

$$\begin{aligned} P(\{\text{n gerade}\}) &= P(\{2,4,6,\dots\}) \overset{\sigma\text{-Additivität}}{=} (1-q)q + (1-q)q^3 + \dots \\ &= (1-q)q\sum_{i\geq 0} q^{2i} = (1-q)q\frac{1}{1-q^2} \\ &= \frac{q}{1+q} \end{aligned}$$

$$P(\{\text{n ungerade}\}) = 1 - P(\{\text{n gerade}\}) = \frac{1}{1+q} > \frac{q}{1+q} \qquad \forall q \in [0,1)$$

1.3.4 Bemerkung[abschließende Bemerkungen]**.**

a) Im höchstens abzählbahren Fall reicht, die $P(\{\omega\})$ zu kennen.
b) Bei der Wahl des richtigen Wahscheinlichkeitsraums gibt es meist verschiedene Möglichkeiten.
c) Der Rest ist geschicktes Zählen.

1.3.5 Definition[Urnenmodelle (Elementare Kombinatorik, Behrends-Skript S.55-58)]**.**

- Urne mit unnummerierten Kugeln $\{1,\dots,N\}$, von denen $n \in \mathbb{N}$ gezogen werden
- Ziehung ist ein Wort $a_1a_2\dots a_n, a_i \in \{1,\dots,N\}$
- Stichprobenraum $\Omega = \{$Menge aller Ziehungen$\}$ Vier Fälle lassen sich unterscheiden:
  (i) Ziehen mit Zurücklegen
    a) Unter Berücksichtigung der Reihenfolge: $|\Omega| = N^n$
    b) Ohne Berücksichtigung der Reihenfolge: $|\Omega| = \binom{N+n-1}{n}$
  (ii) Ziehen ohne Zurücklegen
    a) Unter Berücksichtigung der Reihenfolge: $|\Omega| = N(N-1)\dots(N-n+1)$. Spezialfall: $n = N \Rightarrow |\Omega| = N!$ (gibt die Anzahl der möglichen Permutationen von $\{1,\dots,N\}$ an)
    b) Ohne Berücksichtigung der Reihenfolge: $|\Omega| = \binom{N}{n}$. D.h. $|\Omega|$ ist die Anzahl der n-elementigen Teilmengen von $\{1,\dots,n\}$.

1.3.6 Beispiel[Geburtstagsparadoxon 2]**.** Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 Leute aus der Vorlesung am gleichen Tag Geburtstag haben?
Modellierung à la Laplace:

$$P = \frac{\#\text{günstige Ausgänge}}{\#\text{mögliche Ausgänge}}, N = 365, n = 150.$$

$$\begin{aligned} P(\{\text{keine 2 Personen}\}) &= \frac{N(N-1)\dots(N-n+1)}{N^n} \approx 0 \\ &\Rightarrow P(\{\text{mindestens 2 Personen}\}) = 1 - P(\{\text{keine 2 Personen}\}) \approx 1 \end{aligned}$$

Bemerkung: Bei ca. $n = 50$ Personen beträgt $P(\{\text{mindestens 2 Personen}\}) \approx 99.9\%$.

1.4 Hypergeometrische Verteilung

1.4.1 Definition[Hypergeometrische Verteilung]**.**

- "vergröbertes" Urnenmodell. N Kugeln, von denen r rot und $(N-r)$ weiß sind.
- Wir ziehen n Kugeln ohne Zurücklegen (Typ: "Ziehen ohne Zurücklegen ohne Berücksichtigung der Reihenfolge").
- Wie groß ist die Wahrscheinlichkeit, dass unter den n gezogenen Kugeln k rote sind?
- Modellierung nach Laplace:

$$\begin{aligned} P &= \frac{\#\text{günstige Möglichkeiten}}{\#\text{alle Möglichkeiten}} = \frac{\#\text{k rote und N-k weiße}}{\binom{N}{n}} \\ &= \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}} = H(k,n;N,r) \end{aligned}$$

- $\sum\limits_k H(k,\_;\_,\_) = 1.$

1.4.2 Beispiel[Lotto ohne Schickschnack]**.** $P(\text{k richtige}) = H(k,6,49,6)$

1.4.3 Beispiel. Wahrscheinlichkeit für k Asse beim Skat?

$$P(\text{k Asse}) = H(k,10,32,4) = \begin{cases} 20\% & k=0 \\ 43\% & k=1 \\ 29\% & k=2 \\ 7\% & k=3 \\ 1\% & k=4 \end{cases}$$

1.4.4 Bemerkung.

(i) Die hypergeometrische Verteilung lässt sich leicht auf mehr als 2 Farben (Eigenschaften) verallgemeinern (siehe 2. Übungszettel).

(ii) Ist $N >> n$, so besteht kein Unterschied zwischen Ziehen mit und ohne Zurücklegen.

1.5 Wahrscheinlichkeitsräume mit Dichten

1.5.1 Bemerkung.

(i) Ω höchstens abzählbar $(P(E), E \in \mathcal{E} = 2^\Omega)$ ist durch Abzählen bestimmbar!

(ii) Ω überabzählbar: $P(E)$ ist eine Art von Volumen $\Rightarrow$ Abzählen / Summieren wird zur Integration

1.5.2 Definition[Wahrscheinlichkeitsräume mit Dichten]**.** Wir betrachten Wahrscheinlichkeitsräume $(\Omega, \mathcal{E}, P)$, wobei

(i) $\Omega \subset \mathbb{R}^n$

(ii) $\mathcal{E} \subset 2^\Omega$ die σ-Algebra der Borelmengen ist

(iii) P ein Wahrscheinlichkeitsmaß ist, das durch seine Dichte definiert ist, d.h.

$$\forall E \in \mathcal{E} : P(E) = \int\limits_{E \subset \Omega} f(x)dx, f \geq 0 \text{ (Wahrscheinlichkeitsdichte)}$$

1.5.3 Bemerkung.

(i) Die Borel-σ-Algebra ist die kleinste σ-Algebra, die von den offenen Mengen in Ω erzeugt wird.

(ii) Da f ein Wahrscheinlichkeitsmaß definiert, gilt:

- $P(\Omega) = \int_\Omega f = 1$
- Für $E_1, E_2, \ldots \in \mathcal{E}$ paarweise disjunkt gilt $\int\limits_{E_1 \cup E_2 \cup \ldots} f = \sum\limits_i \int\limits_{E_i} f$

1.5.4 Bemerkung.

- Es reicht anzunehmen, dass f stückweise stetig bzw. mindestens integrabel ist
- Die Dichte f übernimmt hier die Rolle von $P(\{\omega\}), \omega \in \Omega$ im höchstens abzählbaren Fall
- Ist Ω überabzählbar, ist $P(\{\omega\}) = 0 \quad \forall \omega \in \Omega$

1.5.5 Beispiel[Beispiele für Wahrscheinlichkeitsdichten]**.**

(i) Gleichverteilung
Es sei $\Omega = [a, b] \subset \mathbb{R}$ und $f(x) = (b - a)^(- 1)$ eine Wahrscheinlichkeitsdichte. f definiert ein Wahrscheinlichkeitsmaß, denn

- $P(E) = \int_E f(x)dx$ ist σ-additiv wegen der σ-additivität des Lebesguemaßes dx und es gilt $P(\Omega) = 1$
- Für alle Ereignisse $E \in \mathcal{E}$ bezeichnet $P(E)$ die "relative Länge" von $E \subset \Omega$:

$$P([c, d] \subset [a, b]) = \int\limits_c^d (b - a)^{-1} = \frac{d - c}{b - a}$$

- Ist Ω beschränkt, so definiert $f(x) = (\int_\Omega dx)^{-1}$ eine Wahrscheinlichkeitsdichte
- Ist Ω nicht beschränkt, so existiert keine Gleichverteilung, denn $(\int_\Omega dx)^{-1} = 0$.

1.5.6 Bemerkung. Es ist möglich, abzählbar viele Punktereignisse $\omega \in \Omega$ zu entfernen, d.h. von Ω zu $\Omega \setminus E$ überzugehen, ohne den Wahrscheinlichkeitsraum wesentlich zu ändern

(ii) Normalverteilung (Gaußverteilung)
Es sei $\Omega = \mathbb{R}$ und $f \geq 0$ die Dichtefunktion

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

(Gaußdichte zu den Parametern m und σ) f ist normiert, d.h. es gilt $\int\limits_{-\infty}^{\infty} f(x)dx = 1$

(iii) Exponentialverteilung:
Es sei $\Omega = [0, \infty)$ und f die Wahrscheinlichkeitsdichte $f(x) = \lambda e^{-\lambda x}$.

- Die Exponentialverteilung beschreibt die Verteilung der Wartezeiten zwischen 2 seltenen Ereignissen, die durch die Poisson-Verteilung mit Parameter $\lambda > 0$ gegeben sind.
- Die Verteilungsfunktion

$$F(\tau) = P(x < \tau) = \lambda \int_0^\tau e^{-\lambda x} dx$$

beschränkt die Wartezeitverteilung bis zum nächsten Ereignis. Z.B.

$$P(a \leq x \leq b) = F(b) - F(a) = e^{-\lambda b} - e^{-\lambda a}$$

1.6 Bedingte Wahrscheinlichkeiten

1.6.1 Definition[bedingte Wahrscheinlichkeit]**.** Die bedingte Wahrscheinlichkeit von E unter der Annahme F ist

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, P(F) > 0$$

1.6.2 Bemerkung[Interpretation]**.** Bedingte Wahrscheinlichkeit ist das relative Maß von E in F. Informationen verändern Wahrscheinlichkeiten.

1.6.3 Definition[Unabhängigkeit von Ereignissen]**.** Zwei Ereignisse $E, F \in \mathcal{E}$ heißen unabhängig, wenn

$$P(E \cap F) = P(E) \cdot P(F)$$

bzw. (nach Definition von bedingten Wahrscheinlichkeiten), wenn $P(E|F) = P(E), P(F|E) = P(F)$.

1.6.4 Notiz. Unabhängigkeit bedeutet, dass die Kenntnis von F ihnen keine Informationen über die Wahrscheinlichkeit von E liefert (bzw. umgekehrt).

1.6.5 Satz. E, F unabhängig $\Leftrightarrow E, \Omega \setminus F = F^c$ unabhängig.

1.6.6 Beispiel.

- Würfeln: Es sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $E = \{\text{Zahl ist gerade}\}$.
  (i) $F = \{\text{Zahl } \geq 3\}$. Dann ist $P(E) = \frac{1}{2}, P(F) = \frac{2}{3}, P(E \cap F) = \frac{1}{3} = P(E)P(F)$. Also sind E und F unabhängig.
  (ii) $F' = \{\text{Zahl } \geq 4\}$, d.h. $P(F') = \frac{1}{2} \neq \frac{P(E \cap F')}{P(E)}$. $\Rightarrow E, F'$ sind abhängig (F' enthält mehr gerade als ungerade Zahlen).
- Noch einmal Würfeln: Gefragt ist nach der Summe der Augenzahlen bei 2 mal Würfeln. Es sei $E = \{i + j \geq 7 \mid i, j \in \{1, \dots, 6\}\}$. (Z.B. $(6, 1), (6, 2), \dots$ insgesamt 21 aus 36 Möglichkeiten).
  (i) $F = \{i = 5, j\}$, d.h. 5 von 6 Möglichkeiten bleiben $\Leftrightarrow P(E|F) = \frac{5}{6}$. Zum Vergleich: $P(E \cap F) = P(E|F)P(F) = \frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$.

(ii) $F' = \{i \leq 3, j \leq 3\} \Rightarrow P(E|F') = 0$, da $E \cap F' = \emptyset$.

(iii) $F'' = \{i \geq 4, j \geq 4\} \Rightarrow P(E|F) = 1$ $(F \subset E)$.

1.6.7 Satz[Rechenregeln für bedingte Wahrscheinlichkeiten]**.** Es seien $E, F, E_1, E_2, \ldots$ in einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, P)$ und wir nehmen an, dass $P(F) > 0$. Dann gelten:

(i) $E \cap F = \emptyset \Rightarrow P(E|F) = 0$

(ii) $F \subset E \Rightarrow P(E|F) = 1$

(iii) $P(\Omega \setminus E|F) = 1 - P(E|F)$

(iv) $\forall i, j, i \neq j : E_i \cap E_j = \emptyset$, dann ist $P(\bigcup_k E_k|F) = \sum_k P(E_k|F)$ (σ-Additivität)

1.6.8 Bemerkung. Man kann das Wahrscheinlichkeitsmaß $P_F = P(\cdot|F)$ als Wahrscheinlichkeitsmaß auf der Spur-σ-Algebra auffassen:

$$\mathcal{E}_F = \{E \cap F \mid E \in \mathcal{E}\}$$

1.6.9 Satz[Satz von der totalen Wahrscheinlichkeit]**.** Es sei $\Omega = \bigcup_n E_n$ eine disjunkte Vereinigung von Teilmengen. Dann gilt

$$P(F) = \sum_n P(F|E_n) \cdot P(E_n) \ \forall F \in \mathcal{E}$$

Beweis: $F = \bigcup_n (F \cap E_n)$ ist die disjunkte Vereinigung der $F \cup E_n$.

$$P(F) = P(\bigcup_n (F \cap E_n)) = \sum_n P(F \cap E_n) = \sum_n \frac{P(F \cap E_n)P(E_n)}{P(E_n)} = \sum_n P(F|E_n)P(E_n)$$

□

1.6.10 Satz[Satz von Bayes]**.** Der Satz von Byes ist einer der bekanntesten und wichtigsten Sätze der Wahrscheinlichkeitstheorie und der Statistik. Wieder sei $\Omega = \bigcup_n E_n$ eine disjunkte Vereinigung und $F \in \mathcal{E}$ mit $P(F) > 0$. Dann gilt

$$P(E_n|F) = \frac{P(E_n)P(F|E_n)}{P(F)} = \frac{P(E_n)P(F|E_n)}{\sum_n P(E_n)P(F|E_n)}$$

Beweis:

$$\begin{aligned} P(E_n|F) &= \frac{P(E_n \cap F)}{P(F)} \\ &= \frac{P(F|E_n)P(E_n)}{P(F)} && \text{, nach Definition von bedingter Wahrscheinlichkeit} \\ &= \frac{P(F|E_n)P(E_n)}{\sum_n P(F|E_n)P(E_n)} && \text{, nach Satz der totalen Wahrscheinlichkeit} \end{aligned}$$

□

1.6.11 Beispiel. Zwei Kartenstapel mit einen Würfel, wobei Stapel 1 2 rote und 8 schwarze Karten hat, Stapel 2 hat 5 rote und 5 schwarze Karten. Bei Augenzahlen $\{1, 2, 3, 4\}$ wird von Stapel 1 gezogen, sonst von Stapel 2.

(i) Wie wahrscheinlichkeit ist es, dass eine rote Karte gezogen wird? D.h. $F = \{\text{rote Karte}\}$ und

$$P(F) = P(F|E_1 = \{1,2,3,4\})P(E_1) + P(F|E_2 = \{5,6\})P(E_2) = \frac{1}{5} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{3}{10}$$

(ii) Eine rote Karte wird gezogen. Wie wahrscheinlich ist es, dass sie vom Stapel 1 gezogen wurde? Mit Satz von Bayes folgt:

$$P(E_1|F) = \frac{P(E_1 \cap F)}{P(F)} = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{\frac{1}{5} \cdot \frac{2}{3}}{\frac{3}{10}} = \frac{4}{9}$$

1.6.12 Bemerkung[Anwendungen der Bayesschen Formel]**.**

(i) Bestätigung von Theorien: In welchem Maße stützt oder widerlegt die Beobachtung B eines schwarzen Raben die Theorie $T =$ "Alle Raben sind schwarz"?

$$\begin{aligned} P(T|B) &= \frac{P(B|T)P(T)}{P(B)} \\ &= \frac{P(B|T)P(T)}{P(B|T)P(T) + P(B|\neg T)P(\neg T)} \end{aligned}$$

Gegeben sind folgende Informationen:

- $P(B|T) = 1$, da Theorie wahr $\Rightarrow$ die Beobachtung eines schwarzen Raben verwundert nicht.
- $P(T) = P(\neg T) = \frac{1}{2}$
- $P(B|\neg T) = \varepsilon$

Mit zum Beispiel $\varepsilon = \frac{1}{2}$ folgt:

$$P(T|B) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4}} = \frac{2}{3} > P(T) = \frac{1}{2}$$

Vergleiche Curd / Cover (1998), "The Philosophy of Science".

(ii) Medizinische Tests: Es sei T ein Test auf eine bestimmte (seltene) Krankheit K, die mit der Wahrscheinlichkeit $P(K)$ auftritt (Prävalenz). Der Test wird spezifisch genannt, $\neg T$) mit Wahrscheinlichkeit nahe 1 erhält. Die Spezifität ist $P(\neg T|\neg K) = 1 - P(T|\neg K)$. Nach Bayes:

$$\begin{aligned} P(\neg K|T) &= \frac{P(T|\neg K)P(\neg K)}{P(T)} \\ &= \frac{P(T|\neg K)P(\neg K)}{P(T|\neg K)P(\neg K) + P(T|K)P(K)} \end{aligned}$$

Gegeben sind:

- Die Prävalenz $P(K)$ ohne Dunkelziffer ist $P(\neg K) = 1 - P(K)$.

- Ebenso bekannt (aus Daten): $P(T|K) \approx 1$ (Zuverlässigkeit)
- $P(T|\neg K) \approx 0$ ("false positives")

Zahlenbeispiel: Siehe Aigner-Skript.

1.6.13 Beispiel[Ziegenproblem, "Ask Marilyn"]**.** Das Ziegenproblem: Der Kandidat wählt Tür 1 von 3. Der Moderator weiß was hinter den Türen steckt und öffnet Tür 3 mit einer Ziege dahinter. Der Kandidat wird gefragt, ob er bei seiner Wahl bleiben möchte. Die Frage ist: Ist es sinnvoll sich umzuentscheiden oder ist es egal?
Gegeben: $B_i = \{\text{Auto hinter Tür } i\}, i \in \{1,2,3\}$,
$A = \{\text{Moderator zeigt ihnen, dass hinter Tür 3 eine Ziege steht}\}$
Gesucht: $P(B_1|A)$ bzw. $P(B_2|A)$.
Nach dem Satz der totalen Wahrscheinlichkeit: $P(A) = \sum\limits_i P(A|B_i)P(B_i)$, wobei $P(B_i) = \frac{1}{3}$, $P(A|B_1) = \frac{1}{2}$, $P(A|B_2) = 1$, $P(A|B_3) = 0$. $\Rightarrow P(A) = \frac{1}{6} + \frac{1}{3} + 0 = \frac{1}{2}$. Also (nach Bayes):

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{1}{3}$$
$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{2}{3}$$

1.6.14 Bemerkung. Wenn der Moderator eine Präferenz für Tür 3 hat, ändern sich die Wahrscheinlichkeiten auf $P(B_1|A) = P(B_2|A) = \frac{1}{2}$.

1.7 Unabhängigkeit mehrerer Ereignisse

1.7.1 Bemerkung. Bisher für 2 Ereignisse: $P(E \cap F) = P(E)P(F)$. Wie verhält sich Unabhängigkeit bei 3 oder mehr Ereignissen?

1.7.2 Beispiel. Zweifacher Münzwurf mit Ereignissen

$$E = \{\text{K, Z}\} \cup \{\text{Z, K}\}, F = \{\text{K}, n\}, G = \{m, \text{Z}\}$$

(i) E, F, G sind paarweise unabhängig:

$$P(E \cap F) = P(E)P(F), P(E \cap G) = P(E)P(G), \ldots$$

(ii) $P(E \cap F \cap G) \neq P(E)P(F \cap G)$.

1.7.3 Definition[Unabhängigkeit mehrerer Ereignisse]**.** Die Ereignisse $E_1, \ldots, E_n \in \mathcal{E}$ heißen <u>unabhängig</u>, wenn

$$P(E_{i_1} \cap \ldots \cap E_{i_r}) = \prod_{k=1}^{r} P(E_{i_k})$$

für alle $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$

1.7.4 Bemerkung.

(i) Falls $E_1, \ldots, E_n$ unabhängig sind, so sind sie auch paarweise unabhängig. Die Umkehrung gilt <u>nicht</u>.

(ii) Durch Induktion zeigt man leicht:

$$P(E_1 \cap \ldots \cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \cdot \ldots \cdot P(E_n|E_1 \cap \ldots \cap E_{n-1})$$

(iii) Die Definition von Unabhängigkeit ist äquivalent zu:

$$\begin{aligned} &E_1 \text{ ist unabhängig von } E_2, \ldots, E_n \\ \text{und } &E_2 \text{ ist unabhängig von } E_1, E_3, \ldots, E_n \\ \text{und } &\ldots \\ \text{und } &E_n \text{ ist unabhängig von } E_1, \ldots, E_{n-1} \end{aligned}$$

wobei F ist unabhängig von $E_1, \ldots, E_k$ für alle $[E_1, \ldots, E_k]$ getestet werden muss.

2 Zufallsvariablen

2.1 Induzierter Raum und Verteilung

Wir kommen zum wichtigsten Begriff der Wahrscheinlichkeitstheorie. In den meisten Situationen ist man nicht an der gesamten Verteilung interessiert, sondern nur an bestimmten Daten, die vom Experiment abhängen.

2.1.1 Beispiel. 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.

2.1.2 Definition. Sei $(\Omega, \mathcal{E}, P)$ ein Wahrscheinlichkeitsraum, $X : \Omega \to W$, $\mathcal{F}$ 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 abzählbar oder endlich ist, so sei $\mathcal{F} = 2^W$. Wir werden fast ausschließlich $W \subset \mathbb{R}$ betrachten, $\mathcal{F}$ = Borelmengen. X heißt dann "reelle Zufallsvariable".

2.1.3 Definition[inducierter Wahrscheinlichkeitsraum]**.** Der durch X induzierte Wahrscheinlichkeitsraum ist $(W, \mathcal{F}, P_X)$ mit $P_X(F) = P(X^{-1}(F))$. Wir schreiben kurz $P(X \in F)$ für $P(\{\omega \in \Omega \mid X(\omega) \in F\})$. Zum Beispiel $P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$ oder $P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\})$, falls $W = \mathbb{R}$.

2.1.4 Satz. P_X ist Wahrscheinlichkeitsmaß auf $(W, \mathcal{F})$.

Beweis:

- $P_X(W) = P(X^{-1}(W)) = P(\Omega) = 1$
- $P_X(\bigcup_i F_i) = P(X^{-1}(\bigcup_i F_i)) = P(\bigcup_i X^{-1}(F_i)) = \sum_i P(X^{-1}(F_i)) = \sum_i P_X(F_i)$

□

2.1.5 Beispiel.

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

(ii) Zweimaliges Würfeln, $\Omega = \{1,2,3,4,5,6\}^2$ mit Gleichverteilung. Sei X = Augensumme, $W = \{1,2,\ldots,12\}$. Dann ist

$$P(X = s) = \frac{1}{36} \cdot \#\{(i,j) \in \Omega \mid i + j = s\}$$

zum Beispiel $P(X = 10) = \frac{3}{36} = \frac{1}{12}$.

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

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\})$$

Offenbar gilt

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

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

2.1.7 Beispiel.

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

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

(ii) 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

$$F_{I_A}(x) = \begin{cases} 0 & x < 0 \\ 1 - P(A) & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

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 $X_1, \ldots, X_n : \Omega \to \mathbb{R}$, dann ist $(X_1, \ldots, X_n) : \Omega^n \to \mathbb{R}^n$. Die gemeinsame Verteilung ist

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

2.2 Diskrete Zufallsvariablen

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

2.2.1 Definition[Unabhängigkeit]**.** Seien $X, Y : \Omega \to W \subset \mathbb{R}$ reelle Zufallsvariablen. Wir werden X und Y als unabhängig ansehen, falls

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

für alls $A, B \in W$ gilt.

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

$$P(X = x \cap Y = y) = P(X = x)P(Y = y) \quad \forall x, y \in W$$

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

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

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

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

$$\mathbb{E}(X) = \sum_{x} xP(X = x) = \sum_{\omega \in \Omega} X(\omega)P(\omega)$$

falls die Summe absolut konvergiert.

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

$$X(\omega) \leq \mathbb{E}(X), X(\omega') \geq \mathbb{E}(X).$$

Für endliche Mengen Ω oder endlichem Wertebereich existiert $\mathbb{E}(X)$ natürlich immer.

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

$$\mathbb{E}(X) = \frac{1}{|\Omega|} \sum_{\omega} X(\omega)$$

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

$$\mathbb{E}(X) = \frac{x_1 + \ldots + x_n}{n}$$

schreiben. $\mathbb{E}(X)$ ist somit der übliche Durchschnittswert.

2.2.6 Beispiel[Gleichverteilung]**.**

- Sei $\Omega = \{\omega_1, \ldots, \omega_n\}$ und $X(\omega_i) = x_i$ (gleichverteilt). Dann ist $\mathbb{E}(X) = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$.

- Sei nun $Y = X^2$ $(y = x^2)$. dann ist

$$\mathbb{E}(Y) = \mathbb{E}(X^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

2.3 Momente von Zufallsvariablen (ZV)

2.3.1 Definition[Momente einer Zufallsvariablen]**.**

(i) Es sei $k \in \mathbb{N}$ das k-te Moment von X ist $m_k = \mathbb{E}(X^k)$

(ii) Das k-te zentrale (zentrierte) Moment von X ist $\sigma_k = \mathbb{E}((x - \mathbb{E}(X))^k)$

2.3.2 Bemerkung.

- Wichtige Spezialfälle sind der Erwartungswert $m_1 = \mathbb{E}(X)$ und die Varianz $\sigma_2 = \mathbb{E}((x - \mathbb{E}(X))^2)$
- Ist eine Verteilung multimodal (d.h. mehrere Maxima in der Verteilung), so sind Erwartungswert und Varianz nicht sehr aussagekräftig.
- Gaußsche Zufallsvariablen sind solche, die durch eine Gaußverteilung mit Dichte $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ erzeugt werden. Gaußsche Zufallsvariablen sind vollständig durch ihre ersten beiden Momente $m = m_1$, $\sigma = \sigma_2$ bestimmt.

2.4 Eigenschaften der Momente

2.4.1 Satz[Linearität des Erwartungswertes]**.** Es seien $X, X_1, \ldots, X_n : \Omega \to W$ Zufallsvariablen auf Ω mit der Eigenschaft $X = \alpha_1 X_1 + \ldots \alpha_n X_n$. Dann gilt

$$\mathbb{E}(X) = \sum_{i=1}^{n} \alpha_i \mathbb{E}(X_i), \alpha_i \in \mathbb{R}$$

Beweis: Nach Definition des Erwartungswertes gilt

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{\omega \in \Omega} \sum_{i=1}^{n} \alpha_i X_i(\omega)P(\omega) = \sum_{i=1}^{n} \alpha_i \sum_{\omega \in \Omega} X_i(\omega)P(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{E}(X)$$

□

2.4.2 Korollar. Es gilt

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

2.4.3 Satz[Unabhängigkeit von Zufallsvariablen]**.** Es seien $X : \Omega \to V \subset \mathbb{R}$ und $Y : \Omega \to W \subset \mathbb{R}$ zwei unabhängige Zufallsvariablen auf Ω. Dann gilt

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

Beweis: Nach Definition der Unabhängigkeit gilt

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{\omega \in \Omega} X(\omega)Y(\omega)P(\omega) = \sum_{(x,y) \in V \times W} xyP((X = x) \cap (Y = y)) \\
&= \sum_{(x,y) \in V \times W} xyP(X = x)P(Y = y) && \text{nach Unabhängigkeit} \\
&= (\sum_{x \in V} xP(X = x))(\sum_{y \in W} yP(Y = y)) = \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

□

2.4.4 Korollar. Sind X, Y unabhängige Zufallsvariablen, so gilt

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Beweis: Unter Verwendung der Eigenschaften des Erwartungswertes (EW) gilt

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}((X+Y)^2) - (\mathbb{E}(X+Y))^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) + \mathbb{E}(XY) + \mathbb{E}(Y^2) - (\mathbb{E}(X))^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - (\mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 + \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y)\end{aligned}$$

□

2.4.5 Bemerkung.

- Die Umkehrung des Satzes über die Unabhängigkeit von Zufallsvariablen und seinem Korollar gilt nicht, eine Ausnahme sind die Gaußschen Zufallsvariablen.
- Unter linearen Transformationen $X \mapsto aX + b$ transformieren sich die ersten beiden Momente wie
$$\mathbb{E}(X) \mapsto a\mathbb{E}(X) + b, \quad \mathrm{Var}(X) \mapsto a^2\mathrm{Var}(X)$$
- Ist für eine Zufallsvariable X die Varianz $\mathrm{Var}(X) = 0$, so ist $P(X = \mathbb{E}(X)) = 1$. Die Umkehrung gilt auch.

2.4.6 Beispiel.

(i) Momente der Bernoulliverteilung (Münzwurf).

- $\Omega = \{k, z\}$ und $P(k) = p \in [0,1]$, $P(z) = 1 - p$.
- Zufallsvariable $X : \Omega \to \{0,1\}$ mit $X(k) = 1$, $X(z) = 0$ (d.h. $P(X = 1) = p$, $P(X = 0) = 1 - p$).

$$\Rightarrow \begin{cases} \text{Erwartungswert } \mathbb{E}(X) = \sum\limits_x xP(X=x) = 0 \cdot (1-p) + 1 \cdot p = p \\ \text{2-tes Moment } \mathbb{E}(X^2) = \sum\limits_x x^2 P(X=x) = p \end{cases}$$

$$\Rightarrow \text{Varianz } \mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p - p^2$$

(ii) Unabhängige Wiederholungen (Münzwurf).

- Situation wie in (i), aber wir führen n unabhängige Wiederholungen aus.
- Zufallsvariable $X = \#\text{Kopfwürfe}$ und wir schreiben X als
$$X = \sum_{i=1}^{n} x_i, \text{ wobei } x_i = \begin{cases} 1 & \text{, Kopf im } i\text{-ten Wurf} \\ 0 & \text{, sonst} \end{cases}$$
- Wahrscheinlichkeit für k-mal Kopf ist
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$\Rightarrow \mathbb{E}(X) = np$$
$$\mathrm{Var}(X) = n(p - p^2)$$

2.4.7 Notiz. Im Folgenden sei $(\Omega, \mathcal{E}, P)$ ein Wahrscheinlichkeitsraum (Ω höchstens abzählbar) und $X : \Omega \to W \subset \mathbb{R}$ eine reele Zufallsvariable auf Ω.

2.4.8 Beispiel.

(i) Momente müssen nicht existieren, z.B. sei $\Omega = \mathbb{N}$, $X(\omega_k) = k$ mit $P(X = k) = \frac{c}{k^3}$. c ist so gewählt, dass $\sum\limits_{k\in\mathbb{N}} P(X = k) = 1$, d.h. $c = \sum\limits_{k\in\mathbb{N}} \frac{1}{k^3} = \zeta(3)$. Berechne 1. und 2. Moment:

- $\mathbb{E}(X) = \sum\limits_{k\in\mathbb{N}} k\frac{c}{k^3} = \sum\limits_{k\in\mathbb{N}} \frac{c}{k^2} = \frac{\zeta(2)}{\zeta(3)}$
- $\mathbb{E}(X^2) = \sum\limits_{k\in\mathbb{N}} k^2\frac{c}{k^3} = \sum\limits_{k\in\mathbb{N}} \frac{c}{k} = \infty$

(ii) Momente der Poissonverteilung

- $\Omega = \mathbb{N}_0$ und wir betrachten $P(X = n) = \frac{\lambda^n}{n!}e^{-\lambda}$. Berechne wieder 1. und 2. Moment:
  - $\mathbb{E}(X) = \sum\limits_{n=0}^{\infty} n\frac{\lambda^n}{n!}e^{-\lambda} = \lambda\sum\limits_{n=0}^{\infty} \frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda} = \lambda$
  - $\mathbb{E}(X^2) = \sum\limits_{n=0}^{\infty} n^2\frac{\lambda^n}{n!}e^{-\lambda} = \lambda + \lambda^2$, d.h $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
- Berechnung aus der Binomialverteilung:

$$\lim_{\substack{n\to\infty\\ np\to\lambda}} B(k, n, p) = \frac{\lambda^k}{k!}e^{-\lambda}$$

  Aus der letzten Vorlesung wissen wir, dass
  - $\mathbb{E}_B(X) = np, \mathrm{Var}_B(X) = n(p - p^2)$
  - $\mathbb{E}_B(X) \overset{n\to\infty}{\to} \mathbb{E}(X) = \lambda, \mathrm{Var}_B(X) \overset{n\to\infty}{\to} \mathrm{Var}(X) = \lambda$

2.5 Kovarianz und Korrelation

2.5.1 Definition[Kovarianz]**.** Es seien $X, Y : \Omega \to W \subset \mathbb{R}$ zwei Zufallsvariablen. Die Kovarianz von X und Y ist

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

2.5.2 Bemerkung[Kovarianz, Varianz und Unabhängigkeit]**.**

(i) Zwei Zufallsvariablen X, Y heißen unkorreliert, wenn $\mathrm{cor}(X, Y) = 0$. Sind X und Y unabhängig, so sind sie immer unkorreliert. Umkehrung gilt allerdings nicht.

(ii) Offensichtlich gilt $\mathrm{cor}(X, X) = \mathrm{Var}(X)$. Weiter gilt

$$\mathrm{Var}(\alpha X + \beta Y) = \alpha^2\mathrm{Var}(X) + \beta^2\mathrm{Var}(Y) + 2\alpha\beta\mathrm{cov}(X, Y)$$

(iii) $\mathrm{cov}(\cdot, \cdot) : W \times W \to \mathbb{R}$ definiert eine symmetrische, positiv semidefinite Bilinearform, d.h.

- $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$
- $\mathrm{cov}(X, X) \geq 0$

(iv) Mittels Kovarianz lassen sich Skalarprodukte und Normen definieren, für die die üblichen Identitäten gelten.

2.6 Wichtige Ungleichungen für Wahrscheinlichkeitsräume

2.6.1 Satz. Es seien $X, Y : \Omega \to W \subset \mathbb{R}$. Dann gilt die Cauchy-Schwarzsche Ungleichung

$$(\mathbb{E}(XY))^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

Gleichheit gilt, wenn $P(X = \alpha Y) = 1$ für $\alpha \in \mathbb{R}$

Beweis: Wir unterscheiden zwei Fälle:

(i) Sei $\mathbb{E}(X^2) = 0$. Das ist gleichbedeutend mit $P(X = 0) = 1$ bzw. $P(X \neq 0) = 0$. Das heißt, dass

$$\mathbb{E}(XY) = \sum_{(x,y)} xyP((X = x) \cap (Y = y)) = 0$$

Das heißt wir haben Gleichheit und $P(X = 0 \cdot Y) = 1$

(ii) Nun $\mathbb{E}(X^2) > 0$ und wir definieren die Hilfs-Zufallsvariable $Z := Y - \alpha X$, für die gilt

$$0 \leq \mathbb{E}(Z^2) = \mathbb{E}((Y - \alpha X)^2) = \alpha^2\mathbb{E}(X^2) - 2\alpha\mathbb{E}(XY) + \mathbb{E}(Y^2) \quad \forall \alpha \in \mathbb{R}$$

Die Ungleichung gilt insbesondere für $\alpha = \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$, *d.h.*

$$0 \leq \frac{(\mathbb{E}(XY))^2}{\mathbb{E}(X^2)} - 2\frac{(\mathbb{E}(XY))^2}{\mathbb{E}(X^2)} + \mathbb{E}(Y^2) = -\frac{(\mathbb{E}(XY))^2}{\mathbb{E}(X^2)} + \mathbb{E}(Y^2)$$

Das ist gerade die Cauchy-Schwarzsche Ungleichung, wobei Gleichheit genau dann gilt, wenn

$$P(Y = \alpha X) = 1.$$

□

2.6.2 Bemerkung. Anstelle der Kovarianz wird häufig der Korrelationskoeffizient

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

betrachtet. Nach der Cauchy-Schwarzschen Ungleichung ist

$$(\operatorname{cov}(X, Y))^2 \leq \operatorname{Var}(X)\operatorname{Var}(Y)$$

Daraus folgt sofort, dass $-1 \leq \rho(X, Y) \leq 1$ mit Gleichheit, falls X, Y mit Wahrscheinlichkeit 1 linear abhängig sind.

2.6.3 Satz[noch mehr Ungleichungen..]**.**

(i) Markov-Ungleichung: Sei $X \geq 0$. dann gilt

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a} \quad \forall a \in \mathbb{R}$$

(ii) Tschebyschev-Ungleichung:

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{\operatorname{Var}(X)}{a^2} \quad \forall a \in \mathbb{R}$$

Beweis:

(i)

$$\begin{aligned}\mathbb{E}(X) &= \sum_x xP(X=x) = \sum_{x\geq a} xP(X=x) + \sum_{x<a} xP(X=x)\\ &\geq \sum_{x\geq a} xP(X=x)\\ &\geq a\sum_{x\geq a} P(X=x) = aP(X\geq a)\end{aligned}$$

(ii) Wir definieren $Y = |X - \mathbb{E}(X)|^2 \geq 0$. Dann gilt nach (i)

$$P(|X-\mathbb{E}(X)|\geq a) = P(|x-\mathbb{E}(X)|^2 \geq a^2) \leq \frac{\mathbb{E}(Y)}{a^2} = \frac{\mathrm{Var}(X)}{a^2}$$

□

2.6.4 Beispiel[Anwendung der Ungleichungen]**.** Approximation von Erwartungswerten.

- Es sei $X : \Omega \to W \subset \mathbb{R}$ eine reelle Zufallsvariable mit unabhängigen Kopien $X_1, \ldots, X_n$ (z.B. n unabhängige Münzwürfe oder Versuche im Lotto zu gewinnen).
- Ferner sei $Y = X_1 + \ldots + X_n$. Dann gilt

$$\mathbb{E}(Y) = n\mathbb{E}(X),\quad \mathrm{Var}(Y) = n\mathrm{Var}(X)$$

Mit der Tschebyschev-Ungleichung folgt:

$$P(|\frac{X_1+X_2+\ldots+X_n}{n} - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2 n}\mathrm{Var}(X) \xrightarrow{n\to\infty} 0 \quad \forall\varepsilon > 0$$

D.h. die Güte der Schätzung von $\mathbb{E}(X)$ durch unabhängige Ausführungen (Samples) hängt von der Größe des Samples / Anzahl der Versuche n und der Varianz von X ab.

2.7 Zufallsvariable mit Dichtefunktion

Aigner-Skript, S. 34 ff. Wie immer $(\Omega, \mathcal{E}, P)$ sei ein Wahrscheinlichkeitsraum, wobei $\mathcal{E}$ eine geeignete σ-Algebra über Ω (hier: die σ-Algebra der Borelmengen).

2.7.1 Definition[Zufallsvariable mit Dichtefunktion]**.** Die Zufallsvariable $X : \Omega \to W = \mathbb{R}$ wird stetig verteilt mit Verteilungsfunktion $F(x) = P(X \leq x)$, wenn $F(x)$ die folgende Form hat:

$$F(x) = \int_{-\infty}^{x} f(s)ds$$

Dabei ist $f \geq 0$ die Dichte von X, d.h. es gilt

$$\int_{-\infty}^{\infty} f(s)ds = 1$$

2.7.2 Notiz[Berechnung von Wahrscheinlichkeiten]**.**

- Ereignissystem: $\mathcal{F} \subset 2^W$ sei die σ-Algebra der Borelmengen von W.
- Wahrscheinlichkeit für $X \in A$, $A \in \mathcal{F}$

$$F(X \in A) = \int\limits_{A \subset W} f(s)ds$$

- Insbesondere ist für $A = [a,b] \subset W$:

$$P(X \in A) = \int\limits_a^b f(x)ds$$

und es gilt $P(X = x) = 0 \ \forall \{x\} \in \mathcal{F}$

2.7.3 Beispiel. Zufallsvariable auf der Einheitskreisscheibe. $\Omega = \{(\omega_1, \omega_2) \mid \omega_1^2 + \omega_2^2 \leq 1\}$ und wir definieren $X = \sqrt{\omega_1^2 + \omega_2^2}$, also Abstand zum Ursprung. Frage: Welches Wahrscheinlichkeitsmaß induziert $X : \Omega \to [0,1]$?

$$P_X([c,d]) = P(\{X \in [c,d]\}) = \frac{\text{Fläche des Kreisrings} R_{cd}}{\text{Fläche von } \Omega} = \frac{\pi d^2 - \pi c^2}{\pi} = d^2 - c^2$$

Welche Dichtefunktion f gehört zu P_X?

$$f(s) = 2s \Rightarrow P_X([c,d]) = \int\limits_c^d 2sds = d^2 - c^2$$

2.8 Unabhängigkeit stetiger Zufallsvariablen

2.8.1 Definition[gemeinsame Verteilung]**.** Es seien $X, Y : \Omega \to W \subset \mathbb{R}$ zwei Zufallsvariablen mit Verteilungsfunktionen $F_X(x), F_Y(y)$ und Dichten f_X, f_Y. Als gemeinsame Verteilung von X und Y bezeichnen wir die Funktion

$$F(x,y) = P(\{X \leq x\} \cap \{Y \leq y\}) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(x,t)dsdt$$

Dabei ist $f(x,y)$ die gemeinsame Wahrscheinlichkeitsdichte von X und Y.

2.8.2 Definition[Unabhängigkeit]**.** Zwei Zufallsvariablen X, Y heißen unabhängig, wenn die Ereignisse $\{X \leq x\}$ und $\{Y \leq y\}$ unabhängig sind, d.h. wenn

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

2.8.3 Notiz[Zentrale Fragen]**.**

(i) Was ist der Zusammenhang zwischen $f(x,y)$, $f_X(x)$ und $f_Y(y)$?

(ii) was heißt Unabhängigkeit für die Dichte $f(x,y)$?

2.8.4 Lemma. Es seien $X, Y : \Omega \to W \subset \mathbb{R}$ zwei stetig verteilte Zufallsvariablen mit gegebenen Verteilungen F_x, F_y, F und Dichten f_X, f_Y. Dann gilt:

(i) $f_X(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy$, $f_Y(y) = \int\limits_{-\infty}^{\infty} f(x,y)dx$

(ii) X und Y sind genau dann ünabhängig, wenn

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

Mit

$$f \mapsto \begin{cases} f(x) & x \in W \\ 0 & \text{sonst} \end{cases}$$

folgt $W = \mathbb{R}$.

Beweis:

(i)

$$\begin{aligned} F_X(x) = P(X \leq x) &= \lim_{y \to \infty} P(\{X \leq x\} \cap \{Y \leq y\}) \\ &= \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{\infty} f(x,t)dsdt \end{aligned}$$

Nun gilt aber

$$f_X(x) = F_X'(x) = \int\limits_{-\infty}^{\infty} f(x,t)dt$$

(ii) "$\Rightarrow$": X, Y unabhängig, d.h. $F = F_X F_Y$ und es gilt

$$\frac{\delta^2}{\delta x \delta y} F(x,y) = f(x,y) \Rightarrow f(x,y) = f_X(x)f_Y(y)$$

"$\Leftarrow$": Umgekehrt folgt aus $f(x,y) = f_X(x)f_Y(y)$, dass

$$\begin{aligned} F(x,y) &= \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f_X(s)f_Y(t)dsdt = \int\limits_{-\infty}^{x} (\int\limits_{-\infty}^{y} f_Y(t)dt)f_X(s)ds \\ &= (\int\limits_{-\infty}^{x} f_X(s)ds)(\int\limits_{-\infty}^{y} f_Y(t)dt) = F_X(x)F_Y(y) \end{aligned}$$

□

2.8.5 Beispiel[Gaußsche Zufallsvariablen].

- Es seien X_1, X_2 Zufallsvariablen, deren Verteilungsfunktion durch eien Gaußdichte gegeben ist. Bekannt sei die gemeinsame Gaußverteilung von $X = (X_1, X_2)$: Erwartungswert $\mathbb{E}(X_i) = 0$ für $i = 1, 2$. Kovarianz $C_{ij} = \mathbb{E}(X_i X_j)$ für $i, j = 1, 2$.

- Kovarianzmatrix $C = (C_{ij}) \in \mathbb{R}^{2\times 2}$, $C = C^t > 0$.

- gemeinsame Gaußdichte für $x = (x_1, x_2)$:
$$f(x) = \frac{1}{\sqrt{2\pi \det C}} e^{-\frac{1}{2}x^t C^{-1} x} \qquad \text{wobei } x^t A x = \sum_{i,j} A_{ij} x_i x_j$$
- X_1, X_2 unabhängig genau dann, wenn C diagonal ist.
  - 1.Fall: $C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
  - 2.Fall: $C = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$
  - 3.Fall (abhängig): $C = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}$
  - 4.Fall: $C = \begin{pmatrix} 1 & c \\ c & 2 \end{pmatrix}$, $c >> \delta$
- Falls X_1, X_2 abhängig, so ändert sich der Erwartungswert unter der bedingten Wahrscheinlichkeit, dass entweder X_1 oder X_2 bekannt ist.

2.8.6 Definition[Marginalverteilung (Randverteilung)]**.**

$$F_X(x) = \lim_{y \to \infty} P(\{X \leq x\} \cap \{Y \leq y\}) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(s,t) dt ds$$

$\rightarrow$ marginale Dichte $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$ (analog für F_Y, f_Y).

2.9 Erwartungswert und Varianz

2.9.1 Definition. Es sei X eine stetige Zufallsvariable mit Dichtefunktion $f \geq 0$. Dann definieren wir den Erwartungswert als

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

und die Varianz als

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx$$

2.9.2 Beispiel[Gleichverteilung]**.**

- Sei $X : \Omega \rightarrow [a,b] \subset \mathbb{R}$ gleichverteilt mit Dichte
$$f(x) = (b-a)^{-1}$$
- Erwartungswert und Varianz:
$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = [\frac{x^2}{2b-2a}]_a^b = \frac{b^2 - a^2}{2b-2a} = \frac{a+b}{2}$$
$$\mathrm{Var}(X) = \int_a^b \frac{(x - \frac{a+b}{2})^2}{b-a} dx = \frac{(b-a)^2}{12}$$

- Dichte und die Verteilungsfunktion auf $\mathbb{R}$:

$$F(x) = P(X \leq x) = \int\limits_{-\infty}^{x} f(x)dx = \begin{cases} 0 & x \in (-\infty, a) \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x \in (a, \infty) \end{cases}$$

$$f(x) = \begin{cases} 0 & x \in (-\infty, a) \\ \frac{1}{b-a} & x \in [a, b] \\ 0 & x \in (b, \infty) \end{cases}$$

2.9.3 Beispiel[Erwartungswert der Gaußverteilung]**.**

- X sei eine standard-normalverteilte Zufallsvariable, d.h. mit Dichte

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

Was ist $\mathbb{E}(X)$?

$$\mathbb{E}(X) = \int\limits_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} xe^{-\frac{x^2}{2}} dx = 0 \text{ ,da } xe^{-\frac{x^2}{2}} \text{ symmetrisch zum Ursprung}$$

2.10 Eigenschaften von Erwartungswert und Varianz

2.10.1 Satz. Es sei $X : \Omega \to W \subset \mathbb{R}$ eine stetige Zufallsvariable mit Dichtefunktion $f \geq 0$ und $g : W \to W' \subset \mathbb{R}$ eine stetig differenzierbare Funktion mit $g' > 0$. Dann gilt:

$$\mathbb{E}(g(X)) = \int\limits_{W \subset \mathbb{R}} g(x)f(x)dx$$

Beweis: Wir definieren $Y = g(X)$ als neue Zufallsvariable und wir nehmen zunächst an, dass Y die Dichte $h \geq 0$ habe.

$$\Rightarrow \mathbb{E}(Y) = \int\limits_{W'} yh(y)dy$$

Nun ist $g' \neq 0 \Rightarrow g$ ist invertierbar. D.h. $X = g^{-1}(Y)$ und es folgt, dass

$$\mathbb{E}(Y) = \int\limits_{W'} yh(y)dy = \int\limits_{g^{-1}(W)} g(x)h(g(x))g'(x)dx$$

wobei wir $y = g(x)$ und $dy = g'(x)dx$ substituiert haben. Bleibt zu zeigen: $h \geq 0$ existiert (als Dichtefunktion) und $f(x) = h(g(x))g'(x)$.
Nun ist

$$\begin{aligned} F(x) &= P(X \leq x) = P(y \leq g(x)) \\ &= \int\limits_{-\infty}^{g(x)} h(y)dy = \int\limits_{-\infty}^{x} h(g(u))g'(u)du \qquad \text{Substitution: } y = g(x) \end{aligned}$$

Für die Verteilungsfunktion gilt $F'(x) = f(x)$

$$\Rightarrow F'(x) = h(g(x))g'(x) = f(x)$$

Durch Umdrehen der Substitution folgt analog:

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} \geq 0$$

□

2.10.2 Bemerkung.

(i) Ein Nebenprodukt des Satzes war die folgende Transformationsformel für Zufallsvariablen:
Hat X die Dichte $f(x)$, dann hat $Y = g(X)$, $g'(x) \neq 0$ die Dichte

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

(ii) Der Satz gilt gleichermaßen für mehrere Zufallsvariablen $X_1, \ldots, X_n$ mit gemeinesamer Dichte $f(x_1, \ldots, x_n)$:

$$\mathbb{E}(g(X_1, \ldots, X_n)) = \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

2.10.3 Folgerung[Unmittelbare Folgerungen aus dem Satz]**.**

(i) $\mathrm{Var}(X) = \int_W (X - \mathbb{E}(X))^2 f(x) dx = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

(ii) $\mathbb{E}(\alpha X + \beta) = \alpha \mathbb{E}(X) + \beta \quad \forall \alpha, \beta \in \mathbb{R}$

(ii)' $\mathbb{E}(\alpha_1 X_1 + \ldots + \alpha_n X_n) = \alpha_1 \mathbb{E}(X_1) + \ldots + \alpha_n \mathbb{E}(X_n)$, $\alpha_i \in \mathbb{R}$

Beweis: $X_1, \ldots, X_n$ mit gemeinesamer Dichtefunktion $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n) = \sum_{i=1}^n \alpha_i x_i$

$$\begin{aligned}\mathbb{E}(\sum_{i=1}^n \alpha_i x_i) &= \int (\sum_{i=1}^n \alpha_i x_i) f(x_1, \ldots, x_n) dx = \sum_{i=1}^n \alpha_i \int x_i f(x_1, \ldots, x_n) dx_1 \ldots dx_n \\ &= \sum_{i=1}^n \alpha_i \int_W x_i f_{X_i}(x_i) dx_i = \sum_{i=1}^n \alpha_i \mathbb{E}(X_i)\end{aligned}$$

Wobei wir ausgenutzt haben, dass

$$f_{X_i}(x_i) = \int f(x_1, \ldots, x_i, x_{i+1}, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$$

□

(iii) $\mathrm{Var}(\alpha X + \beta) = \alpha^2 \mathrm{Var}(X) \quad \forall \alpha \beta \in \mathbb{R}$

(iv) $P(X \in A) = \mathbb{E}(\chi_A(X)) \quad \forall A \in \mathcal{E}$

2.10.4 Beispiel[Transformation Gaußscher Zufallsvariablen]**.**

- Sei X standard-normalverteilt mit Dichte

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

- Betrachte eine neue Zufallsvariable $Y = \sigma X + \mu$, $\sigma > 0$. Was ist die Dichte von Y?

- Möglichkeit 1: Berechnung von $\mathbb{Y}$, $\mathrm{Var}(Y)$:
  - $\mathbb{E}(Y) = \mathbb{E}(\sigma X + \mu) = \sigma \mathbb{E}(X) + \mu$
  - $\mathrm{Var}(Y) = \sigma^2 \mathrm{Var}(X)$
  - $\Rightarrow h(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$

- Möglichkeit 2: Transformationsformel

$$g(x) = \sigma x + \mu \Rightarrow g'(x) = \sigma > 0, g^{-1}(y) = \frac{y - \mu}{\sigma}$$

Einsetzen:

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

2.11 Summe von Zufallsvariablen (Faltung)

Im Folgenden seien $X, Y : \Omega \to \mathbb{R}$ zwei Zufallsvariablen mit gemeinsamer Dichte $f(x, y)$.
Gesucht: Dichte und die Verteilungsfunktion der Summe $Z = X + Y$.

2.11.1 Beispiel.

(i) Verteilung des arithmetischen Mittels $\frac{X+Y}{2}$

(ii) Toleranzen bei zusammengesetzten Bauteilen

(iii) Summe der Augenzahlen beim Würfeln (diskret)

2.11.2 Notiz[Zentrale Fragen]**.**

(i) In welchen Fällen lässt sich die Verteilung F_{X+Y} bzw. die Dichte f_{X+Y} (elementar) aus den f_X, f_Y bzw. F_X, F_Y berechnen?

(ii) Ist die Verteilung der Summe vom gleichen Typ wie die Verteilung der Summanden?

2.11.3 Satz. Es seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen mit Dichte $f(x, y)$. Dann hat $Z = X + Y$ die Dichtefunktion

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x) dx = \int_{-\infty}^{\infty} f(z - x, x) dx$$

Beweis: Wir berechnen die Verteilungsfunktion F_Z für das Ereignis

$$\{Z \leq z\} = \{(x,y) \mid x+y \leq z\} \subset \mathbb{R} \times \mathbb{R}$$

Ist $F_Z(z) = P(Z \leq z)$ bekannt, dann bekommen wir die Dichte aus $f_Z(z) = F_Z'(z)$.

$$F_Z(z) = \iint\limits_{\{Z \leq z\}} f(x,t) dsdt = \int\limits_{s=-\infty}^{\infty} \int\limits_{t=-\infty}^{z-s} f(s,t) dsdt$$

Neue Variable: $x = s, y = s+t \Leftrightarrow s = x, t = y - x \Rightarrow dsdt = dxdy$ (Determinante $= 1$).

$$\Rightarrow F_Z(z) = \int\limits_{x=-\infty}^{\infty} \int\limits_{y=-\infty}^{z} f(x, y-x) dxdy = \int\limits_{-\infty}^{z} (\int\limits_{-\infty}^{\infty} f(x, y-x) dx) dy$$

$$\Rightarrow f_Z(z) = F_Z'(z) = \int\limits_{-\infty}^{\infty} f(x, z-x) dx$$

□

2.11.4 Korollar. Sind X, Y unabhängig ($\Leftrightarrow f(x,y) = f_X(x) f_Y(y)$), dann ist

$$f_{X+Y}(z) = \int\limits_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int\limits_{-\infty}^{\infty} f_X(z-x) f_Y(x) dx$$

Der Integralausdruck $\int_{-\infty}^{\infty} f(x) g(z-s) ds = (f * g)(z)$ wird "Faltung" genannt.

2.11.5 Bemerkung.

(i) Die Faltung ist assoziativ, d.h.

$$f * g * h = (f * g) * h = f * (g * h)$$

Daher lässt sich der Satz (bzw. seine Folgerung) leicht durch Induktion auf n unabhängige Zufallsvariablen $X_1, \ldots, X_n$ verallgemeinern.

(ii) Im diskreten Fall ist die Verteilung der Summe zweier Zufallsvariablen $X, Y : \Omega \to W \subset \mathbb{N}$ durch die diskrete Faltung

$$P_{X+Y}(k) = \sum_{-\infty}^{\infty} P_X(i) P_Y(k-i), \quad k \in W \subset \mathbb{N}$$

2.11.6 Beispiel[Summe gleichverteilter Zufallsvariablen]**.**

- Seien $X, Y : \Omega \to [0,1]$ unabhängig udn gleichverteilt mit $f_X = f_Y = 1$ auf $[0,1]$ und $f_X = f_Y = 0$ sonst.
- Dichte von $X + Y$?

$$f_{X+Y}(z) = \int\limits_{-\infty}^{\infty} f_X(s) f_Y(z-s) ds$$

- Wir unterscheiden zwei Fälle:
  - $z \in [0, 1]$:
  $$f_{X+Y}(z) = \int_0^z 1 ds = z$$
  - $z \in [1, 2]$:
  $$f_{X+Y}(z) = \int_{z-1}^1 1 ds = 2 - z$$
- D.h. $X + Y$ ist nicht gleichverteilt!

2.11.7 Beispiel[weitere Beispiele..]**.**

- Gaußverteilung: Sind X_1, X_2 unabhängig gaußverteilt mit Erwartungswert μ_1, μ_2 und Varianz σ_1, σ_2 (Notation: $X_i \sim N(\mu_i, \sigma_i^2)$), dann gilt
$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$
- Exponentialverteilung: Sind $X_1, X_2 : \Omega \to [0, \infty)$ unabhängig exponentialverteilt mit Dichte $f_i(x) = \lambda e^{-\lambda x}, \lambda > 0, i \in \{1, 2\}$, dann hat die Summe die Dichte
$$f_{X_1+X_2}(z) = \lambda^2 e^{-\lambda z} z$$
Falls Dichte unterschiedlich mit $\lambda_1 \neq \lambda 2$, dann folgt
$$f_{X_1+X_2}(z) = \lambda_1 \lambda_2 \int_0^z e^{-\lambda_1 s} e^{-\lambda_2 (z-s)} ds = -\frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_1 z} - e^{-\lambda_2 z})$$

2.12 Exponentialverteilung (Wartezeitmodelle)

Modellierung von zufälligen Zeitintervallen (Wartezeiten).

2.12.1 Beispiel.

- Zeit zwischen zwei Anrufen
- Zeit, bis ein Fisch anbeißt
- Zahl der Würfe bis zur nächsten 6

2.12.2 Definition[gedächtnislose Wartezeit]**.** Es sei $X : \Omega \to [0, \infty)$ eine Zufallsvariable auf Ω. X heißt "gedächtnislose Wartezeit", wenn
$$P(X \geq x + t \mid X \geq t) = P(X \geq x)$$

2.12.3 Satz. Folgende Aussagen sind äquivalent:

(i) X ist eine gedächtnislose Wartezeit

(ii) X ist exponentialverteilt mit Dichte $f(x) = \lambda e^{-\lambda x}$ für ein $\lambda > 0$

Beweis: Wir zeigen zunächst, dass die Definition gleichbedeutend ist mit der Aussage $P(X \geq s+t) = P(X \geq s)P(X \geq t)\ \forall x, t \in [0, \infty)$. Nach Definition bedingter Wahrscheinlichkeiten ist

$$P(X \geq s+t \mid X \geq t) = \frac{P(\{X \geq s+t\} \cap \{X \geq t\})}{P(X \geq t)} = \frac{P(X \geq s+t)}{P(X \geq t)}$$

Also "(ii) $\Rightarrow$ (i)":
Verteilungsfunktion $F(x) = P(X \leq x) = \int_0^x f(s)ds = \lambda \int_0^x e^{-\lambda s}ds = 1 - e^{-\lambda x}$

$$\begin{aligned}&\Rightarrow P(X \geq x) = 1 - P(X \leq x) = e^{-\lambda x}\\ &\Rightarrow P(X \geq s+t) = e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t} = P(X \geq t)P(X \geq s)\end{aligned}$$

Nun "(i) $\Rightarrow$ (ii)":
Setze $G(x) = 1 - F(x)$, wobei $F(x) = P(X \leq x)$ ist und sich $G(x) = \int_x^\infty g(s)ds$ schreiben lässt.
Eigenschaften von G:

(a) $G(0) = 1$ (gleichbedeutend mit $P(\Omega) = 1$)

(b) $G(s+t) = G(s)G(t)$ (Definition für die gedächtnislose Wartezeit)

(c) $\lim\limits_{x\to\infty} G(x) = 0$ (da $G = 1 - F$)

Zu zeigen: (a) - (c) bestimmen eindeutig $G(x) = e^{\alpha x}$, $\alpha < 0$ (d.h. $\alpha = -\lambda$). Denn dann folgt

$$f(x) = F'(x) = -G'(x) = -\frac{\delta}{\delta x}(e^{-\lambda x}) = \lambda e^{-\lambda x}$$

<u>Beweis in zwei Schritten:</u>
Wir zeigen zunächst, dass $G(s) > 0 \quad \forall s \geq 0$. Es ist $G(0) > 0$ und wegen der Stetigkeit existiert ein $\varepsilon > 0$, so dass $G(s) > 0 \quad \forall s \in [0, \varepsilon]$.
$\Rightarrow G(s) > 0 \quad \forall s \geq 0$, denn

$$G(s) = \underbrace{G(\frac{s}{n}) \cdot \ldots \cdot G(\frac{s}{n})}_{n\text{-mal}} = G(\frac{s}{n})^n$$

und es gibt ein $n_0 \in \mathbb{N}$, so dass $\frac{s}{n} \in [0, \varepsilon] \quad \forall n > n_0$. Nun gilt, dass $G(1) = e^\alpha > 0$, $\alpha = \log G(1) \in \mathbb{R}$. Definiere nun eine Funktion $\tilde{G}(s) = e^{\alpha s}$. Wir weren zeigen, dass $\tilde{G}(s) = G(s)$ für alle $s = \frac{n}{m} \in \mathbb{Q}^+$.
Es gilt $G(1) = G(\frac{m}{m}) \overset{(b)}{=} G(\frac{1}{m})^m = \tilde{G}(1) = e^\alpha \Leftrightarrow G(\frac{1}{m}) = \tilde{G}(\frac{1}{m})$. Also gilt auch $G(\frac{1}{m})^n = \tilde{G}(\frac{1}{m})^n$ bzw. (wegen (b))

$$G(\frac{n}{m}) = \tilde{G}(\frac{n}{m}) \quad \forall n \in \mathbb{N}_0, m \in \mathbb{N}.$$

Wegen der Stetigkeit gilt $G(s) = \tilde{G}(s) \quad \forall s \geq 0$.
Wegen (c) muss $\alpha < 0$ gelten. Wir setzen $\alpha = -\lambda$, $\lambda > 0$ und sind fertig. $\square$

2.12.4 Bemerkung.

(i) Oft ist die Annahme, dass eine Wartezeit gedächtnislos ist, Unfug. Dennoch ist die Exponentialverteilung oft (meist) eine gute (erste) Näherung.

(ii) "Warten" muss nicht notwendigerweise in Zeiteinheiten gemessen werden.
Beispiel: Strecke bis zum nächsten Unfall auf der Autobahn oder Seiten bis zum nächsten Tippfehler in einem Text.

(iii) Die diskrete Variante der gedächtnislosen Wartezeit wird durch die geometrische Verteilung

$$P(n) = q^{n-1}(1-q) \quad \text{mit } q = 1-p$$

beschrieben.
Beispiel: Wann kommt die nächste 6 beim Würfeln?

2.12.5 Beispiel[Hotline]**.**

- Hotline angerufen - es ist besetzt. Sie wissen aus Erfahrung, dass die durchschnittliche Wartezeit 10 Minuten beträgt.

- Wie groß ist die Wahrscheinlichkeit, nach 5 Mintuen dranzukommen?
  - Annahme: Wartezeit ist exponentialverteilt.

$$f(x) = \lambda e^{-\lambda x} \text{ mit } \lambda = \frac{1}{10 \text{ Minuten}}$$

  - Begründung:

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^\infty x f(x) dx && \text{(durchschnittliche Wartezeit)} \\
&= \lambda \int_0^\infty x e^{-\lambda x} dx = \underbrace{[-xe^{-\lambda x}]_0^\infty}_{=0} + \int_0^\infty e^{-\lambda x} dx \\
&= -\frac{1}{\lambda}[e^{-\lambda x}]_0^\infty = \frac{1}{\lambda}
\end{aligned}$$

  - Was uns interessiert ist:

$$P(X \leq 5) = \frac{1}{10} \int_0^5 e^{-\frac{x}{10}} = [-e^{-\frac{x}{10}}]_0^5 = 1 - e^{-\frac{1}{2}} \approx 0,39$$

- Wie lange muss man warten, um mit 90%-iger Sicherheit an die Reihe zu kommen?

$$P(X \leq T) = F(T) = \frac{1}{10} \int_0^T e^{-\frac{x}{10}} \geq 0,9$$

Also $1 - e^{-\frac{T}{10}} \geq 0,9$, d.h. $T \geq 23$ Minuten.

- Hotline 2-mal angerufen, beide Male besetzt bzw. 1-mal angerufen und sie wissen, dass eine Person vor ihnen ist. Was ist die Verteilung der summierten Wartezeiten? (Satz: $f_{X+Y}(x) = (f_X * f_Y)(x)$).

– Annahme $f_X = f_Y$ mit Parameter $\lambda > 0$:

$$f_{X+Y}(x) = \lambda^2 \int_0^x e^{-\lambda s} e^{-\lambda(x-s)} ds = \lambda^2 e^{-\lambda x} x$$

– f_X, f_Y unterschiedlich mit Parametern $\lambda, \mu > 0$:

$$\tilde{f}_{X+Y}(x) = \frac{\lambda\mu}{\mu - \lambda}(e^{-\lambda x} - e^{-\mu x})$$

– Es lässt sich zeigen, dass

$$\lim_{\mu \to \lambda} \tilde{f}_{X+Y} = f_{X+Y}(x)$$

Beweis mit l'Hospital.

2.13 Mehr zu Unabhängigkeit

Wir betrachten nun den Fall von n unabhängigen Zufallsvariablen $X_1, \ldots, X_n$.

2.13.1 Beispiel. Hotline mit n Leitungen.

2.13.2 Satz. Es seien $X_1, \ldots, X_n$ unabhängig exponentialverteilt mit Parameter $\lambda > 0$. Dann gilt

(i) $X_{\min} = \min\{X_1, \ldots, X_n\}$ ist exponentialverteilt mit Parameter $n \cdot \lambda$.

(ii) $X_{\max} = \max\{X_1, \ldots, X_n\}$ hat die Dichte

$$f_{\max} = n\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{n-1}$$

Beweis: $F_{\min,\max} = P(X_{\min,\max} \leq x)$ und $G_{\min,\max}(x) = 1 - F_{\min,\max}(x)$.

(i) Es gilt

$$\begin{aligned} G_{\min}(x) &= P(X_{\min} \geq x) = P(\{X_1 \geq x\} \cap \ldots \cap \{X_n \geq x\}) = \prod_{i=1}^n P(X_i \geq x) \\ &= \prod_{i=1}^n (1 - F(x)) = e^{-\lambda n x} \qquad \text{siehe letzte VL} \end{aligned}$$

Dichte: $f_{\min}(x) = F'_{\min}(x) = (1 - G_{\min}(x))' = n\lambda e^{-\lambda n x}$

(ii) $F_{\max}(x) = \prod_{i=0}^n P(X_i \leq x) = (F(x))^n = (1 - e^{-\lambda x})^n$ (siehe letzte VL).

$$\Rightarrow f_{\max}(x) = F'_{\max}(x) = n(1 - e^{-\lambda x})^{n-1} \lambda e^{-\lambda x}$$

□

2.13.3 Korollar. Es gilt

(i) $\mathbb{E}(X_{\min}) = \frac{1}{n\lambda} \to 0$ für $n \to \infty$

(ii) $\mathbb{E}(X_{\max}) = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i} \to \infty$ für $n \to \infty$

Beweis: Erwartungswerte ausrechnen!

2.13.4 Beispiel.

(i) 3 unabhängige Kühlkreisläufe in einem AKW:

$$\mathbb{E}(X_{\max}) = \frac{1}{\lambda}(1 + \frac{1}{2} + \frac{1}{3}) = \frac{11}{6\lambda} \approx \frac{2}{\lambda}$$

(ii) Sportlehrer wartet auf 30 Schüler, die sich umziehen:

$$\mathbb{E}(X_{\max}) = 3\text{min}(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{30}) >> 3\text{min}$$

3 Grenzwertsätze

3.1 Konvergenz von Zufallsvariablen

3.1.1 Beispiel. Fairer Würfel, $X = \begin{cases} 1 & \text{falls 6 gewürfelt} \\ 0 & \text{sonst} \end{cases}$, $X_1, X_2, \ldots$ seien unabhängige Kopien von X.

$$\begin{aligned} Y_n := \frac{X_1 + \ldots + X_n}{n} &= \text{relativer Anteil von } n \text{ Würfen, die 6 ergeben} \\ &\approx \frac{1}{6} \text{ für große } n \end{aligned}$$

bzw. $Y_n \overset{n\to\infty}{\to} \frac{1}{6}$, aber was heißt das genau?

3.1.2 Beispiel. X gleichverteilt auf $[0, 1]$, $X_1, X_2, \ldots$ unabhängige Kopien von X. Wie sieht die Dichte von $X_1 + \ldots + X_n$ für verschiedene n aus? Für große n ähnelt sich die Faltung immer mehr der Normalverteilung.
Bei der Dichte von $\frac{X_1+\ldots+X_n}{n}$ für verschiedene n? Es konvergiert gegen $x = 0,5$ (vertikaler Strich).

3.1.3 Erinnerung. $(a_n)_{n\in\mathbb{N}}$ Folge in $\mathbb{R}$.

$$a_n \overset{n\to\infty}{\to} a \quad :\Leftrightarrow \quad \forall\varepsilon > 0\ \exists n_0 \in \mathbb{N} : \forall n \geq n_0 : |a_n - a| < \varepsilon$$

3.1.4 Definition. Seien $X_1, X_2, \ldots$ Zufallsvariablen in $\mathbb{R}$ auf $(\Omega, \mathcal{E}, P)$.

a) $X_n \to X$ fast sicher $:\Leftrightarrow P(\{\omega \in \Omega \mid X_n(\omega) \to X(\omega)\}) = P(X_n \to X) = 1$
Schreibweise: $X_n \overset{\text{f.s.}}{\to} X$

b) $X_n \to X$ in Wahrscheinlichkeit $:\Leftrightarrow P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}) \overset{n\to\infty}{\to} 0 \quad \forall\varepsilon > 0$
Schreibweise (kurz): $X_n \overset{\text{i.W.}}{\to} X \Leftrightarrow P(|X_n - X| > \varepsilon) \overset{n\to\infty}{\to} 0 \quad \forall\varepsilon > 0$

c) $X_n \to X$ in Verteilung $:\Leftrightarrow P(X_n \leq x) \overset{n\to\infty}{\to} P(X \leq x)$ (d.h. $X_{X_n}(x) \to F_X(x)$) für alle $x \in \mathbb{R}$, in denen F_X stetig ist.
Schreibweise: $X_n \overset{\text{i.V.}}{\to} X$

3.1.5 Bemerkung. Ein Ereignis ist "fast sicher", falls $P(\text{Ereignis}) = 1$

3.1.6 Satz.

(i) $X_n \overset{\text{f.s.}}{\to} X \Rightarrow X_n \overset{\text{i.W.}}{\to} X$

(ii) $X_n \overset{\text{i.W.}}{\to} X \Rightarrow X_n \overset{\text{i.V.}}{\to} X$

Beweis:

(i) O.B.d.A. $X = 0$ (sonst nehme $X_n - X$). Wir wissen $P(X_n \to 0) = 1$. Zu zeigen: $\forall \varepsilon\ P(|X_n| > \varepsilon) \overset{n\to\infty}{\to} 0$.
Sei ε fest. Definiere $A_n := \{\omega \in \Omega \mid |X_k(\omega)| \leq \varepsilon \quad \forall k \geq n\}$. Es gilt also

$$A_1 \subset A_2 \subset \ldots$$

Sei $\omega \in \Omega$ mit $X_n(\omega) \to 0$. Dann existiert $n_0 \in \mathbb{N}$ mit $|X_n(\omega)| \leq \varepsilon \quad \forall n \geq n_0$. D.h. $\omega \in A_{n_0}$. Es gilt

$$\underbrace{\{\omega \in \Omega \mid X_n(\omega) \to 0\}}_{\text{hat Wahrscheinlichkeit 1}} \subset \bigcup_{n\in\mathbb{N}} A_n$$

Es gilt außerdem, dass $P(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} P(A_n)$. Zusammen haben wir:

$$1 = P(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} P(A_n)$$

Nun $\{\omega \in \Omega \mid |X_n(\omega)| > \varepsilon\} \subset \Omega \setminus A_n$ und damit $P(|X_n| > \varepsilon) \leq 1 - P(A_n)$.

$$\Rightarrow \lim_{n\to\infty} P(|X_n| > \varepsilon) = \lim_{n\to\infty} 1 - P(A_n) = 0$$

Und damit ist die Implikation gezeigt. Gegenbeispiel für die Rückrichtung:
$X_n = \begin{cases} 1 \\ 0 \end{cases}$ mit $P(X_n = 1) = \frac{1}{n}$ und $P(X_n = 0) = 1 - \frac{1}{n}$.

$$P(X_n > \varepsilon) = \begin{cases} \frac{1}{n} & 0 < \varepsilon < 1 \ \to 0 \\ 0 & \varepsilon \geq 1 \ \to 0 \end{cases} \Rightarrow X_n \overset{\text{i.W.}}{\to} 0$$

Aber $X_n \overset{\text{f.s.}}{\not\to} 0$, denn: Definiere $A_n := \{\omega \in \Omega \mid X_k(\omega) = 0 \text{ für } k \geq n\}$, Eigenschaften von A_n also wie im Beweis. Es lässt sich zeigen:

$$\bigcup_{n\geq 1} A_n = \{\omega \in \Omega \mid X_n(\omega) \to 0\} \text{ und } P(A_n) = 0$$

$\Rightarrow P(\bigcup_{n\in\mathbb{N}} A_n) = 0 \Rightarrow P(X_n \to 0) = 0$

(ii) Sei $X_n \overset{\text{i.W.}}{\to} X$, d.h. $P(|X_n - X| > \varepsilon) \overset{n\to\infty}{\to} 0\ \forall \varepsilon > 0$. Sei $F_n(x) = P(X_n \leq x)$, $F(x) = P(X \leq x)$. Sei $\varepsilon > 0$ fest. Es gilt:

$$F_n(x) = P(X_n \leq x) = \underbrace{P(X_n \leq x \wedge X \leq x + \varepsilon)}_{\leq P(X \leq x+\varepsilon) = F(x+\varepsilon)} + \underbrace{P(X_n \leq x \wedge X > x + \varepsilon)}_{\leq P(|X_n - X| > \varepsilon)}$$

Außerdem:

$$F(x-\varepsilon) = P(X \leq x-\varepsilon) = \underbrace{P(X \leq x-\varepsilon \wedge X_n \leq x)}_{\leq F_n(x)} + \underbrace{P(X \leq x-\varepsilon \wedge X_n > x)}_{\leq P(|X_n - X| > \varepsilon)}$$

Zusammen:

$$F(x-\varepsilon) - \underbrace{P(|X_n - X| > \varepsilon)}_{\overset{n\to\infty}{\to} 0} \leq F_n(x) \leq F(x+\varepsilon) + \underbrace{P(|X_n - X| > \varepsilon)}_{\overset{n\to\infty}{\to} 0}$$

Nach Quetschlemma: $F(x) = \lim\limits_{n\to\infty} F_n(x)$, also $X_n \overset{\text{i.V.}}{\to} X$.

Gegenbeispiel für die Rückrichtung: $X_n = \begin{cases} 0 \\ 1 \end{cases}$ mit $P(X_n = 0) = \frac{1}{2} = P(X_n = 1)$.
$X_1 = \ldots = X_n = X$. Setze $Y = 1 - X$.

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

und $F_Y(x) = P(Y \leq x) = P(X \leq x) = F(x) = F_n(x)$. Aber $|X_n - Y| = 1$, d.h. $P(|X_n - Y| = 1) = 1\ \forall n \Rightarrow P(|X_n - Y| > \varepsilon) = 0\ \forall \varepsilon < 1$ und somit ist keine Konvergenz gegen 0 möglich.

□

3.1.7 Satz[Das schwache Gesetz der großen Zahl]**.** Seien $X_1, X_2, \ldots$ unabhängige Kopien von einer reellen Zufallsvariablen X. Dann gilt:

$$P(|\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}(X)| \geq \varepsilon) \overset{n\to\infty}{\to} 0 \quad \forall \varepsilon > 0$$

d.h. $\frac{X_1+\ldots+X_n}{n} \overset{\text{i.W.}}{\to} \mathbb{E}(X)$. Genauer:

$$P(|\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}(X)| \geq \varepsilon) \leq \underbrace{\frac{\operatorname{Var}(X)}{n\varepsilon^2}}_{\overset{n\to\infty}{\to} 0} \quad \forall \varepsilon > 0$$

Im Beispiel: Würfel $X = \begin{cases} 0 & \text{falls } 6 \\ 1 & \text{sonst} \end{cases} \Rightarrow \mathbb{E}(X) = \frac{1}{6}$.

Beweis: Einfach die Tschebyschev-Ungleichung anwenden. $\bar{X}_n := \frac{X_1+\ldots+X_n}{n}$. $\mathbb{E}(\bar{X}_n) = \mathbb{E}(X)$, $\operatorname{Var}(\bar{X}_n) = \frac{1}{n^2}\sum \operatorname{Var}(X) = \frac{\operatorname{Var}X}{n}$
Tschebyscheff: $P(|\bar{X}_n - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\operatorname{Var}\bar{X}_n}{\varepsilon^2} = \frac{\operatorname{Var}X}{n\varepsilon^2}$ □

3.1.8 Bemerkung[Konvergenzbegriffe, Zusammenstellung]**.**

- Fast sichere Konvergenz:

$$X_n \xrightarrow{\text{f.s.}} X \Leftrightarrow P(\{\omega \in \Omega \mid X_n(\omega) \to X(\omega)\}) = 1$$

- Konvergenz in Wahrscheinlichkeit

$$X_n \xrightarrow{\text{i.W.}} X \Leftrightarrow P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}) \xrightarrow{n\to\infty} 0 \quad \forall \varepsilon > 0$$

- Konvergenz in Verteilung

$$X_n \xrightarrow{\text{i.V.}} X \Leftrightarrow P(X_n \leq x) \xrightarrow{n\to\infty} P(X \leq x)$$

Interpretation: Fast sichere Konvergenz und Konvergenz in Wahrscheinlichkeit sind Aussagen über einzelne Ausführungen eines Zufallsexperiments. Konvergenz im Verteilungssinne betrifft das Verhalten im Mittel.

3.1.9 Beispiel[Schwaches Gesetz der großen Zahl]**.**

- Relative Häufigkeit der "6" beim Würfeln, $X_i = \{\text{"6" beim i-ten Wurf}\}$
- Fairer Würfel: $\mathbb{E}(X_i) = p = \frac{1}{6}$, $\mathrm{Var}(X_i) = p(1-p) = \frac{1}{36}$ (X_i bernoulliverteilt mit Wahrscheinlichkeit $p = \frac{1}{6}$).
- Schwaches Gesetz der großen Zahlen:

$$P(|\underbrace{\frac{X_1 + \ldots + X_n}{n}}_{\text{empirisches Mittel}} - \underbrace{\frac{1}{6}}_{\substack{\text{Grenzwert}\\ p=\mathbb{E}(X)}} | \geq \varepsilon) \leq \frac{\mathrm{Var}(\frac{X_1+\ldots+X_n}{n}))}{n\varepsilon^2} = \underbrace{\frac{1}{n\varepsilon^2}}_{\text{Konvergenzrate}} \cdot \frac{5}{36}$$

3.2 Borel-Cantelli-Lemma(0-1-Gesetz)

Hilfssatz zur Untersuchung der Konvergenz von Zufallsvariablen ($\to$ starkes Gesetz der großen Zahlen).

3.2.1 Notiz[Voraussetzungen]**.**

(i) Es sei $(\Omega, \mathcal{E}, P)$ ein Wahrscheinlichkeitsraum, $A_1, A_2, \ldots$ eine unendliche Folge von Ereignissen mit $A_i \in \mathcal{E}$.
Frage: Welche Ereignisse kommen unendlich oft vor? Etwas präziser: Wir fragen nach $\omega \in \Omega$, die in unendlich vielen A_i liegen, in Formeln:

$$A_\infty = \{\omega \in \Omega \mid \omega \in A_k \text{ für unendlich viele } k\}$$

Beispiel: Auftreten der "6" beim Würfeln. Stichprobenraum $\Omega = \{(\omega_1, \omega_2, \ldots)\}$, $\omega_i \in \{1,2,3,4,5,6\}$, also eine unendliche Folge von Würfen. Wir betrachten die Ereignisfolge $A_1, A_2, \ldots$ mit $A_i = \{\text{"6" beim i-ten Wurf}\}$. D.h.

$$A_\infty = \{\omega \in \Omega \mid \omega_k = 6 \text{ unendlich oft}\}$$

(ii) A_∞ ist ein Ereignis, denn A_∞ lässt sich als

$$A_\infty = \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} B_n \text{ mit } B_n = \bigcup_{k\geq n} A_k$$

schreiben, d.h. $A_\infty \in \mathcal{E}$ und es gilt, dass $B_1 \supset B_2 \supset \ldots$ eine aufsteigende Kette bilden, weshalb

$$P(A_\infty) = P(\bigcap_{n=1}^{\infty}) = \lim_{n\to\infty} P(B_n) \text{ (Stetigkeit von Oben)}$$

3.2.2 Satz[Borel-Cantelli-Lemma]**.**

- Ist $\sum\limits_{k\geq 1} P(A_k) < \infty$, so gilt $P(A_\infty) = 0$
- Sind $A_1, A_2, \dots$ unabhängig und $\sum\limits_{k\geq 1} A_k = \infty$, dann ist $P(A_\infty) = 1$

Beweis:

(i) Wegen der Konvergenz der Reihe $\sum\limits_{k\geq 1} A_k$ gilt, dass $\lim\limits_{n\to\infty} \sum\limits_{k\geq n} P(A_k) = 0$. Wegen $P(B_1) \leq \sum\limits_{k\geq n} P(A_k)$ folgt, dass

$$P(A_\infty) = \lim_{n\to\infty} P(B_n) = 0$$

(ii) Wir werden zeigen, dass $P(A_\infty^c) = 0$. Wegen der Unabhängigkeit von $A_1, A_2, \dots$ sind auch $A_1^c, A_2^c, \dots$ unabhängig und somit ist

$$P(\bigcap_{i=1}^n A_i^c) = \prod_{i=1}^n \underbrace{(1 - P(A_i))}_{=P(A_i^c)} \leq \prod_{i=1}^n e^{-P(A_i)} = e^{-\sum_{i=1}^n P(A_i)}$$

Daher ist

$$\lim_{n\to\infty} P(\bigcap_{i=1}^n A_i^c) = 0$$

Es folgt:

$$P(A_\infty^c) = P(\underbrace{(\bigcap_{n=1}^\infty \bigcup_{k\geq n} A_k)^c}_{=\liminf_{n\to\infty} A_n^c}) = \lim_{n\to\infty} P(\bigcap_{k\geq n} A_k^c)$$

Also gilt, dass $P(A_\infty^c) = 0$ und somit $P(A_\infty) = 1$.

□

3.2.3 Bemerkung. Ohne die Unabhängigkeit der $A_1, A_2, \dots$ ist (ii) falsch: Mit z.B. A_1, so dass $0 < P(A_1) < 1$ und $A_k = A_1$ $\forall k$, folgt, dass $P(A_\infty) = P(\omega \in A_1) \in (0, 1)$.

3.2.4 Beispiel.

(i) Würfeln: $A_i = \{$"6" im i-ten Wurf$\}$ hat die Wahrscheinlichkeit, eine 6 zu würfeln, d.h. $P(A_i) = \frac{1}{6}$. Also ist $\underset{k\geq 1}{P}(A_k) = \infty$ und die "6" tritt fast sicher unendlich oft auf.

(ii) Infinite Monkey Theorem: Affe drückt wohllos (d.h. gleichverteilt) Tasten einer Schreibmaschine und A_i sei das Ergebnis, alle Bände von Shakespeare. Auch hier ist $\underset{k\geq 1}{P}(A_k) = \infty$, d.h. Shakespeare wird ∞ oft geschrieben.

(iii) Urnenexperiment: Seien $U_1, U_2, \dots$ Urnen mit roten und weißen Kugeln, wobei U_n gerade 1 weiße und $n^\alpha - 1$ rote Kugeln hat.

$$P(\underbrace{\{\text{weiße Kugel im i-ten Zug}\}}_{A_i}) = \frac{1}{n^\alpha}$$

a) $\alpha = 1$: $\sum_{k\geq 1} P(A_i) = \sum_{n\geq 1} \frac{1}{n} = \infty$

b) $\alpha > 1$: $\sum_{k\geq 1} P(A_i) < \infty$

d.h. es werden fast sicher nur endlich viele weiße Kugeln gezogen (im Fall $\alpha > 1$).

3.3 Starkes Gesetz der großen Zahl(en)

3.3.1 Beispiel[Urnenexperiment]**.** Seien $U_1, U_2, U_3, \ldots$ Urnen, wobei U_n genau 1 weiße und $n^\alpha - 1$ rote Kugeln ($\alpha \in \{1, 2\}$) hat.

$$\Rightarrow P(\{\text{weiße Kugel aus } U_n \text{ gezogen}\}) = \frac{1}{n^\alpha}$$

Sei $A_k = \{\text{weiße Kugel aus } U_k \text{ gezogen}\} \Rightarrow P(A_k) = \frac{1}{k^\alpha}$. 2 Fälle:

- $\alpha = 1$:
$$\sum_{k\geq 1} P(A_k) = \sum_{k=1}^{\infty} \frac{1}{k} = \infty$$
d.h. es werden fast sicher ∞ viele weiße Kugeln gezogen

- $\alpha = 2$:
$$\sum_{k\geq 1} P(A_k) = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$
d.h. es werden fast sicher nur endlich viele weiße Kugeln gezogen.

Bisherige Situation:
Seien $X_1, X_2, \ldots$ unabhängige Kopien von $X : \Omega \to \mathbb{R}$, deren Erwartungswert $\mathbb{E}(X)$ und die Varianz $\mathrm{Var}(X)$ existieren. Sei ferner $S_n = X_1 + \ldots + X_n$. Dann folgt aus der Tschebyscheff-Ungleichung das schwache Gesetz der großen Zahlen:

$$P(|\frac{S_n}{n} - \mathbb{E}(X)| < \varepsilon) \to 1 \quad \text{für } n \to \infty$$

Mit anderen Worten: Für $n \to \infty$ gilt: $\frac{S_n}{n} \approx \mathbb{E}(X)$
Können wir die stärkere Aussage ($\frac{S_n}{n} \to \mathbb{E}(X)$ fast sicher) auch zeigen?

3.3.2 Satz[starkes Gesetz der großen Zahl(en)]**.** Es gilt $\frac{S_n}{n} \overset{\text{f.s.}}{\to} \mathbb{E}(X)$, d.h.

$$P(\{\omega \in \Omega \mid \frac{S_n(\omega)}{n} \to \mathbb{E}(X)\}) = 1 \text{ für } n \to \infty$$

Beweisidee (Borel-Cantelli-Lemma): O.E. sei $\mathbb{E}(X) = 0$ und wir betrachten eine Folge $Y_1, Y_2, Y_3, \ldots$ auf $(\Omega, \mathcal{E}, P)$. Interessante Ereignisse: $B^c = \Omega \setminus B$, wobei

$$B = \{\omega \in \Omega \mid Y_n(\omega) \to 0\}$$

Vorüberlegungen: Sei $\omega \in \Omega$ fest, dann ist

$$Y_n(\omega) \to 0 \Leftrightarrow |Y_n(\omega)| \geq \frac{1}{k} \quad \forall k \in \mathbb{N}$$

nur endlich oft.
Sei nun $E_k = \{\omega \in \Omega \mid |Y_n(\omega)| > \frac{1}{k} \text{ unendlich oft}\}$. Offenbar gilt $E_1 \subset E_2 \subset E_3 \subset \dots$ und $B = \bigcap_{k \geq 1} E_k^c$. Daher ist $P(B) = P(\bigcap_{k \geq 1} E_k^c) = \lim_{n \to \infty} P(E_n^c)$ (Stetigkeit von P) bzw. $P(B^c) = P(\bigcup_{k \geq 1} E_k) = \lim_{k \to \infty} P(E_k)$.
Also gilt: $P(E_k) = 0 \quad \forall k \in \mathbb{N} \Rightarrow P(B^c) = 0 \Leftrightarrow P(B) = 1$

Beweis *in 3 Schritten.*

(i) Konstruktion einer konvergenten Teilfolge von $\frac{S_n}{n}$

(ii) Abschätzen der weggelassenen Folgenglieder

(iii) Stochastische Variante der Dreiecksungleichung für (i) und (ii)

Also:

(i) Betrachte die Folge $A_{1,k}, A_{2,k}, \dots$ von Ereignissen mit

$$A_{n,k} = \{\omega \in \Omega \mid |Y_n(\omega)| > \frac{1}{k}\}$$

Dann ist gerade $E_k = \{\omega \in \Omega \mid \omega \in A_{n,k} \text{ unendlich oft}\}$.
Konvergente Teilfolge: Setze $\tilde{Y}_n = Y_{n^2} = \frac{S_{n^2}}{n^2}$ und betrachten die $A_{n^2,k}$. Wegen der Unabhängigkeit der X_i in $S_n = X_1 + \dots + X_n$ ist

$$\text{Var}(\frac{S_n}{n}) = \text{Var}(\frac{X_1 + \dots + X_n}{n}) = \frac{1}{n}\text{Var}(X)$$

und mit der Tschebyscheff-Ungleichung folgt:

$$P(A_{n^2,k}) = P(|\frac{S_{n^2}}{n^2}| > \frac{1}{k}) \leq \frac{k^2}{n^2}\text{Var}(X)$$

Borel-Cantelli:

$$\sum_{n \geq 1} P(A_{n^2,k}) \leq \sum_{k \geq 1} \frac{k^2}{n^2}\text{Var}(X) = C(k) \sum_{n \geq 1} \frac{1}{n^2} < \infty$$

$$\Rightarrow P(E_k) = 0 \quad \forall k \in \mathbb{N} \Rightarrow P(\{\omega \in \Omega \mid \frac{S_{n^2}(\omega)}{n^2} \to 0\}) = 1$$

(ii) Sei $m \in \mathbb{N}$ beliebig und $n = n(m)$, so dass

$$n^2 \leq m \leq (n+1)^2$$

Wir betrachten die Folge $\tilde{\tilde{Y}}_m = \frac{S_m - S_{n(m)^2}}{n(m)^2}$ für die sich analog zu (i) zeigen lässt, dass

$$P(\{\omega \in \Omega \mid |\frac{S_m - S_{n(m)^2}}{n(m)^2}| \to 0\}) = 1 \text{ für } m \to \infty$$

(iii) Seien $A = \{\omega \in \Omega \mid \lim\limits_{n\to\infty} \frac{S_{n^2}(\omega)}{n^2} \to 0\}$ und $B = \{\omega \in \Omega \mid \lim\limits_{m\to\infty} \frac{S_m - S_{n(m)^2}}{n(m)^2} \to 0\}$.
(i) und (ii): $P(A) = P(B) = 1$. Außerdem wissen wir, dass

$$P(A \cap B) = \underbrace{P(A) + P(B)}_{2} - \underbrace{P(A \cup B)}_{\leq 1}$$
$$\geq 1 \Rightarrow P(A \cap B) = 1$$

Nun ist

$$|\frac{S_m}{m}| \leq |\frac{S_m}{n(m)^2}| = |\frac{S_m - S_{n(m)^2}}{n(m)^2} + \frac{S_{n(m)^2}}{n(m)^2}| \leq |\frac{S_m - S_{n(m)^2}}{n(m)^2}| + |\frac{S_{n(m)^2}}{n(m)^2}| \to 0 \quad \forall \omega \in A \cap B$$

Also ist

$$A \cap B \subset \{\omega \in \Omega \mid |\frac{S_m(\omega)}{m}| \to 0\}$$

Aber $P(A \cap B) = 1$ und daher

$$P(\{\omega \in \Omega \mid \frac{S_n}{n} \to 0\}) = 1$$

□

3.3.3 Beispiel.

(i) "Die Bank gewinnt immer": Gegeben zwei Spieler A und B, wobei A in der n-ten Runde X_n erhält und B erhält Y_n.
Regel: A gewinnt Runde n, wenn $Z_n = X_n - Y_n > 0$. Annahme: (X_n, Y_n) seien unabhängige Kopien von (X, Y).
Was ist die Wahrscheinlichkeit, dass A nach n Runden mehr Punkte als B hat:

$$p_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{Z_n > 0\}^2}(\omega) \overset{n\to\infty}{\to} \mathbb{E}(\mathbb{1}_{\{z>0\}}) = p_1 = p$$

Konsequenz: Ist $p > \frac{1}{2}$, dann gewinnt A (Bank) langfristig mehr Runden als B.

(ii) Monte-Carlo-Verfahren
- Gesucht: $I(f) = \int_0^1 f(x)dx$ mit $f \in C([0,1])$
- Idee: Berechne Fläche von $G \subset [0,1] \times [0, f_{\max}]$ mit $f_{\max}$ als Maximum von f durch Zählen zufällig gleichverteilter Punkte in G.
- Monte-Carlo-Algorithmus (Fermi, Ulam, v.Neumann; Los Alamos):
  - Ziehe Zufallszahlen X_k, Y_k unabhängig und identisch verteilt, wobei $X_k \sim U(0,1)$, $Y_k \sim U(0, f_{\max})$ ($U(x,y)$ "=" gleichverteilt auf $[x,y]$)
  - Definiere Zählmaß $Z_k = \begin{cases} 1 & Y_k \leq f(x_k) \\ 0 & \text{sonst} \end{cases}$
  - Berechne

$$\mathbb{E}(Z) = 1 \cdot P(Y \leq f(x)) + 0 \cdot P(Y > f(x)) = P(\{X \leq 1\} \cap \{Y \leq f(x)\})\}$$
$$= P(\{(X,Y) \in G\}) = \int_0^1 \int_0^{f(x)} \frac{1}{f_{\max}} dydx = \frac{1}{f_{\max}} \int_0^1 f(x)dx$$

Starkes Gesetz der großen Zahlen liefert:

$$\frac{f_{\max}}{n}\sum_{k=1}^{n} Z_k \overset{\text{f.s.}}{\to} \int_0^1 f(x)dx$$

3.4 Zentraler Grenzwertsatz

- Das wichtigste Resultat der Wahrscheinlichkeitstheorie
- Der zentrale Grenzwertsatz ist eine Aussage über Konvergenz im Verteilungssinne (starkes/schwaches Gesetz der großen Zahlen: Konvergenz in/mit Wahrscheinlichkeit).

3.4.1 Beispiel[Konvergenz in Verteilung]**.**

(i) Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilte Zufallsvariablen, wobei X_n binomialverteilt zum Parameter $p_n \in [0,1]$ ist, d.h.

$$P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$$

Wir nehmen an, dass p_n der Gestalt sei, dass

$$\lim_{n\to\infty} np_n = \lambda > 0$$

Dann gilt

$$\lim_{n\to\infty}\binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$$

Also mit anderen Worten

$$X_n \overset{\text{i.V.}}{\to} \frac{\lambda^k}{k!}e^{-\lambda}$$

(ii) Approximation der hypergeometrischen Verteilung durch die Binomialverteilung:

$$N, N-M, M >> n \Rightarrow \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \approx \binom{n}{k}\frac{M}{N}^k(1-\frac{M}{N})^{n-k}$$

(iii) Es seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte Kopien einer Zufallsvariablen X. Dann gilt (falls $\mathbb{E}(X)$ und $\mathrm{Var}(X)$ existieren)

- $\frac{X_1+\ldots+X_n}{n} \overset{\text{f.s.}}{\to} \mathbb{E}(X) =: m$
- $\mathrm{Var}(\frac{X_1+\ldots+X_n}{n}) = \frac{\mathrm{Var}(X)}{n} =: \frac{\sigma^2}{n}$

D.h. die Varianz von $\frac{X_1+\ldots+X_n}{n}$ nimmt mit $\frac{1}{n}$ ab.
<u>Normierte Zufallsvariablen:</u> $\tilde{X_n} = \frac{X_n - m}{\sigma} \Rightarrow \mathbb{E}(\tilde{X_n}) = 0$, $\mathrm{Var}(\tilde{X_n}) = 1$.
<u>Frage:</u> $\frac{\tilde{X_1}+\ldots+\tilde{X_n}}{\sqrt{n}} \overset{\text{i.V.}}{\to} ?$

3.4.2 Satz[zentraler Grenzwertsatz]**.** Es seien $X_1, X_2, \ldots$ unabhängige Kopien einer normierten Zufallsvariablen X, d.h. $\mathbb{E}(X) = 0$ und $\mathrm{Var}(X) = 1$. Dann gilt

$$\frac{X_1+\ldots+X_n}{\sqrt{n}} \overset{\text{i.V.}}{\to} \underbrace{N(0,1)}_{\text{Normalverteilung mit EW 0 und Var 1}}$$

$$\Leftrightarrow \lim_{n\to\infty} P(a \leq \frac{X_1(\omega)+\ldots+X_n(\omega)}{\sqrt{n}} \leq b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx$$

3.4.3 Beispiel. n unabhängige Würfe einer fairen Münze ($p = \frac{1}{2}$).

$$\text{Normierte Verteilung}(\#\text{Kopf}) = \frac{\#\text{Kopf} - \overbrace{\frac{n}{2}}^{\text{Erwartungswert}}}{\frac{1}{2}\sqrt{n}} \xrightarrow{\text{i.V.}} N(0,1)$$

3.4.4 Korollar[Satz von Moivre-Laplace]**.** Es seien $X_1, X_2, \dots$ unabhängig bernoulliverteilt mit Erfolgswahrscheinlichkeit $p \in (0,1)$. Dann gilt

$$\frac{X_1 + \ldots + X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\text{i.V.}} N(0,1)$$

Beweis: Normierung: $X_k \mapsto \frac{X_k - \mathbb{E}(X_k)}{\sqrt{\text{Var}(X_k)}}$, $\mathbb{E}(X_k) = p$, $\text{Var}(X_k) = p(1-p)$ □

Beweis *vom zentralen Grenzwertsatz.* Wir betrachten die momentenerzeugenden Funktionen (MEF) von X

$$\begin{aligned}\mathbb{E}(e^{sX}) &= \mathbb{E}(1 + sX + \frac{s^2}{2}X^2 + \ldots) \\ &= 1 + s\mathbb{E}(X) + \frac{s^2}{2}\mathbb{E}(X^2) + \ldots\end{aligned}$$

D.h. $\frac{d^k}{ds^k}\mathbb{E}(e^{sX})\Big|_{s=0} = \mathbb{E}(X^k)$ k-tes Moment, wobei wir annehmen, dass alle Momente existieren.

Beweisidee:

- Konvergenz der MEF
- Limes der MEF ist die MEF von $N(0,1)$

Berechne die MEF von $Z_n = X_1 + \ldots + X_n$:

$$\begin{aligned}\mathbb{E}(e^{s\frac{Z_n}{\sqrt{n}}}) &= \mathbb{E}(e^{\frac{s}{\sqrt{n}}(X_1+\ldots+X_n)}) = \mathbb{E}e^{\frac{s}{\sqrt{n}}X_1} \cdot \ldots \\ &= \prod_{i=1}^{n} \mathbb{E}(e^{\frac{s}{\sqrt{n}}X_i}) = (\mathbb{E}(e^{\frac{s}{\sqrt{n}}X}))^n = (\mathbb{E}(1 + \frac{s}{\sqrt{n}}X + \frac{s^2}{2n}X^2 + \ldots))^n \\ &\xrightarrow{n\to\infty} (1 + \frac{s^2}{2n} + O(n^{-\frac{3}{2}})) = e^{\frac{s^2}{2}}\end{aligned}$$

Berechne die MEF von N(0,1)

$$\begin{aligned}\mathbb{E}_{N(0,1)}(e^{sX}) &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{(x-s)^2}{2}} e^{\frac{s^2}{2}}\, dx \\ &= e^{\frac{s^2}{2}}\end{aligned}$$

Aus der Konvergenz aller Momente (in Form der MEF) folgt Konvergenz im Verteilungssinne. □

3.5 Anwendungen von Grenzwertsätzen

3.5.1 Beispiel[Messung physikalischer Größen]**.**

- Messung einer Größe G_0 (z.B. eine Naturkonstante) mittels einer Messapparatur A; diese habe einen Messfehler ΔA_G
- Wiederholte unabhängige Messungen $G_1, G_2, G_3, \ldots$, nach starkem Gesetz der großen Zahlen gilt:
$$\frac{G_1 + \ldots + G_n}{n} \xrightarrow{\text{f.s.}} \mathbb{E}(G) = G_0$$
- Modellierung des Messfehlers Annahmen:
  - Messfehler beschränkt, d.h. $\Delta A_G < \infty$ und symmetrisch um G_0 verteilt, also
$$G_0 - \Delta A_G \leq G_i \leq G_0 + \Delta A_G$$
  - Messfehler seien gleichverteilt im Intervall $[G_0 - \Delta A_G, G_0 + \Delta A_G]$ (typische "zero-knowledge"-Annahme).
  - G_i gleichverteilt in $[G_0 - \Delta A_G, G_0 + \Delta A_G] \Rightarrow \mathbb{E}(G_i) = G_0$, $\text{Var}(G_i) = \frac{(\Delta A)^2}{3}$
- Wie muss nun n gewählt haben, damit $\frac{G_1+\ldots+G_n}{n} \approx G_0$ bis auf einen Fehler $\varepsilon > 0$?
- Zentraler Grenzwertsatz: $|\frac{G_1+\ldots+G_n}{n} - G_0| \overset{!}{\leq} \varepsilon$.
$$\Rightarrow P(|\frac{G_1 + \ldots + G_n - nG_0}{\Delta A_G \sqrt{\frac{n}{3}}}| \leq \delta) \approx \frac{1}{\sqrt{2\pi}} \int_{-\delta}^{\delta} e^{-\frac{x^2}{2}} dx$$
$$\Leftrightarrow P(|\frac{G_1 + \ldots + G_n}{n} - G_0| \leq \varepsilon) \approx \frac{1}{\sqrt{2\pi}} \int_{-\frac{\varepsilon}{\Delta A_G}\sqrt{3n}}^{\frac{\varepsilon}{\Delta A_G}\sqrt{3n}} e^{-\frac{x^2}{2}} dx \quad \text{mit } \delta = \frac{\varepsilon}{\Delta A_G}\sqrt{3n}$$
- Wir möchten, dass unsere Schätzung (Approximation) $S_n = \frac{G_1+\ldots+G_n}{n} \approx G_0$ bis auf einen Fehler $\varepsilon > 0$ exakt ist, und zwar mit 99%-iger Sicherheit. Wie groß muss n sein?
$$P(|S_n - G_0| \leq \varepsilon) \approx \frac{1}{\sqrt{2\pi}} \int_{-\frac{\varepsilon}{\Delta A_G}\sqrt{3n}}^{\frac{\varepsilon}{\Delta A_G}\sqrt{3n}} e^{-\frac{x^2}{2}} dx \overset{!}{=} 0,99$$

  Also nach n auflösen: In Tabellen steht so etwas wie $P(X \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx = F(a)$, also
$$P(a \leq X \leq b) = F(b) - F(a)$$
  Es gilt zudem, dass
$$0,99 = \frac{1}{\sqrt{2\pi}} \int_{-a(n)}^{a(n)} e^{-\frac{x^2}{2}} dx \Leftrightarrow \frac{0,99}{2} = \frac{1}{\sqrt{2\pi}} \int_{0}^{a(n)} e^{-\frac{x^2}{2}} dx$$

D.h. bestimme $a^* > a(n)$ aus Tabelle und löse $a^* = \frac{\varepsilon}{\Delta A_G}\sqrt{3n}$ auf, d.h. $n = \frac{1}{3}(\frac{a^* \Delta A_G}{\varepsilon})^2$

- Exemplarische Tabelle:

$F(Z)$	1	2	3	4
$0,0$	$0,5000$			
$0,1$		$\ddots$		
$0,2$			$\ddots$	
$\vdots$				1

3.5.2 Beispiel[Brownsche Bewegung]**.**

- Studien von Blütenpollen in Lösung (Robert Brown, 1827), Beobachtung: Pollen machen eine Zitterbewegung. Erklärung durch Einstein (1905) und Smoluchowski (1906) durch ein Partikelmodell: Wassermoleküle stoßen zufällig gegen die Polle $\Rightarrow$ Zitterbewegung.
- Pro Zeiteinheit $\Delta t = 1$ hat man N zufällige, unabhängige Stöße mit Wasserteilchen
- In Formeln:
  - Position des Teilchens vor Beobachtung: $x(0)$
  - Position des Teilchens nach der Zeit t: $x(t) = x(0) + \sum\limits_{i=1}^{Nt} \xi_i$, wobei ξ_i unabhängig und identisch verteilte Zufallsvariablen sind, die die Zufallsstöße beschreiben.
  - Modellannahme:
    - $\mathbb{E}(\xi_i) = 0$ (Stöße gleichverteilt aus allen Richtungen)
    - $\mathrm{Var}(\xi_i) = \mathbb{E}(\xi_i^2) = \frac{1}{N}$
- Zentraler Grenzwertsatz:
  - normierte Zufallsvariablen: $\xi_i = \frac{1}{\sqrt{N}}\eta_i$ mit $\mathbb{E}(\eta_i) = 0, \mathbb{E}(\eta_i^2) = 1$
  - Also:

$$x(t) = x(0) + \frac{1}{\sqrt{N}} \sum_{i=1}^{Nt} \eta_i \xrightarrow{N\to\infty} N(x(0), t)$$

4 Statistik

Zentrale Frage: Verteilungseigenschaften (z.B. Erwartungswert, Varianz, Verteilungstyp) zu gegebene Daten (z.B. Temperatur, Parteienpräferenzen) herausfinden.

Bemerkung[Klassifikation von Daten]**.**

- Nicht-parametrische Statistik: Nominale und ordinale Daten wie z.B. Parteien oder Ranglisten.
- Parametrische Statistik: Alle Daten, die metrisierbar sind, wie z.B. Niederschlagsmengen oder Körpergrößen.

Bemerkung[Qualität von Daten]**.** Sind die Daten unabhängig oder abhängig, stationär oder zeitabhängig, ...

4.1 Parametrische Statistik

4.1.1 Bemerkung[Typische Fragestellungen der parametrischen Statistik]**.**

(i) Schätzung von Parametern, z.B. Erwartungswert μ und Varianz σ^2 von $N(\mu, \sigma^2)$

(ii) Hypothesentest, z.B. die Antwort auf die Frage "Sind die Daten normalverteilt?" oder "Ist Medikament A besser als B?"

(iii) Konfidenzintervalle: Frage nach der Güte von Schätzungen bzw. von Tests.

4.1.2 Bemerkung[Struktur von parametrischen Verfahren]**.**

- Daten / Beobachtungen: $x \in X$ mit $X \subset \mathbb{R}$ bzw. $X \subset \mathbb{R}^n$. X wird die Stichprobenmenge mit Ereignisalgebra $\mathcal{E}_X$
- Familie von Verteilungen (Kandidatenverteilungen): $P_\Theta = \{P_\vartheta \mid \vartheta \in \Theta\}$ mit z.B. $\Theta \subset \mathbb{R}$
- Statistisches Modell: $M = (X, \mathcal{E}_X, P_\Theta)$ (Wahrscheinlichkeitsraum)

4.1.3 Beispiel[Familie von Verteilungen]**.**

- Messung mit Messfehler: Annahme, dass Fehler unabhängig und gleichverteilt aus $[0, \Delta A]$ kommen, d.h.

$$P_\Theta = \{f_\vartheta(x) = \frac{1}{\vartheta} \mid \vartheta > 0\} = \{P_\vartheta(X \leq a) = \frac{a}{\vartheta} \mid \vartheta > 0 \wedge a \in [0, \vartheta]\}$$

- Binomialverteilung (z.B. beim wiederholten Würfeln)

$$P_\Theta = \{P_\vartheta(X = k) = \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k} \mid \vartheta \in [0,1]\}$$

4.2 Parameterschätzung

Gegeben: M=(X,$\mathcal{E}_X$, P_Θ), wobei P_Θ fest mit unbekanntem $\vartheta \in \Theta \subset \mathbb{R}$.
Die Abbildung $T_n : X^n \to \Theta$ (n-fache Messung) mit $T_n(x_1, \dots, x_n) = \widehat{\vartheta}$ wird Schätzer genannt. T_n ist eine Zufallsvariable.

4.2.1 Definition. Es sei T_n ein Schätzer für $\vartheta \in \Theta$. Dann heißt T_n

- erwartungstreu, wenn $\mathbb{E}(T_n) = \vartheta$
- asymptotisch erwartungstreu, wenn $\mathbb{E}(T_n) \xrightarrow{n \to \infty} \vartheta$
- konsistent, wenn $T_n \xrightarrow{\text{i.W.}} \vartheta$ für $n \to \infty$

4.2.2 Definition. Ein Schätzer T_n heißt bester Schätzer, wenn

- T_n ist erwartungstreu und
- $\mathrm{Var}(T_n) \leq \mathrm{Var}(\widehat{T_n})$ für alle erwartungstreuen Schätzer $\widehat{T_n}$

4.2.3 Beispiel[Fortsetzung vom Beispiel mit der physikalischen Größenmessung]**.** Bestimmung des maximalen Messfehlers ΔA mit bekanntem G_0.

- Modellierungsannahme: Fehler seien unhabhängig und identisch gleichverteilt im Intervall $[0, \Delta A]$ (also ist ΔA unser unbekanntes ϑ).
- Problem: Gegeben seien $G_1, \ldots, G_n$ aus $[0, \Delta A]$. Wie groß ist das Intervall?
- Verschiedene Schätzer für $\vartheta = \Delta A$
  - $T_n = 2\frac{G_1+\ldots+G_n}{n} \xrightarrow{\text{f.s.}} \Delta A$, da $\mathbb{E}(G_i) = \frac{1}{\Delta A}\int_0^{\Delta A} x dx = \frac{\Delta A}{2}$.
    Eigenschaften von diesem T_n
    * T_n ist konsistent, da aus $T_n \xrightarrow{\text{f.s.}} \Delta A$ folgt, dass $T_n \xrightarrow{\text{i.W.}} \Delta A$.
    * T_n ist erwartungstreu, denn
    $$\mathbb{E}(T_n) = \frac{2}{n}\mathbb{E}(G_1 + \ldots + G_n) = \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}(G_i) = \Delta A$$
  - $\tilde{T}_n = \max\{G_1, \ldots, G_n\}$. Eigenschaften von diesem $\tilde{T}_n$:
    * $P(\tilde{T}_n \leq \Delta A - \varepsilon) = \prod\limits_{i=1}^{n} P(G_i \leq \Delta A - \varepsilon) = (\frac{1}{\Delta A}\int\limits_0^{\Delta A - \varepsilon} dx)^n = (1 - \frac{\varepsilon}{\Delta A})^n \to 0 \quad \forall \varepsilon \in (0, \Delta A]$, d.h. der Schätzer ist konsistent.
    * $\tilde{T}_n$ ist nicht erwartugstreu: Es gilt $\tilde{T}_n \leq \Delta A$ und es gilt
    $$\begin{aligned}\mathbb{E}(\tilde{T}_n) &= \int\limits_0^{\Delta A} x\frac{\delta}{\delta a}\big|_{a=x} P(\tilde{T}_n \leq a)dx = \int\limits_0^{\Delta A} x\frac{nx^{n-1}}{(\Delta A)^n}dx \\ &= \frac{n}{(\Delta A)^n}\int\limits_0^{\Delta A} x^n dx = \frac{n}{n+1}\Delta A < \Delta A \quad \forall n \in \mathbb{N}\end{aligned}$$
    d.h. $\tilde{T}_n$ unterschätzt ΔA systematisch
  - Neuer Versuch: $\tilde{\tilde{T}}_n = \frac{n+1}{n}\tilde{T}_n$
    * $\tilde{\tilde{T}}_n$ ist erwartungstreu, denn
    $$\mathbb{E}(\tilde{\tilde{T}}_n) = \frac{n+1}{n}\mathbb{E}(\tilde{T}_n) = \Delta A$$
    * $\tilde{\tilde{T}}_n$ ist sogar ein besserer Schätzer als T_n im Sinne von:
    $$\mathrm{Var}(\tilde{\tilde{T}}_n) < \mathrm{Var}(T_n) \quad \forall n \geq 2$$
    * Berechnen der Varianzen von T_n, $\tilde{\tilde{T}}_n$:
    $$\begin{aligned}\mathrm{Var}(T_n) &= \mathrm{Var}(\frac{2}{n}(G_1 + \ldots + G_n)) = \frac{4}{n^2}\mathrm{Var}(G_1 + \ldots + G_n) = \frac{4}{n}\mathrm{Var}(G_i) \\ &= \frac{(\Delta A)^2}{3n} \\ \mathrm{Var}(\tilde{\tilde{T}}_n) &= (\frac{n+1}{n})^2\mathrm{Var}(\tilde{T}_n) = (\frac{n+1}{n})^2(\mathbb{E}(\tilde{T}_n^2) - (\mathbb{E}(\tilde{T}_n))^2) \\ &= \frac{(\Delta A)^2}{n(n+2)} < \frac{(\Delta A)^2}{3n} \quad \text{für } n \geq 2\end{aligned}$$
    Kleinere Varianz bedeutet schnellere Konvergenz gegen den Parameter.

4.3 Maximum-Likelihood-Schätzer (MLS)

Idee: Für gegebene Daten $x \in X$ wählt man $\vartheta \in \Theta$ so, dass P_ϑ die Daten am besten erklärt, d.h. man maximiert die Wahrscheinlichkeit (engl. Likelihood) $x \in X$ mittels P_ϑ zu erzeugen.

4.3.1 Definition. Für ein festes Datum $x \in X$ heißt die Funktion

$$\vartheta \mapsto L(x \mid \vartheta) = \begin{cases} P_\vartheta(X = x) & X \text{ diskret} \\ f_\vartheta(x) = P_\vartheta(X \leq x) & X \text{ stetig verteilt} \end{cases}$$

die Likelihood des Parameters $\vartheta \in \Theta$.

4.3.2 Definition. Die Abbildung $T_n^{\text{MLS}}(X_1, \ldots, X_n) = \underset{\vartheta \in \Theta}{\operatorname{argmax}} \; L(X_1, \ldots, X_n \mid \vartheta) = L^{-1}(\{\max_{\vartheta \in \Theta} L(X_1, \ldots, X_n \mid \vartheta)\})$ wird MLS des Parameters ϑ genannt.

4.3.3 Bemerkung.

(i) Die Likelihood-Funktion ist die Reinterpretation der Wahrscheinlichkeit $x \mapsto P_\vartheta(\cdot)$ als Funktion $\vartheta \mapsto P_\vartheta(\cdot)$. Warnung: P_ϑ ist kein Wahrscheinlichkeitsmaß, kann insb. also auch größer als 1 werden.

(ii) Anstelle der Likelihood wird oft die Log-Likelihood

$$l(x \mid \vartheta) = \log L(x \mid \vartheta)$$

verwendet. Z.B.

$$L(x \mid \mu) = e^{-\frac{(x-\mu)^2}{2}}, \; l(x \mid \mu) = -\frac{1}{2}(x - \mu)^2$$

4.3.4 Beispiel.

(i) Fortsetzung des Messfehlerbeispiels: Seien $G_1, \ldots, G_n$ unabhängig und identisch gleichverteilt aus $[0, \Delta A]$. Was ist der MLS für ΔA?
Daten $g_1, \ldots, g_n \in [0, \Delta A]$ haben gemeinsame Dichte

$$f_{\Delta A}(g_1, \ldots, g_n) = \prod_{i=1}^{n} f_{\Delta A}(g_i) = \begin{cases} \frac{1}{(\Delta A)^n} & \text{falls } \Delta A \geq \max\{g_1, \ldots, g_n\} \\ 0 & \text{sonst} \end{cases}$$

MLS: Finde ΔA, das $\frac{1}{(\Delta A)^n}$ maximal werden lässt und $\geq \max g_1, \ldots, g_n$ ist.

$$\Rightarrow T_n^{\text{MLS}}(g_1, \ldots, g_n) = \max g_1, \ldots, g_n$$

(ii) n-facher Münzwurf: Erfolgswahrscheinlichkeit $p \in (0, 1)$. Das Datum $x \in X$ sei die Anzahl der Erfolge, d.h.

$$L(x \mid p) = P_p(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Einfacher: Maximieren der Log-Likelihood:

$$l(x \mid p) = \log \binom{x}{n} + x \log p + (n - x) \log(1 - p)$$

Erste Ableitung 0 setzen:

$$0 = \frac{\delta l}{\delta p} = \frac{x}{p} - \frac{n-x}{1-p} \Leftrightarrow p = \frac{x}{n}$$

$(\frac{\delta^2 l}{\delta p^2} < 0$ d.h. $p = \frac{x}{n} = \underset{p\in(0,1)}{\operatorname{argmax}}\, l(x \mid p))$.
$T_n^{\text{MLS}} = \frac{x}{n}$ ist konsistent und erwartungstreu: Schreibe x als $x = Y_1 + \ldots + Y_n$ mit $Y_i \in 0, 1$ mit $P(Y_i = 1) = 1 - P(Y_i = 0) = p$

$$\Rightarrow \mathbb{E}(T_n^{\text{MLS}}) = \frac{1}{n}\mathbb{E}(Y_1 + \ldots Y_n) = \frac{n}{n}\mathbb{E}(Y_i) = p \text{ (erwartungstreu)}$$

$$\frac{x}{n} = \frac{Y_1 + \ldots + Y_n}{n} \xrightarrow{\text{f.s.}} \mathbb{E}(Y_i) = p \text{ (konsistent: } \frac{x}{n} \xrightarrow{\text{f.s.}} p \Rightarrow \frac{x}{n} \xrightarrow{\text{i.W.}} p)$$

4.3.5 Bemerkung[Abschließende Bemerkungen]**.**

(i) Ein Schätzer sollte mindestens konsistent bzw. asymptotisch erwartungstreu sein, ansonsten gilt: Anything goes!

(ii) Erwartungstreue bedeutet, dass ein Schätzer im Mittel das Richtige tut (d.h. weder über- noch unterschätzt).

(iii) MLS bedeutet, dass das Modell $M = \{X, \mathcal{E}_X, P_\theta\}$ die vorhandenen Daten am besten erklärt (im Sinne der Maximierung der Wahrscheinlichkeit). MLS sind immer konsistent (Beweis schwierig).

(iv) Für normalverteilte Daten entspricht die MLS der Minimierung der Fehlerquadrate (Ausgleichsproblem):

$$\underset{\mu,\sigma}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \underset{\mu,\sigma}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^2} + n\log\sigma + \text{const}$$

4.4 Parameterschätzung: Konfidenzintervalle

Im Folgenden sei $M = \{X, \mathcal{E}_X, P_\Theta\}$ ein Modell zu gegebenen Beobachtungen $x_1, \ldots, x_n \in X$.
Bisher: Punktschätzung von $\vartheta \in \Theta \subset \mathbb{R}$ z.B. via Maximum-Likelihood-Schätzung

$$T_n(x_1, \ldots, x_n) = \underset{\vartheta\in\Theta}{\operatorname{argmax}}\, L(x_1, \ldots, x_n \mid \vartheta)$$

mit der Likelihood-Funktion $L(x \mid \vartheta) = \begin{cases} P_\vartheta(X = x) & \text{diskret} \\ f_\vartheta(x) & \text{stetig} \end{cases}$.

4.4.1 Beispiel. n-facher Münzwurf, $x = \#$Kopf mit unbekanntem Parameter $p \in (0, 1)$

$$L(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Schätzer: relative Häufigkeit $\frac{d}{dp}\log L(x \mid p) = 0 \Leftrightarrow p = \frac{x}{n}$.
Probleme:

- Maximum nicht eindeutig
- $T_n(x)$ sehr sensitiv gegenüber x

Jetzt: Schätzung eines Bereichs (Intervall), in dem $\vartheta \in \Theta$ mit hoher Wahrscheinlichkeit liegt.

4.4.2 Definition. Eine Abbildung $C : X \to 2^{\Theta}, x \mapsto C(x) \subset \Theta$ hießt Konfidenzintervall zum Irrtumsniveau $\alpha \in (0,1)$, wenn

$$\forall \vartheta \in \Theta \quad P_\vartheta(\{x \in X \mid \vartheta \in C(x)\}) \geq 1-\alpha \Leftrightarrow \inf_\vartheta P_\vartheta(\{x \in X \mid \vartheta \in C(x)\}) \geq 1-\alpha$$

4.4.3 Notiz[Problem]**.** Wie rechnet man P_ϑ und damit dann $C(x)$ aus?

4.4.4 Beispiel.

- Wieder Münzwurf mit dem Schätzer $T_n(x) = \frac{x}{n}$ für p (bester Schätzer).
- Wir setzen $C(x) = (\frac{x}{n} - \varepsilon, \frac{x}{n} + \varepsilon)$ mit $\varepsilon > 0$.
- Wie groß muss $\varepsilon = \varepsilon(\alpha)$ gewählt werden?
- Hier: $p \in C(x) \Leftrightarrow |\frac{x}{n} - p| < \varepsilon$, d.h.

$$P_\vartheta(\{x \in X \mid p \in C(x)\}) \geq 1-\alpha \Leftrightarrow P_\vartheta(|\frac{x}{n} - p| \geq \varepsilon) \leq \alpha$$

Zwei Möglichkeiten, P_ϑ zu berechnen:

(i) Tschebyscheff-Ungleichung:

$$P(|\frac{x}{n} - p| \geq \varepsilon) \leq \frac{\mathrm{Var}(\frac{x}{n})}{\varepsilon^2} = \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}, \text{ da } p(1-p) \text{ Parabel mit Max. bei } \frac{1}{4}$$

mit x Summe von bernoulliverteilten Zufallsvariablen.
Konfidenzintervall: $P_\vartheta(|\frac{x}{n} - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2} \leq \alpha$, d.h.

$$\frac{1}{4n\varepsilon^2} \leq \alpha \Rightarrow \epsilon = \frac{1}{2\sqrt{n\alpha}}$$

tut es. Zahlenbeispiel: $\alpha = 0.0025$, $n = 10^5$, $\varepsilon = 0,01$ (also Schätzung bis auf zwei Nachkommastellen genau).

(ii) Genauer: Zentraler Grenzwertsatz (Moivre-Laplace).

$$P_\vartheta(|\frac{x}{n} - p| < \varepsilon) = P_\vartheta(|\frac{x-np}{\sqrt{np(1-p)}}| < \varepsilon\sqrt{\frac{n}{p(1-p)}}) \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-\varepsilon\sqrt{\frac{n}{p(1-p)}}}^{\varepsilon\sqrt{\frac{n}{p(1-p)}}} e^{-\frac{x^2}{2}}\,dx$$

nach zentralem Grenzwertsatz für $n >> 1$. Berechnung von $\varepsilon = \varepsilon(\alpha)$ mittels Normalverteilungstabelle:

$$\frac{1}{\sqrt{2\pi}} \int\limits_{-a}^{a} e^{-\frac{x^2}{2}}\,dx = \frac{2}{\sqrt{2\pi}} \int\limits_{0}^{a} e^{-\frac{x^2}{2}}\,dx = \frac{2}{\sqrt{2\pi}}(\int\limits_{-\infty}^{a} \ldots - \underbrace{\int\limits_{-\infty}^{0} \ldots}_{=\frac{1}{2}}) = 2\underbrace{F(X \leq a)}_{\text{s. Tabelle}} - 1$$

Zahlenbeispiel: $F(X \leq a) \geq 1 - \frac{\alpha}{2}$. $\alpha = 0.025$, bestimme a, so dass $F(X \leq a) \geq 0,9875$.

$$a = \varepsilon \sqrt{\frac{n}{p(1-p)}} \geq 2\varepsilon\sqrt{n}$$

Für $n = 10^5$ ergibt sich $\varepsilon \approx 10^{-5}$, also genau auf 5 Nachkommastellen.

4.4.5 Bemerkung.

(i) Die Wahl eines Konfidenzintervalls ist nicht eindeutig; in der Regel hängt es von den Approximationen für P_ϑ ab (vgl. Beispiel 2 Nachkommastellen vs 5 Nachkommastellen).

(ii) In der Praxis soll das Konfidenzintervall möglichst klein gewählt werden. Wie?

- Beim Münzwurf für festes p und n gilt: $\sum_{x=0}^{x_{\min}-1} B(x;n,p)$, $\sum_{x=x_{\max}+1}^{n} B(x;n,p) \leq \frac{\alpha}{2}$, d.h. #Kopf liegt mit Wahrscheinlichkeit $\geq 1-\alpha$ in $\{x_{\min}, \ldots, x_{\max}\}$
- $C = \{(x,p) \in X \times \Theta \mid p \in C(x)\}$
- $C(p) = \{x \in X \mid (x,p) \in C\}$, es gilt $\inf_p P_p(C(p)) \geq 1-\alpha$
- Wähle nun $C(p)$ möglichst klein (d.h. $x_{\min}$ möglichst groß und $x_{\max}$ möglichst klein) und setze
$$C(x) := \{p \in \Theta \mid x \in C(p)\}$$
als Konfidenzintervall.

4.4.6 Beispiel. 2 normalverteilte Zufallsvariablen mit $X \sim N(\mu, 1)$, $Y \sim N(\lambda, 1)$ mit unbekanntem $\mu, \lambda \in \mathbb{R}$ und zwei Schätzungen:

$$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n}$$
$$\bar{Y}_m = \frac{Y_1 + \ldots + Y_m}{m}$$

Es sei $\bar{X}_n \neq \bar{Y}_m$. Folgt daraus, dass $\mu \neq \lambda$ (für ein gewisses $\alpha \in (0,1)$)?

4.5 Von Konfidenzintervallen zu Hypothesentests

4.5.1 Beispiel[Signifikanz unterschiedlicher Schätzungen]**.**

- Seien $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ unabhängig mit bekannter Varianz $\sigma_X^2 = \sigma_Y^2 = \sigma^2$.
- Gegeben seien Schätzungen $\bar{X}_n = \frac{X_1+\ldots+X_n}{n}$ und $\bar{Y}_m = \frac{Y_1+\ldots+Y_m}{m}$
- Nun sei $\bar{X}_n \neq \bar{Y}_m$. Können wir davon ausgehen, dass $\mu_X \neq \mu_Y$ innerhalb der statistischen Schwankungen der Schätzer?
- Müssen also testen, ob die Konfidenzintervalle von $\bar{X}_n$ und $\bar{Y}_m$ überlappen.
- Idee: Konfidenzintervall für $\vartheta = \mu_X - \mu_Y$

$$\bar{X}_n - \bar{Y}_m \sim N(\underbrace{\mu_X - \mu_Y}_{\text{Linearität}}, \sigma^2(\frac{1}{n} + \frac{1}{m})) \quad \text{,da } \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \text{ falls X,Y unabh.}$$

Also: $P_{\vartheta=0}(\{|\bar{X}_n - \bar{Y}_m| \leq \varepsilon\}) = P_{\mu_X=\mu_Y}(|\frac{(\bar{X}_n-\bar{Y}_m)\sqrt{nm}}{\sigma\sqrt{n+m}}| \leq \frac{\varepsilon}{\sigma}\sqrt{\frac{nm}{n+m}}) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\varepsilon}{\sigma}\sqrt{\cdots}}^{\frac{\varepsilon}{\sigma}\sqrt{\cdots}} e^{-\frac{x^2}{2}} dx$

- Setze ein Irrtumsniveau $0 < \alpha < 1$ fest und berechne $\varepsilon = \varepsilon(\alpha)$, so dass $P_{\mu_X=\mu_Y}(\ldots) \geq 1-\alpha$ wird. (Tabelle)
- Dann ist es plausibel anzunehmen, dass $\mu_X = \mu_Y$, wenn
$$|\bar{X}_n - \bar{Y}_m| \leq \varepsilon(\alpha)$$
- Interpretation: $[-\varepsilon, \varepsilon]$ ist der Bereich, der von $\bar{X}_n - \bar{Y}_m$ mit Wahrscheinlichkeit $1-\alpha$ getroffen wird, wenn $\mu_X = \mu_Y$.

4.6 Hypothesentests

(i) Für ein gegebenes Modell $M = \{X, \mathcal{E}_X, P_\Theta\}$ seien:

(ii) $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$ und die Aussagen
- $H_0 =$ "$\vartheta \in \Theta_0$" (Nullhypothese)
- $H_1 =$ "$\vartheta \in \Theta_1$" (Gegenhypothese)

(Im Beispiel: $\Theta = \mathbb{R}^2$, $\Theta_0 = \{\mu_X \neq \mu_Y\}$, $\Theta_1 = \{\mu_X = \mu_Y\}$)

(iii) Für ein vorgegebenes Irrtumsniveau $\alpha \in (0,1)$ definieren wir:
- Fehler (Irrtum) 1. Art: $\vartheta \in \Theta_0$, aber H_0 wird verworfen ($\Leftrightarrow H_1$ wird anerkannt)
- Fehler (Irrtum) 2. Art: $\vartheta \in \Theta_1$, aber H_0 wird akzeptiert ($\Leftrightarrow H_1$ wird verworfen)

Unser Test soll so sein, dass ein Fehler 1. Art mit Wahrscheinlichkeit $\leq \alpha$ vorkommt.

(iv) Entscheidungsregeln: $\varphi : X \to [0,1]$ ("Test" genannt):
- deterministisch: $\varphi(x) = \begin{cases} 0 & \Rightarrow H_0 \text{ wird angenommen} \\ 1 & \Rightarrow H_0 \text{ wird verworfen} \end{cases}$
- randomisiert: H_1 wird mit Wahrscheinlichkeit $\varphi(x) \in [0,1]$ angenommen.

4.6.1 Bemerkung.

(i) Im Beispiel waren $H_0 =$ "$\mu_X = \mu_Y$", $H_1 =$ "$\mu_X \neq \mu_Y$" und φ ein deterministischer Test
$$\varphi(x_1, \ldots, x_n, y_1, \ldots, y_m) = \begin{cases} 0 & , \bar{X}_n - \bar{Y}_m \in [-\varepsilon(\alpha), \varepsilon(\alpha)] \text{ zu } \alpha > 0 \\ 1 & , \text{sonst, dh wenn } \bar{X}_n - \bar{Y}_m \text{ außerhalb des Konf.intervalls liegen} \end{cases}$$

(ii) Die Hypothesen H_0 und H_1 sind nicht gleichberechtigt, und es ist Konvention, H_0 und H_1 so zu wählen, dass der Fehler 1. Art der gravierendere ist. Bsp: $H_0 =$ "Es brennt", $H_1 =$ "Blinder Alarm" oder $H_0 =$ "Der Angeklagte ist unschuldig", $H_1 =$ "..schuldig".

(iii) Der eigentliche Test $\varphi(x)$ wird erst nach dem Festlegen der Entscheidungsregel (d.h. nach Def. von φ) durchgeführt.

Was ist ein (guter) Test?

4.6.2 Definition[Gütefunktion]**.** Es sei $\varphi : X \to [0,1]$ gegeben.
Wir nennen $R = \{x \in X \mid \varphi(x) = 1\}$ die Ablehnungsregion (engl. "rejection region") und definieren die Gütefunktion

$$G_\varphi : \Theta \to [0,1], G_\varphi(\vartheta) = \mathbb{E}_\vartheta(\varphi(x))$$

4.6.3 Definition. Gilt $G_\varphi(\vartheta) \leq \alpha \quad \forall \vartheta \in \Theta_0$, dann heißt φ Test zum Irrtumsniveau $\alpha \in (0,1)$.

4.6.4 Bemerkung.

(i) Ist der Test deterministisch, so gilt $G_\varphi(\vartheta) = \mathbb{E}_\vartheta(\chi_R(x)) = P_\vartheta(X \in R)$

(ii) $G_\varphi(\vartheta)$ gibt den Fehler 1. Art an, wenn $\vartheta \in \Theta_0$; für $\vartheta \in \Theta_1$ ist $1 - G_\varphi(\vartheta)$ die Wahrscheinlichkeit für den Fehler 2. Art.

(iii) Idealerweise soll $G_\varphi(\vartheta)$ für $\vartheta \in \Theta_1$ möglichst nahe an 1 liegen.

4.6.5 Beispiel.

- H. behauptet, hellsehen zu können. Wir stellen H. auf die Probe und lassen H. die Farbe von verdeckten Spielkarten bestimmen.
- ad (i): Modell $n = 32$ (Spielkarten), $X = \{0, 1, \ldots, 32\}$

$$P_\vartheta(X = x) = \binom{32}{x} \vartheta^x (x - \vartheta)^{32-x}$$

- ad (ii): $\Theta = \{\frac{1}{2}\} \cup (\frac{1}{2}, 1]$ und $H_0 =$ "H. kann nicht hellsehen" $=$ "$\vartheta = \frac{1}{2}$"
  und $H_1 =$ "H. kann hellsehen" $=$ "$\vartheta \in (\frac{1}{2}, 1]$"

4.6.6 Bemerkung. Es seien

(i) ein parametrisches Modell $M = \{X, \mathcal{E}_X, P_\Theta\}$

(ii) Parameterbereich $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$ und Hypothesen $H_0 =$ "$\vartheta \in \Theta_0$", $H_1 =$ "$\vartheta \in \Theta_1$"

(iii) ein Irrtumsniveau $\alpha \in (0,1)$ (beschränkt die Wahrscheinlichkeit für Fehler 1. Art)

(iv) Entscheidungsregel (Test) $\varphi : X \to [0,1]$, die angibt, wann H_0 verworfen und H_1 akzeptiert wird (z.B. $\varphi(x) = \begin{cases} 0 & \Rightarrow H_0 \\ 1 & \Rightarrow H_1 \end{cases}$ deterministisch).

Zur Erinnerung: Gütefunktion $G_\varphi(\vartheta) = \mathbb{E}_\vartheta(\varphi)$. φ Test zum Niveau $\alpha \Leftrightarrow G_\varphi(\vartheta \in \Theta_0) \leq \alpha$

4.6.7 Beispiel. H. behauptet, hellsehen zu können. Wir lassen H. dieFarbe von $n = 32$ verdeckten Spielkarten erraten.

ad (i) Modell: $X = \{0, 1, \ldots, 32\} \ni x$ Anzahl der Treffer

$$P_\vartheta(X = x) = \binom{32}{x} \vartheta^x (1 - \vartheta)^{32-x}$$

ad (ii) $\Theta = \{\frac{1}{2}\} \cup (\frac{1}{2}, 1]$ mit $H_0 =$ "$\vartheta = \frac{1}{2}$" (H. kann nicht hellsehen) und $H_1 =$ "$\vartheta \in (\frac{1}{2}, 1]$" (H. kann hellsehen).

ad (iii) Setze $\alpha = 0.05$ für Fehler 1. Art.

ad (iv) Wähle einen deterministischen Test

$$\varphi(x) = \begin{cases} 0 & , x < c \Rightarrow H_0 \\ 1 & , x \geq c \Rightarrow H_1 \end{cases}$$

wobei c so gewählt ist, dass "Wahrscheinlichkeit für Fehler 1. Art" $= G_\varphi(\frac{1}{2}) \leq \alpha$.

$$G_\varphi(\frac{1}{2}) = P_{\vartheta=\frac{1}{2}}(X \geq c) = \sum_{x=c}^{32} \binom{32}{x} (\frac{1}{2})^x \cdot (\frac{1}{2})^{32-x} = \sum_{x=c}^{32} \binom{32}{x} (\frac{1}{2})^x \leq \alpha$$

ist für $c \geq 19$ erfüllt, d.h. H. muss mindestens 19 Karten richtig erraten.

- Fehler 2. Art: Z.B. sei $\vartheta = 0.6 \in \Theta_1$. Dann ist $P(\text{Fehler 2. Art}) = 1 - G_\varphi(0.6) = 1 - \sum_{x=19}^{32} \binom{32}{x} 0.6^x \cdot 0.4^{32-x} \approx 0.4$
- Müsste H. alle 32 Karten richtig erraten, entspräche das einem Irrtumsniveau $\tilde{\alpha} = (\frac{1}{2})^{32}$, selbst für $\vartheta = 0.99$ (d.h. H. kann ziemlich gut hellsehen), beträgt die Wahrscheinlichkeit für den Fehler 2. Art $P(\text{Fehler 2. Art}) = 1 - 0.99^{32} \approx 0.28$

4.7 Likelihood-Ratio-Tests

Auch genannt Likelihood-Quotienten-Tests bzw. Alternativtests. Es geht um Testverfahren mit der Eigenschaft

$$P(\text{Fehler 1. Art}) \leq \alpha \wedge P(\text{Fehler 2. Art}) \text{ minimal (bester Test)}$$

Vereinfachende Annahme: $\Theta = \{\vartheta_0\} \cup \{\vartheta_1\}$ bzw. $\Theta = \{0\} \cup \{1\}$ dh. wir betrachten Modelle der Form $M = \{X, \mathcal{E}_X, P_0 \cup P_1\}$ und Hypothesen $H_0 =$ "$\vartheta = 0$" bzw. $H_1 =$ "$\vartheta = 1$".
Idee: Man verwirft die Hypothese mit der geringeren Likelihood.

$$\text{Likelihood } L(x \mid \vartheta) = \begin{cases} P_\vartheta(X = x) & X \text{ diskret} \\ \frac{d}{dx} P_\vartheta(X \leq x) & X \text{ stetig} \end{cases}$$

4.7.1 Definition. Es sei $L_\vartheta(x) = L(x \mid \vartheta)$ eine Likelihood-Funktion. Dann heißt

$$R(x) = \begin{cases} \frac{L_1(x)}{L_0(x)} & , \text{falls } L_0(x) \neq 0 \\ \infty & , \text{falls } L_0(x) = 0 \end{cases}$$

Likelihood-Quotient (engl.: "likelihood ratio").

4.7.2 Definition[Neyman-Pearson-Tests (NP-Test)]**.** Es sei $\{X, \mathcal{E}_X, P_0 \cup P_1\}$ fest mit Hypothesen $H_0 =$ "$\vartheta = 0$", $H_1 =$ "$\vartheta = 1$". Die Abbildung $\varphi^* : X \to [0, 1]$ heißt Neyman-Pearson-Test zum Irrtumsniveau $\alpha \in (0, 1)$, wenn es ein $c^* > 0$ gibt, so dass

$$\varphi^*(x) = \begin{cases} 0 & (\Rightarrow H_0), R(x) < c^* \\ 1 & (\Rightarrow H_1), R(x) > c^* \\ \gamma & (H_1 \text{ mit Wahrscheinlichkeit } \gamma), R(x) = c^* \end{cases}$$

wobei γ so gewählt ist, dass

$$\alpha = \underbrace{\mathbb{E}_{\vartheta=0}(\varphi^*)}_{\text{Gütefunktion}} = P_{\vartheta=0}(\{x \mid R(x) > c^*\}) + \gamma P_{\vartheta=0}(\{x \mid R(x) = c^*\})$$

(Wahrscheinlichkeit für den Fehler 1. Art)

4.7.3 Satz[Neyman-Pearson-Lemma]**.** Sei $R(x)$ ein Likelihood-Quotient mit $L_0(x)+L_1(x) > 0 \; \forall x \in X$. Dann gilt:

(i) Jeder NP-Test zum Niveau α ist bester Test.

(ii) Es gibt immer einen NP-Test.

(iii) Jeder beste Test zum Niveau α ist fast sicher (d.h. bis auf Nullmenten $\subset X$) ein NP-Test.

Beweis:

(i) Wir müssen zeigen, dass $1 - \mathbb{E}_{\vartheta=1}(\varphi^*)$ minimal ist. D.h. wir müssen zeigen, dass $\mathbb{E}_{\vartheta=1}(\varphi^*) \geq \mathbb{E}_1(\varphi)$, wobei φ beliebiger Test zum Niveau α ist (d.h. $\mathbb{E}_0(\varphi) \leq \alpha$). O.E. nehmen wir an, dass $x \in X$ stetig sei und setzen

$$g(x) = (\varphi^*(x) - \varphi(x))(L_1(x) - c^* L_0(x)) \geq 0$$

Unter der Annahme $g \geq 0$ gilt dann

$$\begin{aligned} 0 \leq \int g(x)dx &= \int (\varphi^*(x) - \varphi(x))(L_1(x) - c^* L_0(x))dx \\ &= \int (\varphi^*(x) - \varphi(x)) f_1(x)dx - c^* \int (\varphi^*(x) - \varphi(x)) f_0(x)dx \\ &= \mathbb{E}_1(\varphi^* - \varphi) - c^* \mathbb{E}_0(\varphi^* - \varphi) \end{aligned}$$

D.h. $\mathbb{E}_1(\varphi^*) \geq \mathbb{E}_1(\varphi) \Rightarrow \varphi^*$ ist bester Test.
Es bleibt zu zeigen, dass $g \geq 0$. Wir unterscheiden drei Fälle:

a) $R = \frac{L_1}{L_0} < c^* \Leftrightarrow L_1 - c^* L_0 < 0 \wedge \varphi^* = 0$. Folglich ist $g \geq 0$

b) $R = \frac{L_1}{L_0} > c^* \wedge L_0 + L_1 > 0 \Rightarrow L_1 - c^* L_0 > 0 \wedge \varphi^* = 1 \Rightarrow g \geq 0$

c) $R = \frac{L_1}{L_0} = c^* \Rightarrow g = 0$

(ii) Wir müssen zeigen, dass $c^* > 0$ und $\gamma \in [0,1]$ existieren.

- Sei $c \geq 0$ und definiere $\alpha(c) = P_0(R > c)$ und $\bar{\alpha}(c) = P_0(R \geq c) \geq \alpha(c)$. Dann gilt $\alpha(c) = 1 - P_\gamma(R \leq c)$ mit $\alpha(0) = 1$ und $\lim\limits_{c \to \infty} \alpha(c) = 0$.
- $\alpha(c)$ erbt die Rechtsstetigkeit. Wir können daher setzen:

$$c^* = \inf\{c > 0 \mid \alpha(c) \leq \alpha\}$$

so dass $\bar{\alpha}(c^*) \geq \alpha \geq \alpha(c^*)$. Wir betrachten zwei Fälle:

a) $\alpha(c) = \bar{\alpha}(c)$. Dann $\bar{\alpha}(c) - \alpha(c) = P_0(R \geq c) - P_0(R > c) = P_0(R = c) = 0$ und wir können $\gamma = 0$ setzen, so dass $\varphi^*(x) = \chi_{\{R(x) > c^*\}}(x)$. In diesem Test ist $\mathbb{E}_0(\varphi^*) = \mathbb{E}_0(\chi_{\{R(x)>c^*\}}) = \alpha$

b) $\alpha(c^*) < \bar{\alpha}(c^*)$. Dann können wir $\gamma = \frac{\alpha - \alpha(c^*)}{\bar{\alpha}(c^*) - \alpha(c^*)} = \frac{\alpha - P_0(R > c^*)}{P_0(R = c^*)}$ setzen, denn dann ist $\mathbb{E}_0(\varphi^*) = \gamma P_0(R = c^*) + P_0(R > c^*) = \alpha$

(iii) Sei φ ein bester Test mit $\mathbb{E}_0(\varphi) = \alpha$. Wir müssen zeigen, dass

$$\varphi^*(x) = \varphi(x)$$

bis auf Nullmengen in X.
Da φ bester Test ist, gilt $\mathbb{E}_1(\varphi) = \mathbb{E}_1(\varphi^*)$. Aus (i) wissen wir zudem, dass $\mathbb{E}_1(\varphi^* - \varphi) \geq \int g \geq 0$. Also gilt $\int g = 0$, d.h. $g(x) = 0$ bis auf Nullmengen.
Nun ist $g = (\varphi^* - \varphi)(L_1 - c^* L_0)$. Für fast alle $x \in \{y \mid R(y) \neq c^*\}$ muss also $\varphi^*(x) = \varphi(x)$ sein.
Für $R(x) = c^*$, d.h. $L_1 - c^* L_0 = 0$ können zwei Fälle auftreten:

a) $\{x \mid R(x) = c^*\}$ ist eine Nullmenge.

b) $\{x \mid R(x) = c^*\}$ hat positives Maß, dann ist wegen $\mathbb{E}_0(\varphi) = \gamma P_0(R = c^*) + P_0(R > c^*) = \alpha$ und damit $\gamma = 0$ oder $\gamma = \frac{\alpha - P_0(R > c^*)}{P_0(R = c^*)}$, d.h. φ ist NP-Test.

□

4.8 Anwendungen Hypothesentests

- Alternativtests mit Modell $M = \{X, \mathbb{E}_X, P_0 \cup P_1\}$ mit Hypothesen $H_0 =$ "$\vartheta = 0$", $H_1 =$ "$\vartheta = 1$"

- Neyman-Pearson-Lemma: Alle Tests φ^* der Form

$$\varphi^*(x) = \begin{cases} 0 & , R(x) < c^* \\ 1 & , R(x) > c^* \\ \gamma & , R(x) = c^* \end{cases}$$

  mit R als Likelihood-Quotient, $c^* > 0$ und $\gamma \in [0,1]$ mit der Eigenschaft

$$\underbrace{\gamma P_0(\{x \mid R(x) = c^*\}) + P_0(\{x \mid R(x) > c^*\})}_{= \mathbb{E}_0(\varphi^*)} = x$$

  sind beste Tests. Umgekehrt sind alle besten Tests fast sicher von der oben genannten Form.

- $L(x \mid 1) >> L(x \mid 0) \Rightarrow$ wähle H_1

- $L(x \mid 0) >> L(x \mid 1) \Rightarrow$ wähle H_0

4.8.1 Beispiel[Filterung varrauschter Signale]**.** Problem: $\sigma^2 >> \mu_0 - \mu_1$.
Annahme: Signal ist normalverteilt mit unbekanntem Erwartungswert $\mu \in \mathbb{R}$ und bekannter Varianz σ^2.

(i) Modellierung: n unabhängige Messungen $x_1, \ldots, x_n \in \mathbb{R} = X$ und wir haben die Kandidatenverteilung $f_\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

(ii) Schätzer für μ: $\bar{x}_n = \frac{x_1 + \ldots + x_n}{n} \xrightarrow{\text{f.s.}} \mu$ für $n \to \infty$ (erwartungstreuer, konsistenter Maximum-Likelihood-Schätzer).

(iii) Definiere Beobachtungsdatum: $\bar{x}_n = \frac{x_1+\ldots+x_n}{n}$. Wollen die Likelihood-Funktion bzw. -Quotient haben:

$$R(x) = \frac{L(x \mid \mu_1)}{L(x \mid \mu_0)}$$

wir brauchen also die Likelihood-Funktion (Dichtefunktion) von $\bar{x}_n$. Wir wissen, dass $x_i \sim N(\mu, \sigma^2)$. Also gilt:

$$\bar{x}_n \sim N(\mu, \frac{\sigma^2}{n})$$

$\Leftrightarrow \bar{x}_n$ hat die Dichte $f_\mu(x) = \frac{1}{\sigma}\sqrt{\frac{n}{2\pi}}e^{-\frac{n(x-\mu)^2}{2\sigma^2}}$. Also ist

$$R(x) = \frac{f_{\mu_1}(x)}{f_{\mu_0}(x)} = e^{-\frac{n}{2\sigma^2}(\mu_1^2-\mu_0^2-2x(\mu_1-\mu_0))} = e^{-\frac{n(\mu_1-\mu_0)}{\sigma^2}(\frac{\mu_1+\mu_0}{2}-x)}$$

Z.B. $\mu_0 = 0, \mu_1 = 1$: $R(x) = e^{-\frac{n}{\sigma^2}(\frac{1}{2}-x)}$

(iv) Ein Neyman-Pearson-Test mit $c^* = 1$ wäre also $\varphi^*(x) = \mathbb{1}_{\{R(x)>1\}} = \mathbb{1}_{\{x>\frac{1}{2}\}}$ (das Ereignis $\{x \mid R(x) = 1\}$ hat Wahrscheinlichkeit 0, d.h. $P_0(\{x \mid R(x) = 1\}) = 0$, so dass $\gamma = 0$ ist).

(v) Irrtumsniveau von φ^*: Fehler 1. Art hat die Wahrscheinlichkeit

$$\alpha = \mathbb{E}_0(\varphi^*) = P_0(\{x \mid R(x) > 1\}) = P_0(\{x \mid x > \frac{1}{2}\})$$

Fehler 2. Art hat Wahrscheinlichkeit

$$\beta = 1 - \mathbb{E}_1(\varphi^*) = 1 - P_1(\{x \mid x > \frac{1}{2}\}) = P_1(\{x \mid x \leq \frac{1}{2}\}) = P_0(\{x \mid x \leq -\frac{1}{2}\})$$

Denn: $X \mapsto X + a \Rightarrow \mathbb{E}(X) \mapsto \mathbb{E}(X) + a$. Nun:

$$\beta = P_0(\{x \mid x \leq -\frac{1}{2}\}) = P_0(x \geq \frac{1}{2}) = P_0(x > \frac{1}{2}) = \alpha$$

In Zahlen:

a) $n = \sigma^2$, $\alpha = 1 - P_0(x \leq \frac{1}{2}) = 1 - F(\frac{1}{2}) \approx 0.31$

b) $\sigma^2 = 1$, $n = 100$: $\alpha = \beta \approx 0$.
Tabelle: Umrechnen von normalverteilten ZV mit Varianz $\frac{1}{n}$ in standardnormalverteilte ZV

$$x \sim N(0,1) \Leftrightarrow Y = \frac{1}{\sqrt{n}}X \sim N(0, \frac{1}{n})$$

4.8.2 Beispiel[Sequentielle Tests]**.**

- Wiederholte Beobachtung, solange

$$A < R(x) < B$$

und wähle H_0 wenn $R(x) \leq A$ bzw. H_1 wenn $R(x) \geq B$.