Wahrscheinlichkeitstheorie und Mathematische Statistik

Prof. Dr. rer. nat. Martin Redmann
Sommersemester 2025

Inhaltsverzeichnis

Diese Mitschrift basiert auf der gleichnamigen Vorlesung *Wahrscheinlichkeitstheorie und Mathematische Statistik*, gehalten im Sommersemester 2025 an der Universität Rostock.

Alle Rechte an Inhalt und Struktur der Lehrveranstaltung liegen bei dem Modulverantwortlichen, Prof. Dr. rer. nat. Martin Redmann, sowie der Universität Rostock.

Diese Mitschrift dient ausschließlich zu Lern- und Dokumentationszwecken. Eine kommerzielle Nutzung oder Weiterverbreitung ohne Zustimmung ist nicht gestattet.

0 Einleitung

Motivation

In der Praxis sind viele Größen, mit denen wir arbeiten, nicht exakt bekannt. Dazu zählen etwa die Wahrscheinlichkeit $\mathbb{P}(A)$ eines Ereignisses A, die Verteilungsfunktion F_X einer Zufallsgröße X, ihr Erwartungswert $\mathbb{E}[X]$ oder die Varianz $D^2[X]$. (Letztere misst die durchschnittliche Abweichung von X gegenüber ihrem Erwartungswert und dient damit als Maß für das Risiko)

Wenn wir diese Kenngrößen nicht wissen, bleibt uns nichts anderes übrig, als sie anhand von Beobachtungen zu schätzen. Es stellt sich also die grundlegende Frage: Was genau sind Beobachtungen im mathematischen Sinn, und wie lassen sich daraus fundierte Erkenntnisse ableiten?

Beispiel 0.1 (Versicherung und Produktion).

- $X \triangleq$ Schadenshöhe eines Autos in der Kfz-Versicherung
- $X \triangleq$ Anzahl fehlerhafter Teile pro Tag in einer Produktionsfirma

In beiden Fällen kann das jeweilige Unternehmen konkrete Daten beobachten oder messen: Wie hoch sind die Schäden? Wie viele fehlerhafte Teile wurden produziert?

Diese Beobachtungen erlauben Rückschlüsse auf die Verteilung F, den Erwartungswert $\mathbb{E}[X]$ und die Varianz $\mathbb{D}^2[X]$ von X, Größen welche bei Planung und Vorhersagen essenziell sind.

Was sind Beobachtungen im mathematischen Sinne?

Formal gesprochen handelt es sich bei Beobachtungen um Zufallsgrößen $X_1, \ldots, X_n$, die unabhängig und identisch verteilt sind mit gemeinsamer Verteilungsfunktion F. Diese heißen dann *unabhängige Beobachtungen* einer Zufallsgröße X bzw. eines zugrunde liegenden Zufallsvorgangs.

Die zugehörigen Realisierungen $x_1, \ldots, x_n \in \mathbb{R}$, also konkrete numerische Ergebnisse wie $X_i(\omega) = x_i$ für ein festes, aber unbekanntes ω, nennen wir *konkrete unabhängige* Beobachtungen von X.

Beispiel 0.2 (Schadensbeobachtung).
Eine Versicherung beobachtet n Schäden aus n Unfällen. Die Zufallsgrößen $X_1, \ldots, X_n$ sind a priori ungewiss, es ist also im Voraus unbekannt wie hoch ein Schaden sein wird. Sobald alle Unfälle eingetreten sind, erhalten wir dann konkrete Werte $x_1, \ldots, x_n$.

Die x_i können als Realisierungen der Schadenshöhen von den Zufallsgrößen X_i interpretiert werden.

Wann sind Beobachtungen unabhängig?

Unabhängigkeit ist dann gegeben, wenn zum Beispiel eine der folgenden Bedingungen erfüllt sind:

- Stichprobenziehung mit Zurücklegen: Der Auswahlvorgang hat keinen Einfluss auf die Grundgesamtheit.
- Jedes Element der Grundgesamtheit hat die gleiche Wahrscheinlichkeit, in die Stichprobe aufgenommen zu werden (z. B. bei einer telefonischen Umfrage).
- Es liegt kein Clustering vor, z. B. nicht nur Fahrgäste einer einzelnen Straßenbahn kontrollieren, da die i-te Ziehung sonst Einfluss auf die $(i+1)$-te Ziehung haben könnte.

Weitere praktische Beispiele

Beispiel 0.3 (Glücksspiel).
Ein Spieler wirft eine möglicherweise verzerrte Münze. Abhängig vom Ergebnis erhält er eine Auszahlung X:

- „Kopf“ $\to$ Gewinn von 2 €
- „Zahl“ $\to$ Verlust von 1 €

Um zu entscheiden, ob sich die Teilnahme lohnt, müssen wir die Wahrscheinlichkeit p für „Kopf“, sowie den Erwartungswert des Spiels $\mathbb{E}[X]$ und die zugehörige Varianz $D^2[X]$ kennen.

Beispiel 0.4 (Schraubenproduktion).
In einem Betrieb werden Schrauben mit Soll-Länge 5 cm produziert. Zur Qualitätskontrolle entnimmt man n Schrauben aus dem laufenden Prozess. Die i-te Messung wird durch die Zufallsgröße X_i beschrieben.

Ziel ist es nun, auf Grundlage dieser Messungen Aussagen über die Verteilung, den Erwartungswert oder die Varianz der Länge X zu treffen.

Verknüpfung zur Wahrscheinlichkeitstheorie

Die mathematische Statistik baut in vielen Aspekten direkt auf Ergebnissen der Wahrscheinlichkeitstheorie auf. Zu den wichtigsten theoretischen Sätzen gehören:

- **Gesetz der großen Zahlen**: garantiert Konvergenz des Stichprobenmittels,
- **Zentrale Grenzwertsatz**: Normalverteilung als Grenzverteilung.

Darüber hinaus benötigen wir weitere wichtige Begriffe der Wahrscheinlichkeitsrechnung:

- Ereignisse und Wahrscheinlichkeiten
- Zufallsgrößen und deren Verteilungen
- Konvergenzbegriffe für Zufallsgrößen
- charakteristische Funktionen
- Erwartungswert und Varianz
- stochastische Unabhängigkeit

Wir beginnen deshalb mit einer gezielten Wiederholung bzw. Erweiterung dieser Konzepte.

Zufallsgrößen als messbare Abbildungen

Eine Zufallsgröße X ist eine Abbildung

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

auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, wobei X messbar sein muss. Über diese Abbildung lässt sich eine Verteilung auf $\mathbb{R}$ induzieren.

Verteilung einer Zufallsgröße

Die Verteilung μ_X einer Zufallsgröße X ist das durch X induzierte Maß auf $\mathbb{R}$ definiert:

$$\mu_X(B) := \mathbb{P}\big(\{\omega \in \Omega \mid X(\omega) \in B\}\big) \quad \text{für } B \in \mathcal{B}(\mathbb{R}).$$

Dies ist auch als *Bildmaß* von $\mathbb{P}$ unter X bekannt.

Verteilungsfunktion

Die Verteilungsfunktion $F_X \colon \mathbb{R} \to [0,1]$ ist gegeben durch

$$F_X(x) := \mathbb{P}(X \leq x) = \mu_X\big((-\infty, x]\big).$$

Erwartungswert als Integral

Der Erwartungswert einer Zufallsgröße X wird als Lebesgue-Integral definiert:

$$\mathbb{E}[X] := \int_\Omega X(\omega)\ \mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} x\, \mu_X(\mathrm{d}x).$$

Dies kann auch als *Stieltjes-Integral* über F_X geschrieben werden kann:

$$\mathbb{E}[X] = \int_\mathbb{R} x\ \mathrm{d}F_X(x).$$

Dieser Zugang erlaubt es, Erwartungswerte und andere Kennzahlen auch für diskrete, stetige oder allgemeinere Verteilungen einheitlich zu behandeln:

Definition 0.5 (Stieltjes-Integral).
Sei $F\colon [a,b] \to \mathbb{R}$ eine monotone, rechtsseitig stetige Funktion (z. B. eine Verteilungsfunktion) und $g\colon [a,b] \to \mathbb{R}$ eine beschränkte, stetige Funktion. Dann definiert man das *Stieltjes-Integral* von g bezüglich F durch

$$\int_a^b g(x)\ \mathrm{d}F(x)$$

als den Grenzwert geeigneter Riemann-Stieltjes-Summen:
Sei $\big((x_i^n)_{i=0}^{N_n}\big)_{n\in\mathbb{N}}$ so gewählt, dass

$$a = x_0^n < x_1^n < \cdots < x_{X_N}^n = b \quad \text{und} \quad \max_{i=0,\ldots,N_n-1} |x_{i+1}^n - x_i^n| \xrightarrow{n\to\infty} 0$$

dann definieren wir im Falle der Existenz

$$\int_a^b g(x)\ \mathrm{d}F(x) := \lim_{n\to\infty} \sum_{i=0}^{N_n-1} g(x_i^n) \cdot \big(F(x_{i+1}^n) - F(x_i)^n)\big)$$

und das *uneigentliche Stieltjes-Integral* entsprechend durch

$$\int_\mathbb{R} g(x)\ \mathrm{d}F(x) := \lim_{\substack{a\to-\infty \\ b\to\infty}} \int_a^b \int_a^b g(x)\ \mathrm{d}F(x)$$

Dieses Integral verallgemeinert das gewöhnliche Riemann-Integral, indem nicht über ein Maß bezüglich der Länge (wie bei dx), sondern bezüglich der Änderung einer Funktion F integriert wird.

1 Ergänzugen/Wiederholung Wahrscheinlichkeitstheorie

1.1 Gesetz der großen Zahlen

Lemma 1.1 (Ungleichung von Markov).
Es sei X eine integrierbare Zufallsgröße, d. h. $\mathbb{E}|X| < \infty$, dann gilt für alle $\varepsilon > 0$:

$$\mathbb{P}\big(\{\omega : |X(\omega)| \geq \varepsilon\}\big) \leq \frac{1}{\varepsilon}\mathbb{E}|X|$$

Beweis.
Verwenden wir die Definition des Erwartungswertes mittels Stieltjes-Integral so ergibt sich die Ungleichung durch Anwendung der Linearität und Montonie des Integrals:

$$\begin{aligned}
\mathbb{E}|X| &= \int_{\mathbb{R}} |x| \, \mathrm{d}F_X(x) \\
&= \underbrace{\int_{\{x:|x|<\varepsilon\}} |x| \, \mathrm{d}F_X(x)}_{\geq 0} + \int_{\{x:|x|\geq\varepsilon\}} |x| \, \mathrm{d}F_X(x) \\
&\geq \int_{\{x:|x|\geq\varepsilon\}} |x| \, \mathrm{d}F_X(x) \\
&\geq \int_{\{x:|x|\geq\varepsilon\}} \varepsilon \, \mathrm{d}F_X(x) \\
&= \varepsilon \cdot \int_{\{x:|x|\geq\varepsilon\}} \mathrm{d}\mu_X(x) \\
&= \varepsilon \cdot \mu_X\big(\{x : |x| \geq \varepsilon\}\big) \\
&= \varepsilon \cdot \mathbb{P}\Big(\big\{\omega : X(\omega) \in \{x : |x| \geq \varepsilon\}\big\}\Big) \\
&= \varepsilon \cdot \mathbb{P}\big(\{\omega : |X(\omega)| \geq \varepsilon\}\big)
\end{aligned}$$

□

Bemerkung 1.2. Aus Lemma 1.1 folgt

$$\begin{aligned}
\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) = \mathbb{P}\big(\underbrace{|X - \mathbb{E}[X]|^2}_{\text{neue Zufallsgröße}} \geq \varepsilon^2\big) \\
&\leq \frac{1}{\varepsilon^2}\mathbb{E}|X - \mathbb{E}[X]|^2 \\
&= \frac{1}{\varepsilon^2} D^2[X]
\end{aligned}$$

Analog lässt sich auch die *Tschebyscheffsche Ungleichung* herleiten, für ein beliebiges $p > 0$ gilt

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{E|X|^p}{\varepsilon^p}$$

Definition 1.3 (Fast sichere Konvergenz). Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Eine Folge (X_n) von Zufallsgrößen konvergiert fast sicher gegen eine Zufallsgröße X, falls eine Nullmenge $N \in \mathcal{F}$, d.h. $\mathbb{P}(N) = 0$, existiert, so dass

$$\lim_{n\to\infty} d(X_n(\omega), X(\omega)) = 0 \qquad \text{für alle } \omega \in \Omega\backslash N$$

äquivalent dazu konvergiert eine Folge reeller Zufallsgrößen fast sicher, wenn

$$\mathbb{P}\Big(\Big\{\omega \in \Omega \;\Big|\; \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}\Big) = 1$$

Definition 1.4 (Stochastische Konvergenz).
Eine Folge (X_n) von Zufallsgrößen *konvergiert stochastisch* (oder in Wahrscheinlichkeit) gegen eine Zufallsgröße X, falls

$$\mathbb{P}\big(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\big) \xrightarrow{n\to\infty} 0 \quad \text{für alle } \varepsilon \geq 0$$

oder dazu äquivalent, falls

$$\mathbb{P}\big(\{\omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\}\big) \xrightarrow{n\to\infty} 1 \quad \text{für alle } \varepsilon > 0$$

Wir schreiben im folgenden auch $X_n \xrightarrow{\mathbb{P}} X$ mit $n \to \infty$ oder sagen, dass X_n mit Wkt. 1 konvergiert.

Definition 1.5 (Definition des schwachen Gesetzes der großen Zahlen).
Eine Folge integrierbarer Zufallsgrößen (X_n), d. h. $\mathbb{E}|X_n| < \infty \;\forall\, n$ genügt genau dann dem schwachen Gesetz der großen Zahlen, wenn

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k]) \xrightarrow{\mathbb{P}} 0, \quad \text{für } n \to \infty.$$

Ein Spezialfall bilden hier Beobachtungen: Seien $X_1, \ldots, X_n$ unabhängige Beobachtungen einer Zufallsgröße X, so gilt $\mathbb{E}[X_k] = \mathbb{E}[X]$ und der Term vereinfacht sich durch

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k]) = \frac{1}{n}\sum_{k=1}^{n} X_k - \mathbb{E}[X]$$

Satz 1.6 (Schwaches Gesetz der großen Zahlen, $L^1(\Omega, \mathcal{F}, \mathbb{P})$ Version).
Es sei (X_n) eine Folge von Zufallsgrößen X_n, welche unabhängig und gleichverteilt wie eine integrierbare Zufallsgröße X sind. Dann genügt (X_n) dem schwachen Gesetz der großen Zahlen (gemäß Definition 1.5), d. h. konkret, dass

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathbb{P}} \mathbb{E}[X], \quad \text{für } n \to \infty.$$

Ohne Beweis.

1.2 Charakteristische Funktionen

Bisher wurden Zufallsgrößen immer entweder durch ihre Verteilungsfunktion oder ihre Verteilungsdichte charakterisiert. Eine weitere solche Charakterisierung bildet die charakteristische Funktion. Sie dient als wichtiges Hilfsmittel für Rechnungen und Beweise (zum Beispiel beim zentralen Grenzwertsatz).

Definition 1.7 (Charakteristische Funktion). Sei X eine Zufallsgröße mit Verteilungsfunktion F, dann heißt für $t \in \mathbb{R}$

$$\Psi(t) := \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}F(x)$$

die charakteristische Funktion von X.

Für die Berechnung ergeben sich zwei Fälle:
Ist X diskret, so gilt

$$\Psi(t) = \sum_k e^{itx_k} \cdot p_k,$$

wobei (x_k) die Werte von X mit $p_k = \mathbb{P}(X = x_k)$ sind.
Ist X absolut stetig, so gilt

$$\Psi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,\mathrm{d}x,$$

wobei f die Dichte von X ist.

Im Fall von Zufallsvektoren $\vec{X} = (X_1, \dots, X_n)^T$ wird die charakteristische Funktion über das Skalarprodukt definiert:

$$\Psi(t) := \mathbb{E}[e^{i\cdot\langle t, \vec{X}\rangle}] \qquad \text{mit } t = (t_1, \dots, t_n)^T$$

Die Eindeutigkeit der charakteristischen Funktion kommt daher, dass sie die Fouriertransformierte der Verteilung von X ist, d.h. im absolutstetigen Fall die Transformation der Funktion f und im diskreten Fall die Transformation der Folge p_k. Bekannterweise ist die Fouriertransformation invertierbar und damit dient sie als Charakterisierung der Zufallsgröße. Diese Inverse können wir verwenden um aus bekannter charakteristischen Funktion die Dichte / Einzelwahrscheinlichkeiten der zugehörigen Zufallsgröße zu bestimmen:

Satz 1.8 (Inverse der charakteristischen Funktion). Sei X eine Zufallsgröße mit charakteristischen Funktion Ψ. Dann ergibt sich mittels inverser Fouriertransformation :

i)

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \Psi(t)\,\mathrm{d}t$$

ii)

$$p_k = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itx_k} \Psi(t)\,\mathrm{d}t$$

(wieder jeweils im absolutstetigen oder diskreten Fall.)

ohne Beweis.

Es folgen einige wichtige Eigenschaften der charakteristischen Funktion:

Lemma 1.9.

1. Die charakteristische Funktion Ψ ist wohldefiniert.
2. $\Psi(0) = 1$ und $|\Psi(t)| \leq 1$.
3. $\Psi_{aX+b} = e^{itb} \cdot \Psi(at)$ für $a, b, t \in \mathbb{R}$.
4. Wenn X endliches n-tes Moment hat, dann gilt für $k = 1, \dots, n$:

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} \Psi(t)\Big|_{t=0} = i^k \cdot \mathbb{E}[X^k]$$

5. Ψ ist gleichmäßig stetig.
6. Wenn X und Y unabhängig sind, dann gilt $\Psi_{X+Y} = \Psi_X \cdot \Psi_Y$

Beweis.

1.
$$\int_{-\infty}^{\infty} \underbrace{|e^{itx}|}_{=1} \, \mathrm{d}F(x) = \int_{-\infty}^{\infty} \mathrm{d}F(x) = F(x)\Big|_{\infty}^{\infty} = 1 < \infty$$

2.
$$\Psi(0) = \int_{-\infty}^{\infty} \underbrace{e^{i\cdot 0\cdot x}}_{=1} \mathrm{d}F(x) = 1$$
$$|\Psi(t)| \leq \int_{-\infty}^{\infty} |e^{i\cdot t\cdot x}| \, \mathrm{d}F(x) = 1$$

3.
$$\Psi_{aX+b} = \mathbb{E}[\exp(it\cdot(aX+b))] = \mathbb{E}[\exp(itaX+itb)] = e^{itb}\cdot\mathbb{E}[\exp(i\cdot(at)\cdot X)] = e^{ibt}\cdot\Psi(at)$$

4.
$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\Psi(t)\Big|_{t=0} = \frac{\mathrm{d}^k}{\mathrm{d}t^k}\int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}F(x)\Big|_{t=0} = \int_{-\infty}^{\infty} \frac{\mathrm{d}^k}{\mathrm{d}t^k} e^{itx}\Big|_{t=0} \mathrm{d}F(x) = i^k \cdot \int_{-\infty}^{\infty} x^k \,\mathrm{d}F(x) = i^k \cdot \mathbb{E}[X^k]$$

Das Vertauschen von Integral und Ableitung ist (nach Leibniz Regel) erlaubt, da
$$\left|\frac{\mathrm{d}^k}{\mathrm{d}t^k} e^{itx}\right| = |i^k x^k e^{itx}| = |x^k|$$
und
$$\int_{-\infty}^{\infty} |x^k| \, \mathrm{d}F(x) = \mathbb{E}|X|^k < \infty$$

5. Als Erinnerung: Gleichmäßig stetig heißt, dass $\sup_t |\Psi(t+h) - \Psi(t)| \to 0$ für $h \to 0$:
$$|\Psi(t+h)-\Psi(t)| = \left|\mathbb{E}\big[e^{i(t+h)X}\big] - \mathbb{E}\big[e^{itX}\big]\right| = \left|\mathbb{E}\big[e^{itX}\cdot(e^{ihX}-1)\big]\right| \leq \mathbb{E}\big[|e^{itX}|\cdot|e^{ihX}-1|\big] = \mathbb{E}|e^{ihX}-1|$$
Demnach gilt für das Supremum:
$$\sup_t |\Psi(t+h) - \Psi(t)| \leq \mathbb{E}|e^{ihX} - 1| \to 0 \quad \text{für } h \to 0$$

6.
$$\Psi_{X+Y}(t) = \mathbb{E}\big[e^{it(X+Y)}\big] = \mathbb{E}\big[e^{itX}\cdot e^{itY}\big] = \mathbb{E}\big[e^{itX}\big]\cdot\mathbb{E}\big[e^{itX}\big] = \Psi_X(t)\cdot\Psi_Y(t)$$

□

Korollar 1.10. Aus Lemma 1.9 (4) ergibt sich folgende Reihenentwicklung:
$$\Psi(t) = \sum_{k=0}^{\infty} \frac{i^k\cdot\mathbb{E}[X^k]}{k!}\cdot t^k$$

Beispiel 1.11 (charakteristische Funktion der Poisonverteilung). Sei X poissonverteilt mit Parameter $\lambda > 0$ ($X \sim \mathrm{Poi}(\lambda)$), dann gilt:
$$\Psi(t) = \mathbb{E}[e^{itX}] = \sum_{k=0}^{\infty} e^{itk}\cdot\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\cdot\sum_{k=0}^{\infty}\frac{(\lambda\cdot e^{it})^k}{k!} = e^{-\lambda}\cdot e^{\lambda\cdot e^{it}} = \exp\big(\lambda(1-e^{it})\big)$$

Beispiel 1.12 (charakteristische Funktion der Normalverteilung). Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ normalverteilt. Zur Bestimmung von Ψ_X betrachten wir $Z \sim \mathcal{N}(0,1)$, da $X = \sigma \cdot Z + \mu$.
Für die Standardnormalverteilung gilt:

$$\mathbb{E}[Z^k] = \begin{cases} 0 & k \text{ ungerade} \\ (k-1)!! & k \text{ gerade}^1 \end{cases}$$

Aus Korollar 1.10 folgt damit

$$\Psi_Z(t) = \sum_{k=0}^{\infty} \frac{(it)^k \cdot \mathbb{E}[Z^k]}{k!} = \sum_{k=0}^{\infty} (it)^{2k} \cdot \frac{1 \cdot 3 \cdot 5 \cdot \dots \cdot (2k-1)}{1 \cdot 2 \dots (2k-1) \cdot 2k} = \sum_{k=0}^{\infty} \frac{(it)^{2k}}{2 \cdot 4 \dots 2k}$$

Aus $i^2 = -1$ und $2 \cdot 4 \dots 2k = 2^k \cdot k!$ ergibt sich

$$\Psi_Z(t) = \sum_{k=0}^{\infty} \frac{(it)^{2k}}{2 \cdot 4 \dots 2k} = \sum_{k=0}^{\infty} dfrac(-t^2 \,/\, 2)^k k! = e^{-t^2 \,/\, 2}$$

Für $\Psi_X(t)$ ergibt sich nach Lemma 1.9 (3)

$$\Psi_X(t) = \Psi_{\sigma Z + \mu}(t) = e^{it\mu} \cdot \Psi_Z(\sigma t) = e^{it\mu} \cdot e^{-\frac{\sigma^2 t^2}{2}}$$

Beispiel 1.13 (Summe von Normalverteilungen). Es seien $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ unabhängig. Nach Lemma 1.9 (6) gilt:

$$\Psi_{X+Y}(t) = \Psi_X(t) \cdot \Psi_Y(t) = e^{it\mu_1} \cdot e^{-\frac{\sigma_1^2 t^2}{2}} \cdot e^{it\mu_2} \cdot e^{-\frac{\sigma_2^2 t^2}{2}} = e^{it(\mu_1+\mu_2)} \cdot e^{-\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

Wegen der Eindeutigkeit der charakteristischen Funktion folgt daraus, dass $X+Y \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Satz 1.14. Sei (F_n) eine Folge von Verteilungsfunktionen mit zugehörigen charakteristischen Funktionen (Ψ_n) und sei F eine Verteilungsfunktion mit charakteristischen Funktion (Ψ). Dann gilt:

$$F_n \to F \text{ punktweise}^2 \iff \Psi_n \to \Psi \text{punktweise}$$

ohne Beweis.

1.3 Zentraler Grenzwertsatz und Konvergenz

Definition 1.15 (Konvergenz in Verteilung). Eine Folge von Zufallsgrößen (X_n) mit Verteilungsfunktionen (F_{X_n}) konvergiert in Verteilung gegen eine Zufallsgröße X, wenn

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

an allen Stetigkeitsstellen der Verteilungsfunktion F_X von X. Wir schreiben dann $X_n \xrightarrow{d} X$.

Eine äquivalente Definition besagt, dass $X_n \xrightarrow{d} X$ falls $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ für alle stetigen, beschränkten Funktionen g. Daraus folgt insbesondere, dass für derartige g auch $g(X_n) \xrightarrow{d} g(X)$.

Satz 1.16. Wenn $X_n \xrightarrow{\mathbb{P}} X$, dann gilt $X_n \xrightarrow{d} X$.
Die Umkehrung gilt im Allgemeinen nicht, außer $X = c$ ist konstant.

Definition 1.17. Wir führen folgende Schreibweisen ein:

[2] an allen Stetigkeitsstellen

1. $S_n := \sum_{i=1}^{n} X_i$
2. $D^2[X] := \mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$

Korollar 1.18. Für die Varianz der Summe von unabhängigen, gleichverteilten (i.i.d.) Zufallsgröße gilt

$$D^2[S_n] = D^2\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} D^2[X_i] = n \cdot D^2[X_1]$$

Satz 1.19 (Zentraler Grenzwertsatz). Sei (X_n) eine Folge von i.i.d. Zufallsgrößen mit $D^2[X_1] < \infty$, so gilt

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{D^2[S_n]}} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$$

Beweis. Es gilt für die charakteristische Funktion von $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{D^2[S_n]}}$:

$$\begin{aligned}
\Psi_n(t) &= \mathbb{E}\left[\exp\left(it \cdot \frac{S_n - \mathbb{E}[S_n]}{\sqrt{D^2[S_n]}}\right)\right] \\
&= \mathbb{E}\left[\exp\left(i \cdot \frac{t}{\sqrt{n \cdot D^2[X_1]}} \cdot \sum_{j=1}^{n}(X_j - \mathbb{E}[X_1])\right)\right] \\
&= \mathbb{E}\left[\prod_{j=1}^{n} \exp\left(i \cdot \frac{t}{\sqrt{n \cdot D^2[X_1]}} \cdot (X_j - \mathbb{E}[X_1])\right)\right] \\
&= \prod_{j=1}^{n} \mathbb{E}\left[\exp\left(i \cdot \frac{t}{\sqrt{n \cdot D^2[X_1]}} \cdot (X_j - \mathbb{E}[X_1])\right)\right] \\
&= \mathbb{E}\left[\exp\left(i \cdot \frac{t}{\sqrt{n \cdot D^2[X_1]}} \cdot (X_1 - \mathbb{E}[X_1])\right)\right]^n \\
&= \Psi_Y\left(\frac{t}{\sqrt{n \cdot D^2[X_1]}}\right)^n
\end{aligned}$$

wobei $Y = X_1 - \mathbb{E}[X_1]$.

Unter Verwendung der Taylor-Formel ergibt sich

$$\begin{aligned}
\Psi_Y\left(\frac{t}{\sqrt{n \cdot D^2[X_1]}}\right) &= \underbrace{\Psi_Y(0)}_{=0} + \underbrace{\Psi_Y'(0)}_{=i\mathbb{E}Y=0} \cdot \frac{t}{\sqrt{n \cdot D^2[X_1]}} + \underbrace{\Psi_Y''(0)}_{=i^2\mathbb{E}Y^2=-D^2[X_1]} \cdot \frac{t^2}{2n \cdot D^2[X_1]} + o\left(\frac{t^2}{n}\right) \\
&= 1 + \frac{-\frac{t^2}{2} + o(\frac{t^2}{n})}{n}
\end{aligned}$$

Da $o(\frac{t^2}{n}) \to 0$ für $n \to \infty$ folgt

$$\Psi_n(t) = \left(1 + \frac{-\frac{t^2}{2} + o(\frac{t^2}{n})}{n}\right)^n \to e^{-t^2/2}$$

Dies ist genau die charakteristische Funktion der Standardnormalverteilung und damit folgt nach Satz 1.14 die Konvergenz in Verteilung gegen $\mathcal{N}(0,1)$. □

Bemerkung 1.20. Äquivalent umgeformt besagt der der zentrale Grenzwertsatz, dass wir Summen von i.i.d. Zufallsgrößen durch eine Normalverteilung annähern können, denn Umformungen ergeben

$$S_n \approx \mathcal{N}\left(\mathbb{E}[S_n], D^2[S_n]\right)$$

und damit

$$\mathbb{P}(S_n \leq x) \approx \Phi\left(\frac{y - \mathbb{E}[S_m]}{\sqrt{D^2[S_n]}}\right).$$

Beispiel 1.21 (Normalapproximation). Seien $X_1, \ldots, X_n$ Bernoulli verteilt zum Parameter 1, d.h.

$$X_k := \begin{cases} 1, & \text{Ereignis} A \text{trifft ein, Wahrscheinlichkeit: } p = \mathbb{P}(A). \\ 0, & \text{Ereignis} A \text{trifft nicht ein, Wahrscheinlichkeit: } 1 - p = \mathbb{P}(A^c). \end{cases}$$

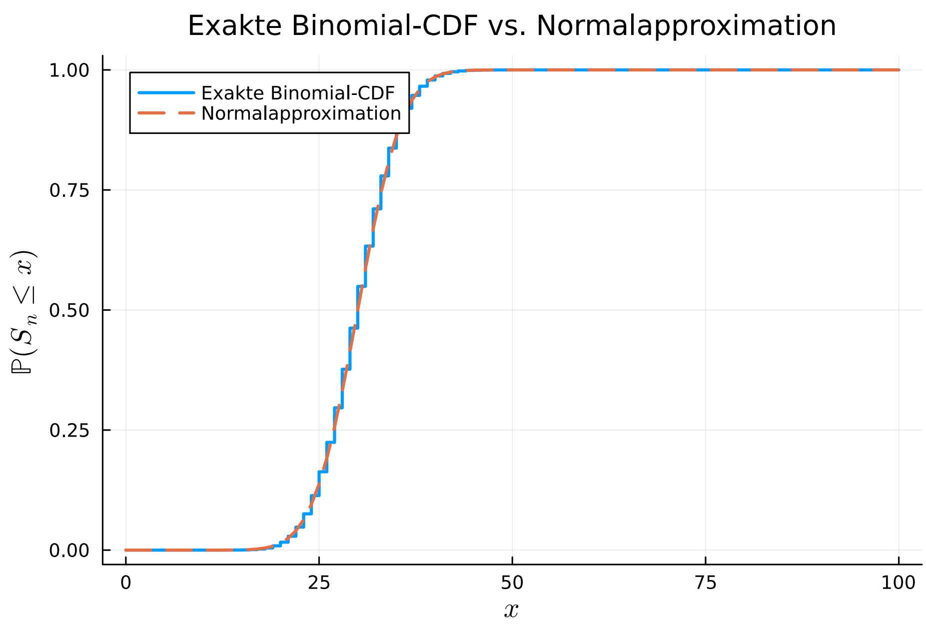
Die Summe $S_n \sim \mathrm{Bin}(n,p)$ gibt demnach die Anzahl an Erfolgen (A trifft ein) nach n unabhängigen Versuchen an. Dabei gilt:

$$\mathbb{P}(S_n \leq x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Dabei ist der Term $\binom{n}{k} = n \cdot (n-1) \ldots (n-k+1)$ für große n teuer zu berechnen. Nutzen wir stattdessen die Annäherung durch die Standardnormalverteilung so ergibt sich:

$$S_n \approx \mathcal{N}(np, np(1-p)) \implies \mathbb{P}(S_n \leq x) = \frac{1}{\sqrt{2\pi n^2 p^2 (1-p)^2}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{x - np}{np(1-p)}\right)^2\right)$$

Bereits für das Beispiel $n = 100$, $p = 0.5$ ergibt sich folgender Vergleich:



mit einem maximalen quadratischen Fehler von unter 0.003.

Beispiel 1.22. Als Anwendungsbeispiel betrachten wir folgenden Kontext:

40 % der Menschen in Deutschland haben Blutgruppe Null. Es wird eine Stichprobe von $n = 1000$ zufällig ausgewählten Personen gezogen. Wie große ist die Wahrscheinlichkeit, dass zwischen 30 % bis 50 % der Personen die Blutgruppe Null haben?

Wir definieren Y als die Anzahl der Personen mit Blutgruppe Null aus den $n = 1000$ zufällig befragen Personen, dann ist $Y \sim \mathrm{Bin}(0.4, 1000)$. Für die Wahrscheinlichkeit, dass zwischen 38 % bis 42 % der Befragten, also zwischen 380 und 420 Personen, die Blutgruppe Null haben, ergibt sich

$$\begin{aligned}
\mathbb{P}(380 < Y \leq 420) &= \mathbb{P}(Y \leq 380) - \mathbb{P}(Y \leq 420) \\
&\approx \Phi\left(\frac{420 - 1000 \cdot 0.4}{\sqrt{1000 \cdot 0.4 \cdot (1-0.4)}}\right) - \Phi\left(\frac{380 - 1000 \cdot 0.4}{\sqrt{1000 \cdot 0.4 \cdot (1-0.4)}}\right) \\
&= \Phi\left(\frac{20}{\sqrt{240}}\right) - \Phi\left(-\frac{20}{\sqrt{240}}\right) \\
&= 2\Phi\left(\frac{20}{\sqrt{240}}\right) - 1 \approx 0.803
\end{aligned}$$

Beispiel 1.23 (Stichprobenumfang bei Umfragen). Ein unbekannter Anteil p von Menschen wählt eine Partei. Dieses p soll in einer Umfrage ermittelt werden.

Gefragt ist, wie groß dem Umfang n einer Stichprobe sein muss, so dass die Umfrage aussagekräftig ist.

Dabei sei $X_1, \ldots, X_n$ unsere mathematische Stichprobe (Bernoulli verteilt mit Parameter p). Es ergibt sich dabei, dass S_n die Anzahl der Wähler gesuchter Partei aus n Befragten ist.

S_n ist binomialverteilt und $\hat{p}_n = \frac{S_n}{n}$ ist die relative Häufigkeit des Ereignisses und dient als Schätzer für den echten Anteil p.

Unser Ziel ist, dass wir mit einer Wahrscheinlichkeit von mindestens 95 % eine Abweichung von maximal 2 % zum echten Anteil haben, d.h. wir suchen n, so dass $\mathbb{P}(|p - \hat{p}_n| \leq 0.02) \geq 0.95$.

Es gilt

$$\mathbb{P}(|p - \hat{p}_n| \leq 0.02) = \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq 0.02\right) = \mathbb{P}\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leq \frac{0.02n}{\sqrt{np(1-p)}}\right)$$

Durch $p(1-p) = \frac{1}{4} - (p - \frac{1}{2})^2 \leq \frac{1}{4}$ ergibt sich

$$\mathbb{P}(|p - \hat{p}_n| \leq 0.02) \geq \mathbb{P}\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leq \frac{0.02n}{\sqrt{n\frac{1}{4}}}\right) = \mathbb{P}\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leq 0.04 \cdot \sqrt{n}\right)$$

Durch unsere Annäherung erhalten wir

$$\mathbb{P}(|p - \hat{p}_n| \leq 0.02) \geq \Phi(0.04\sqrt{n}) - \Phi(0.04\sqrt{n}) = 2\Phi(0.04\sqrt{n}) - 1$$

Damit $\mathbb{P}(|p - \hat{p}_n| \leq 0.02) \geq 0.95$ muss demnach $\Phi(0.04\sqrt{n}) \geq 0.975$, also $0.04\sqrt{n} \geq 1.96$ bzw. $n \geq 2401$.

Satz 1.24 (Slutsky's Theorem). Seien (X_n) und (Y_n) Folgen von zufälligen Vektoren mit $X_n \xrightarrow{d} X$ und $Y_n \xrightarrow{d} Y$, wobei X und Y ebenfalls zufällige Vektoren sind.

1. Im Allgemeinen gilt nicht, dass $(X_n + Y_n)$ gegen $X + Y$, oder dass $(X_n^T Y_n)$ gegen $X^T Y$ in Verteilung konvergiert.

2. Im Fall, dass $Y = c$ konstant ist, gilt jedoch $X_n + Y_n \xrightarrow{d} X + c$ und $X_n^T Y_n \xrightarrow{d} X^T c$.

Beweis. Für ersten genügt ein Gegenbeispiel, sei hierfür (Z_i) eine beliebige Folge von zentrierten Zufallsgrößen (d.h. $\mathbb{E}Z_n = 0$) und sei $S_n = \sum_{i=1}^n Z_n$, so gilt nach dem zentralen Grenzwertsatz:

$$X_n := \frac{S_n}{\sqrt{D^2[S_n]}} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$$
$$Y_n := \frac{-S_n}{\sqrt{D^2[S_n]}} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$$

Offensichtlich gilt jedoch $X_n + Y_n = 0 \xrightarrow{d} 0$, aber $Z + Z \sim \prime, \triangle$.

Das gleiche Beispiel fungiert auch für den Fall $(X_n^T Y_n)$. Hierbei gilt

$$X_n \cdot Y_n = -\left(\frac{S_n}{\sqrt{D^2[S_n]}}\right)^2 \xrightarrow{d} -Z^2$$

Jedoch haben Z^2 und $-Z^2$ verschiedene Verteilungen, da $\mathbb{P}(Z^2 \leq 0) = 0 \neq 1 = \mathbb{P}(-Z^2 \leq 0)$.

Der zweite Teil des Beweises sei ausgelassen. □

Satz 1.25. Seien (X_n) und (Y_n) unabhängige Folgen von Zufallsgrößen mit $X_n \xrightarrow{d} X$ und $Y_n \xrightarrow{d} Y$, wobei X und Y ebenfalls unabhängige Zufallsgrößen sind, so gilt

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ Y \end{pmatrix}$$

Sollte die Unabhängigkeit nicht gelten, so ist die Konvergenz im Allgemeinen nicht gegeben.

Beweis. Für die Verteilungsfunktion von unabhängigen Zufallsgrößen gilt:

$$F_{X_n,Y_n}(x,y) = F_{X_n}(x) \cdot F_{Y_n}(y) \xrightarrow{n\to\infty} F_X(x) \cdot F_Y(x) = F_{X,Y}(x,y)$$

Als Gegenbeispiel dient $X, Y \sim \mathcal{N}(0,1)$ mit $Y_n \equiv X_n \equiv X$. Es folgt[3]

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ X \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right) \neq \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \sim \begin{pmatrix} X \\ Y \end{pmatrix}$$

[3]Für die Normalverteilung gilt mit $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, dass $X = (X_1, \ldots, X_n)^T$ wieder normalverteilt ist, mit $(\mathbb{E}[X])_i = \mu_i$ und $(D^2[X])_{ij} = \mathrm{Cov}(X_i, X_j)$.

Aufgaben:

Aufgabe 1A.
Gegeben seien unabhängige und identisch verteilte Zufallsvariablen $X_1, X_2, \dots$, wobei X_1 exponentialverteilt mit Parameter $\lambda = 1$ sei, d.h. Lebesguedichte $f(x) = \mathbb{1}_{(0,\infty)} \exp(-x)$ besitzt. Zeigen Sie, dass $Y_n = \max\{X_1, \dots, X_n\} - \ln(n)$ in Verteilung konvergiert und bestimmen Sie die Grenzverteilung.

Aufgabe 1B.
Es sei $a \in \mathbb{R}$ und $(X_n)_{n\in\mathbb{N}}$ eine Folge von quadratintegrierbaren Zufallsgrößen. Zeigen Sie:

a) Gilt $\mathbb{E}[X_n] \xrightarrow{n\to\infty} a$ und $\mathbb{V}\mathrm{ar}[X_n] \xrightarrow{n\to\infty} 0$, so folgt $X_n \xrightarrow{\mathbb{P}} a$.

a) Gilt $\mathbb{E}[X_n] \xrightarrow{n\to\infty} a$ und $\sum_{n=1}^{\infty} \mathbb{V}\mathrm{ar}[X_n] < \infty$, so folgt $X_n \xrightarrow{f.s.} a$.

Aufgabe 1C.
Es sei $(Y_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch exponentialverteilter Zufallsgrößen mit Parameter $\lambda > 0$. Zeigen Sie:

a) Für $X_n = Y_n \,/\, \ln n$ gilt $X_n \xrightarrow{\mathbb{P}} 0$ aber nicht $X_n \xrightarrow{f.s.} 0$.

b) Für $X_n = Y_n \,/\, (\ln n)^2$ gilt $X_n \xrightarrow{\mathbb{P}} 0$ und $X_n \xrightarrow{f.s.} 0$.

Aufgabe 1D.
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und X eine weitere Zufallsvariable.

a) Falls $X_n \xrightarrow{f.s.} X$, so folgt $X_n \xrightarrow{\mathbb{P}} X$.

b) Es gelte $X_n \xrightarrow{d} c$, wobei c eine deterministische Konstante ist. Dann folgt $X_n \xrightarrow{\mathbb{P}} c$.

Aufgabe 1E.
Beweisen Sie folgende Verallgemeinerungen der Tschebyscheffschen Ungleichungen:

a) Unter der Voraussetzungen, dass $\mathbb{E}[e^{aX}]$ für $a > 0$ existiert, ist

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[e^{aX}]}{e^{a\varepsilon}}, \qquad (\varepsilon > 0).$$

b) Für die positiv und nichtfallende Funktion f existiere der Erwartungswert $\mathbb{E}[f(|X - \mathbb{E}X|)]$. Dann gilt

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathbb{E}[f(|X - \mathbb{E}X|)]}{f(\varepsilon)}$$

Aufgabe 1F.
Es sei (Y_n) eine Folge unabhängiger auf $(0,1)$ gleichverteilter Zufallsgrößen. Zeigen Sie, dass

$$X_n := \left(\frac{1}{Y_1} \cdots \frac{1}{Y_n}\right)^{\frac{1}{n}} \xrightarrow{f.s.} e$$

gilt.

Aufgabe 1G.
Zeigen Sie, dass der Grenzwert in Wahrscheinlichkeit und der fast sichere Grenzwert linear sind.

Aufgabe 1H.
Beweisen Sie den Satz von Caesaro-Konvergenz, d.h. aus der Konvergenz $x_n \to x \in \mathbb{R}$ einer Zahlenfolge (x_n) folgt $\frac{1}{n}\sum_{i=1}^{n} x_n \to x$ $(n \to \infty)$.

Aufgabe 1I.
Berechnen Sie die charakteristischen Funktionen der folgenden Zufallsgrößen:

a) X sie auf dem Intervall $[-a, a]$ für ein $a > 0$ gleichverteilt,

b) Y sei zweipunktverteilt gegeben durch $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 0.5$,

c) $Z \sim \mathrm{Bin}(n, p)$, d.h. Z sei eine binomialverteilte Zufallsgröße mit Parametern $n \in \mathbb{N}$ und $p \in (0, 1)$.

Aufgabe 1J.
Zeigen Sie mit Hilfe des zentralen Grenzwertsatzes durch die Wahl geeigneter Zufallsgrößen X_n in Satz 1.19, dass

$$\lim_{n\to\infty} e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!} = 0.5.$$

2 Stichproben und Stichprobenfunktion

Definition 2.1 (Stichprobe). Es sei X eine Zufallsgröße auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit der Verteilungsfunktion F. Ein zufälliger Vektor $(X_1, \ldots, X_n)^T$ heißt (mathematische) Stichprobe, bzw. unabhängige Beobachtung von X, falls

1. $F_{X_1} = \cdots = F_{X_n} = F$,
2. $X_1, \ldots, X_n$ sind unabhängig,

d.h. $X_1, \ldots, X_n$ sind unabhängig und identisch verteilt wie X.
Eine Realisierung $(X_1(\omega), \ldots, X_n(\omega))^T = (x_1, \ldots, x_n)^T$ für ein $\omega \in \Omega$, heißt konkrete Stichprobe / konkrete Beobachtung von X.

Definition 2.2 (Stichprobenfunktion). Eine Borel-messbare Abbildung $\phi : \mathbb{R}^n \to \mathbb{R}^m$ heißt Stichprobenfunktion, insbesondere heißt eine Abbildung $\varphi(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^m$ mit $\varphi(X_1, \ldots, X_n)(\omega) = \varphi(X_1(\omega), \ldots, X_n($
Statistik oder auch zufällige Stichprobenfunktion.

Definition 2.3 (Stichprobenmittel). Sei $(X_1, \ldots, X_n)$ eine mathematische Stichprobe, dann heißt die Zufallsvariable

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Stichprobenmittel bzw. empirischer Erwartungswert. Demnach ist für eine konkrete Stichprobe $(x_1, \ldots, x_n)$ der Wert $\overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$ das konkrete Stichprobenmittel.

Satz 2.4. Für das Stichprobenmittel gilt:

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}[X], \qquad D^2[\overline{X}_n] = \frac{D^2[X]}{n}$$

Beweis. Beide Gleichungen folgen aus der Linearität von Erwartungswert und Varianz (wegen Unabhängigkeit):

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \cdot n \cdot \mathbb{E}[X] = E[X]$$

$$D^2[\overline{X}_n] = D^2\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} D^2[X_i] = \frac{1}{n^2} \cdot n \cdot D^2[X] = \frac{D^2[X]}{n}$$

□

Der empirische Erwartungswert nähert sich dem echten Erwartungswert an, d.h.

$$\mathbb{E}|\overline{X}_n - \mathbb{E}[X]| = \mathbb{E}|\overline{X}_n - \mathbb{E}[\overline{X}_n]| = D^2[\overline{X}_n] = \frac{D^2[X]}{n} \xrightarrow{n \to \infty} 0$$

Tatsächlich nähert sich $\overline{X}_n$ dem Erwartungswert sogar fast sicher an:

Satz 2.5. Der empirische Erwartungswert konvergiert fast sicher gegen den Erwartungswert, d.h.

$$\mathbb{P}\left(\lim_{n \to \infty} \overline{X}_n = \mathbb{E}[X]\right) = 1$$

Außerdem gilt

$$\lim_{n \to \infty} \mathbb{P}\left(\sqrt{n} \cdot \frac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{D^2[X]}} \leq x\right) = \Phi(x)$$

für jedes $x \in \mathbb{R}$. (Dabei ist Φ die Verteilungsfunktion der Standardnormalverteilung)

Beweis.
Der erste Teil des Satzes folgt direkt aus dem Gesetz der großen Zahlen.

Für den zweiten Teil des Satzes verwenden wir den zentralen Grenzwertsatz, denn es gilt $S_n = n \cdot \overline{X}_n$ und damit

$$\sqrt{n} \cdot \frac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{D^2[X]}} = \sqrt{n} \cdot \frac{\frac{1}{n}(S_n - \mathbb{E}[S_n])}{\sqrt{D^2[X]}} = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{D^2[S_n]}}$$

Es folgt

$$\mathbb{P}\left(\sqrt{n} \cdot \frac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{D^2[X]}} \leq x\right) = \mathbb{P}\left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{D^2[S_n]}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x)$$

□

Der zweite Teil des Satzes liefert eine Auskunft über die Abweichung von $\overline{X}_n$ gegenüber $\mathbb{E}[X]$:

$$\mathbb{P}(|\overline{X}_n - \mathbb{E}[X]| > \varepsilon) \approx 2 - 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{D^2[X]}}\right)$$

Beispiel 2.6 (Schätzen von unbekanntem Parameter).

(a) Eine Versicherung weiß, dass die Anzahl der Schäden X pro Tag poissonverteilt mit Parameter $\lambda = \mathbb{E}[X]$ ist. Zwar ist λ unbekannt, jedoch kann dieser anhand von Beobachtungen $X_1, \ldots, X_n$ geschätzt werden:

$$\lambda \approx \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{für } n \text{ groß genug.}$$

(b) Die Körpergröße X einer zufällig ausgewählten Person in Rostock ist normalverteilt und hat damit die Verteilungsfunktion

$$\Phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \mathrm{d}t$$

Um eine vollständige Charakterisierung zu tätigen, benötigen wir noch Schätzwerte für die Parameter μ und σ^2 durch

$$\mu = \mathbb{E}[X] \approx \frac{1}{n}\sum_{i=1}^{n} X_i$$

(c) Ein Supermarkt weiß, dass die Zeit X bis zur Ankunft eines neuen Kunden an der Supermarktkasse exponentialverteilt mit Parameter λ ist. Die Wartezeiten werden n mal beobachtet, so dass

$$\lambda = \frac{1}{\mathbb{E}[X]} \approx \frac{n}{\sum_{i=1}^{n} X_i}$$

Für die Charakterisierung einiger Verteilungen benötigen wir neben dem Erwartungswert auch noch eine Schätzung für die Varianz $D^2[X]$.

Wir betrachten hierfür eine weitere Stichprobenfunktion:

Definition 2.7 (Stichprobenvarianz). Sei $X_1, \ldots, X_n$ eine mathematische Stichprobe, dann heißt die Zufallsgröße

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Stichprobenvarianz bzw. empirische Varianz. Für eine konkrete Stichprobe $x_1, \ldots, x_n$ ist demnach $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ die konkrete Stichprobenvarianz.

Satz 2.8. Für die Stichprobenvarianz gilt

$$\mathbb{E}[S_n^2] = D^2[X]$$

Existiert zusätzlich das 4. Moment von X, d.h. $\mathbb{E}X^4 < \infty$, so ergibt sich

$$D^2[S_n^2] = \frac{1}{n}\left(\mathbb{E}(X - \mathbb{E}[X])^4 - \frac{n-3}{n-1}(D^2[X])\right)$$

Beweis.
Der erste Teil des Satzes ergibt mit dem Ansatz

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left((X_i - \mathbb{E}[X]) - \frac{1}{n}\sum_{j=1}^{n}(X_j - \mathbb{E}[X])\right)^2$$

Da der Erwartungswert von $X_i - \mathbb{E}[X]$ durch 0 gegeben ist, lässt sich ab nun o.B.d.A. annehmen, dass $\mathbb{E}[X] = 0$ gilt. Es ergibt sich die Umformung zu

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\cdot\overline{X}_n^2\right)$$

und damit für den Erwartungswert

$$\mathbb{E}[S_n^2] = \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}[X_i^2] - n\cdot\mathbb{E}[\overline{X}_n^2]\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}D^2[X] - D^2[X]\right) = D^2[X]$$

Der zweite Teil des Satzes sei ausgelassen. □

Der Satz garantiert die Konvergenz im quadratischen Mittel. Auch hier lässt sicher zusätzlich noch fast sichere Konvergenz folgern:

Satz 2.9. Die empirische Varianz konvergiert fast sicher gegen die Varianz, d.h.

$$\mathbb{P}\left(\lim_{n\to\infty} S_n^2 = D^2[X]\right) = 1$$

Falls $\mathbb{E}[X^4] < \infty$ so gilt weiter

$$\lim_{n\to\infty}\mathbb{P}\left(\sqrt{n}\cdot\frac{S_n^2 - D^2[X]}{\sqrt{\mathbb{E}(X - \mathbb{E}[X])^4 - (D^2[X])^2}} \leq x\right) = \Phi(x)$$

Beweis. Aus der Unabhängigkeit von $X_1, \dots, X_n$ folgt auch die von $X_1^2, \dots, X_n^2$ und somit liefert das Gesetz der großen Zahlen, dass $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{f.s.} \mathbb{E}[X^2]$ und nach aus Satz 2.5 folgt zusätzlich $\overline{X}_n^2 \xrightarrow{f.s.} (\mathbb{E}[X])^2$.

Analog zum Beweis von Satz 2.8 ergibt sich damit

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\cdot\overline{X}_n^2\right) = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}_n^2\right) \xrightarrow{n\to\infty} 1\cdot(\mathbb{E}[X^2] - (E[X])^2) = D^2[X]$$

Der zweite Teil sei erneut ohne Beweis. □

Beispiel 2.10 (Weiterführung von Beispiel 2.6). Wir sind nun in der Lage für eine Normalverteilung den zweiten Parameter σ^2 zu schätzen:

$$\sigma^2 = D^2[X] \approx \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Satz 2.11. Für den empirischen Erwartungswert und Varianz ergibt sich folgende verwandte Form des zentralen Grenzwertsatzes:

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{n}\cdot\frac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{S_n^2}} \le x\right) = \Phi(x)$$

Beweis. Die Konvergenz folgt direkt aus dem zentralen Grenzwertsatz, denn

$$\begin{aligned}\sqrt{n}\cdot\frac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{S_n^2}} &= \sqrt{n}\cdot\frac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{D^2[X]}}\cdot\frac{\sqrt{D^2[X]}}{\sqrt{S_n^2}} = \sqrt{n}\cdot\frac{\frac{1}{n}S_n - \frac{1}{n}\mathbb{E}[S_n]}{\frac{1}{n}\sqrt{D^2[S_n]}}\cdot\sqrt{\frac{D^2[X]}{S_n^2}}\\ &= \frac{S_n - \mathbb{E}[S_n]}{\sqrt{D^2[S_n]}}\cdot\underbrace{\sqrt{\frac{D^2[X]}{S_n^2}}}_{\xrightarrow{d}1} \xrightarrow{d} Z\sim\mathcal{N}(0,1)\end{aligned}$$

Bemerkung 2.12. Dieser Satz ermöglicht es sogenannte Konfidenzintervalle (später mehr) anzugeben, d.h. einen Bereich für welchen die Wahrscheinlichkeit, dass $|\overline{X}_n - \mathbb{E}[X]|$ darin liegt, einen festen Wert annimmt. Sei hierfür $z_\alpha \in (0,1)$ das α-Quantil der Standardnormalverteilung, d.h. $\Phi(z_\alpha) = \alpha$, dann gilt

$$\mathbb{P}\left(|\overline{X}_n - \mathbb{E}[X]| < \frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\right) \approx 1-\alpha$$

Satz 2.13. Für eine mathematische Stichprobe $X_1,\dots,X_n$ von $X\sim\mathcal{N}(\mu,\sigma^2)$ sind $\overline{X}_n$ und S_n^2 unabhängige Zufallsgrößen und es gilt

$$\overline{X}_n \sim \mathcal{N}(\mu, \tfrac{\sigma^2}{n}),\qquad \frac{(n-1)S_n^2}{\sigma^2}\sim\chi^2_{n-1},\qquad \frac{\sqrt{n}(\overline{X}_n-\mu)}{S_n}\sim t_{n-1}$$

Beweis. Ohne Beweis.

Bemerkung 2.14.

1. χ^2_n ist die Chi-Quadrat Verteilung, sie ist die Summe von n unabhängigen quadratischen Standardnormalverteilung, d.h.

$$X_1,\dots,X_n\sim\mathcal{N}(0,1) \implies \sum_{i=1}^n X_i^2 \sim \chi^2_n$$

   Sie hat folgende Dichtefunktion:

$$f_{\chi^2_n}(x) = \frac{x^{k/2-1}e^{-k/2}}{2^{k/2}\Gamma(\frac{k}{2})}\mathbb{1}_{(0,\infty)}$$

2. t_n ist die studentsche t-Verteilung. Für Zufallsgrößen $X\sim\mathcal{N}(0,1)$ und $U_n\sim\chi^2_n$ gilt

$$\frac{X}{\sqrt{\frac{U_n}{n}}}\sim t_n$$

   Die Dichte ist gegeben durch

$$f_{t_n} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Beispiel 2.15 (Schätzer einer Einzelwahrscheinlichkeit). Gegeben seien (X_n) unabhängig und Bernoulliverteilt mit Parameter $p = \mathbb{P}(A)$ für ein Ereignis A, d.h.

$$\mathbb{P}(X_n = 1) = p, \quad \mathbb{P}(X_n = 0) = 1 - p$$

Für Ergebnisse $\omega = (\omega_1, \omega_2, \dots)$ gilt

$$X_n(\omega) = 1_{\{}A\}(\omega_n) = \begin{cases} 1 & \omega_n \in A \\ 0 & \omega_n \notin A \end{cases}$$

dabei bedeutet $\omega_n \in A$, dass das Ereignis A in der n-ten Wiederholung des Experiments eintrifft.

Die absolute Häufigkeit $H_n(A) = \sum_{k=1}^n X_k$ beschreibt wie oft das Ereignis A in den ersten n Versuchen eintrifft, die relative Häufigkeit ist gegeben durch $h_n(A) = \frac{H_n(A)}{n}$ und es gilt

$$h_n(A) = \frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{f.s.} \mathbb{E}[X] = \mathbb{P}(A)$$

Beispiel 2.16. 5 Mal wird der zufällige Versuch Würfeln n-fach wiederholt. Sei zudem $A = \{1, 2\}$. Wir würfeln am Computer für verschiedene n:

n	$h_n(A)$				
6	$\frac{3}{6}$	$\frac{5}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{2}{6}$
$6 \cdot 10^3$	0.34567	0.33017	0.32800	0.33033	0.33067
$6 \cdot 10^6$	0.33325	0.33311	0.33346	0.33330	0.33319

Ein intuitiver Schätzer der Varianz ergibt sich aus

$$\tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_j)^2$$

Auch für diesen Schätzer gilt $\tilde{S}_n^2 \xrightarrow{f.s.} D^2[X]$, da $\tilde{S}_n^2 = \frac{n-1}{n} S_b^2$. Es gilt jedoch

$$\mathbb{E}[\tilde{S}_n^2] = \frac{n-1}{n}\mathbb{E}[S_n^2] = \frac{n-1}{n} D^2[X] < D^2[X]$$

Für den Fall das der echte Wert von $\mathbb{E}[X]$ bekannt ist, funktioniert dieser Ansatz, denn für

$$\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X])^2$$

ergibt sich

$$\begin{aligned}
\mathbb{E}[\hat{S}_n^2] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i - \mathbb{E}[X])^2 = \frac{1}{n}\left(\sum_{i=1}^n \mathbb{E}[X_i^2] - 2\mathbb{E}[X]\sum_{i=1}^n \mathbb{E}[X_i] + n(\mathbb{E}[X])^2\right) \\
&= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 = D^2[X] \\
\hat{S}_n^2 &= \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\mathbb{E}[X] \cdot \frac{1}{n}\sum_{i=1}^n X_i + (\mathbb{E}[X])^2 \xrightarrow{f.s.} D^2[X]
\end{aligned}$$

Definition 2.17 (Momentschätzer). Sei $X_1, \dots, X_n$ eine Stichprobe einer Zufallsgröße X mit $\mathbb{E}|X|^k < \infty$.

1. Das k-te Moment $\mathbb{E}[X^k]$ kann durch

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

geschätzt werden. Dieser Schätzer ist erwartungstreu, d.h. $\mathbb{E}[\hat{m}_k] = \mathbb{E}[X^k]$, und konvergiert fast sicher gegen das k-te Moment.

2. Das k-te zentrierte Moment $\mathbb{E}(X - \mathbb{E}[X])^k$ kann durch

$$\hat{\mu}_k = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^k$$

geschätzt werden. Auch dieser Schätzer ist erwartungstreu und konvergiert fast sicher.

Sei $(X_1, Y_1), \ldots, (X_n, Y_n)$ eine Stichprobe eines Zufallsvektors (X, Y) mit $D^2[X], D^2[Y] < \infty$, so ist

$$\hat{C}_{X,Y} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$$

die empirische Kovarianz, ein Schätzer für $\operatorname{cov}(X, Y)$. Der Korrelationskoeffizient wird demnach durch den empirischen Korrelationskoeffizient

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2}}$$

der Korrelationskoeffizient geschätzt.

Beispiel 2.18 (Empirischer Korrelationskoeffizient). Nun steht X für die Note eines zufällig ausgewählten Schülers in Mathematik und Y für dessen Note in Physik. Lehrer fragen sich, wie groß der Zusammenhang zwischen X und Y ist und nehmen eine konkrete Stichprobe $(x_1, y_1), (x_2, y_2), \ldots, (x_{10}, y_{10})$:

i	1	2	3	4	5	6	7	8	9	10
x_i	1	3	2	1	5	2	3	4	4	3
y_i	1	3	1	3	4	2	3	3	5	3

Das arithmetische Mittel beträgt $\overline{x}_n = \overline{y}_n = 2.8$.

Der empirische Korrelationskoeffizient berechnet sich zu

$$\hat{\rho}_{XY} = \frac{(1-2.8)(1-2.8) + (3-2.8)(3-2.8) + (2-2.8)(1-2.8) + \ldots}{\sqrt{(1-2.8)^2 + (3-2.8)^2 + \ldots}\sqrt{(1-2.8)^2 + (3-2.8)^2 + \ldots}} = 0.73$$

Da die Schätzung nahe 1 liegt, scheint ein großer Zusammenhang zwischen Mathe- und Physiknoten zu bestehen.

Bemerkung 2.19 (Schätzer, als Integralnäherung). Sei $g : [a, b] \to \mathbb{R}$ eine stetige Funktion, gesucht ist der Wert des Integrals $\int_a^b g(x)\,\mathrm{d}x$.

Durch die Verwendung einer auf $[a, b]$ gleichverteilten Zufallsgröße X ergibt sich

$$\int_a^b g(x)\,\mathrm{d}x = (b-a)\cdot\int_a^b f_X(x)g(x)\,\mathrm{d}x = (b-a)\mathbb{E}[g(X)]$$

Sei nun $X_1, \ldots, X_n$ eine Stichprobe von X, so gilt

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{f.s.} \mathbb{E}[g(X)], \quad \text{d.h.} \quad \int_a^b g(x)\,\mathrm{d}x \approx (b-a)\cdot\frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

Solche auf $[a, b]$ generierten gleichverteilten Zufallszahlen (z.B. am Rechner) ermöglichen demnach eine (Monte-Carlo) Approximation des Integrals.

Definition 2.20 (Ordnungsstatistik). Sei $X_1, \dots, X_n$ eine Stichprobe und

$$\phi(x_1, \dots, x_n) = (x_{(1)}, \dots, x_{(n)}) \qquad \text{mit } x_{(i)} = \min\left\{x_j \,:\, \#\{k : x_k \le x_j\} \ge i\right\}$$

die Permutation, welche $x_1, \dots, x_n$ der Größe nach ordnen, d.h. $x_{(1)} \le \dots \le x_{(n)}$.
Die Zufallsvariablen $X_{(1)}, \dots, X_{(n)}$, gegeben durch

$$(X_{(1)}(\omega), \dots, X_{(n)}(\omega)) = \phi(X_1(\omega), \dots, X_n(\omega)), \quad \forall \omega \in \Omega$$

Definition 2.21. Für eine Ordnungsstatistik $X_{(1)}, \dots, X_{(n)}$ sei $R_n := X_{(n)} - X_{(1)}$ die Stichprobenspannweite und

$$M_n := \begin{cases} X_{(\frac{n+1}{2})}, & n \text{ ungerade} \\ \frac{1}{2} X_{(\frac{n}{2})} + \frac{1}{2} X_{(\frac{n}{2}+1)}, & n \text{ gerade} \end{cases}$$

der Stichprobenmedian.

Der Stichprobenmedian dient als alternative zum klassischen Mittelwert. Etwa die Hälfte der Stichprobe liegt unter M_n und hat den Vorteil, dass extremale Werte bei $X_{(1)}$ oder $X_{(n)}$ ihn weniger verzerren, als es bei $\overline{X}_n$ passieren würde.

Satz 2.22. Sei $X_1, \dots, X_n$ eine Stichprobe zu einer diskreten Zufallsgröße X mit aufsteigendem Wertebereich $x_0 < \dots < x_m$ für $m \in \mathbb{N} \cup \{\infty\}$, so gilt

$$\mathbb{P}(X_{(i)} \le x_j) = \sum_{k=i}^{n} \binom{n}{n} P_j^k \cdot (1 - P_j)^{n-k}$$

wobei $P_j = \mathbb{P}(X \le x_j) = \sum_{k=0}^{j} \mathbb{P}(X = x_k)$.

Beweis. Die Ereignisse $\{X \le x_j\}$ und $\{X > x_j\}$ beschreiben Erfolg und Misserfolg eines bernoulliverteiltem Ereignis mit Erfolgswahrscheinlichkeit P_j. Damit $X_{(i)} \le x_j$ muss dieses Ereignis mindestens i Mal eintreffen. Dabei ist das Eintreffen von k-fachem Erfolg binomialverteilt.

Satz 2.23. Sei $X_1, \dots, X_n$ eine Stichprobe zu einer Zufallsgröße mit Dichtefunktion f und zugehöriger Verteilungsfunktion $F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$, dann ist die Dichte der Ordnungsstatistik $X_{(i)}$ gegeben durch

$$f_{X_{(i)}}(x) = \binom{n}{i} \cdot i \cdot f(x) \cdot F(x)^{i-1} \cdot F(x)^{n-i}$$

ohne Beweis.

Beispiel 2.24. Seien die Stichprobenvariablen $X_1, \dots, X_n$ gleichverteilt auf dem Intervall $[0, b]$. Dann haben diese die folgende Dichte- und Verteilungsfunktion

$$f(x) = \frac{1}{b} \cdot \mathbb{1}_{[0,b]}(x), \qquad F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{b}, & x \in [0, b] \\ 1, & x > 1 \end{cases}$$

Für die Ordnungsstatistik ergibt sich:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} \cdot \frac{x^{i-1}}{b^n} \cdot (b - x)^{n-1} \cdot \mathbb{1}_{[0,b]}(x)$$

und es folgt

$$\mathbb{E}[X_{(i)}] = \int_0^b \frac{n!}{(i-1)!(n-i)!} \cdot \frac{x^i}{b^n} \cdot (b-x)^{n-1}\,\mathrm{d}x = \frac{ib}{n+1}$$
$$D^2[X_{(i)}] = \dots = \frac{i(n-i+1)b^2}{(n+1)^2(n+2)}$$

Definition 2.25. Sei $X_1, \dots, X_n$ eine Stichprobe der Zufallsgröße X, dann ist

$$W_n(x) = \frac{1}{n}|\{X_i \,:\, X_i \leq x\}|$$

die empirische Verteilungsfunktion von X.

Diese Definition gibt die relative Häufigkeit des Ereignis $\{X \leq x\}$ an und liefert damit

$$W_n(x) \approx \mathbb{P}(\{X \leq x\}) = F_X(x)$$

Satz 2.26 (Hauptsatz der Statistik). Sei W_n die empirische Verteilung einer Stichprobe $X_1, \dots, X_n$ der Zufallsgröße X mit Verteilungsfunktion F, dann gilt für alle $x \in \mathbb{R}$:

1. Die Zufallsgröße $nW_n(x)$ ist binomialverteilt mit Parametern n und $p = F(x)$, d.h.

$$\mathbb{P}(nW_n(x) = k) = \binom{n}{k} \cdot F(x)^k \cdot (1 - F(x))^{n-k} \quad \text{für alle } k = 0, 1, \dots, n$$

2. $\mathbb{E}[W_n(x)] = F(x)$
3. $D^2[W_n(x)] = \frac{1}{n}F(x)(1 - F(x))$
4. $W_n(x) \xrightarrow{f.s.} F(x)$
5. Falls $0 < F(x) < 1$, so gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{n}\frac{W_n(x) - F(x)}{\sqrt{F(x)(1 - F(x))}} \leq y\right) = \Phi(y)$$

Beweis.

1. Für $p = F(x)$ sei $Y_i(x)$ definiert als

$$Y_i(x) = \begin{cases} 1, & X_i \leq x, \text{ mit Wkt. } p \\ 0, & X_i > x, \text{ mit Wkt. } 1 - p \end{cases}$$

dann sind $Y_1(x), \dots, Y_n(x)$ unabhängig und Bernoulliverteilt mit

$$n \cdot W_n(x) = |\{X_i \leq x\}| = \sum_{i=1}^n Y_i(x) \sim \text{Bin}(n, p)$$

2. Mit $Y_i(x)$ aus Teil 1 folgt

$$\mathbb{E}[W_n(x)] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i(x)] = \frac{1}{n} \cdot np = p = F(x)$$

3. Analog zu 2. folgt

$$D^2[W_n(x)] = \frac{1}{n^2}\sum_{i=1}^{n} D^2[Y_i(x)] = \frac{1}{n^2}\cdot np(1-p) = \frac{1}{n}F(x)(1-F(x))$$

4. Nach dem Gesetz der großen Zahlen gilt

$$W_n(X) = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{f.s.} \mathbb{E}[Y_1(x)] = p = F(x)$$

5. Nach dem zentralen Grenzwertsatz folgt mit $W_n(x) = \frac{1}{n}S_n(x)$, dass

$$\begin{aligned}\sqrt{n}\frac{W_n(x)-F(x)}{\sqrt{F(x)(1-F(x))}} &= \sqrt{n}\frac{W_n(x)-\mathbb{E}[W_n(x)]}{\sqrt{n\cdot D^2[W_n(x)]}} = \sqrt{n}\frac{\frac{1}{n}S_n(x)-\frac{1}{n}\mathbb{E}[S_n(x)]}{\sqrt{\frac{1}{n}D^2[S_n(x)]}}\\ &= \frac{S_n(x)-\mathbb{E}[S_n(x)]}{\sqrt{D^2[S_n(x)]}} \xrightarrow{d} Z\sim\mathcal{N}(0,1)\end{aligned}$$

Die zweite und dritte Aussage des Satzes garantiert uns, dass $W_n(x)$ ein geeigneter Schätzer ist (d.h. erwartungstreu und Quadratmittelfehler geht gegen Null). Die fünfte Aussage erlaubt uns erneut die Bestimmung eines Konfidenzintervalls:

$$\mathbb{P}(|W_n(x)-F(x)| > \varepsilon) \approx 2 - 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{F(x)(1-F(x))}}\right)$$

Satz 2.27 (Satz von Gliwenko-Cantelli). Die Folge (W_n) der empirischen Verteilungsfunktionen konvergiert fast sicher gleichmäßig in x, d.h.

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}}|W_n(x)-F(x)| \xrightarrow{n\to\infty} 0\right) = 1$$

ohne Beweis.

Beispiel 2.28 (Empirische Verteilungsfunktion). Eine Maschine produziert Teile, die eine gewisse Länge haben sollen. Jedoch weicht die Länge bei der Produktion immer leicht ab. Dieser Fehler ist eine Zufallsgröße X (Abweichung in mm). Konkrete Messungen $x_1, x_2, \dots, x_{50}$ der Abweichungen sehen wie folgt aus:

0.46	0.47	2.46	-0.32	-0.07
0.06	**-2.52**	-0.53	-0.19	0.54
1.49	-0.35	-0.63	0.70	0.93
1.02	-0.47	1.28	**3.56**	0.57
1.39	-0.65	0.05	0.32	2.95
0.30	-0.29	1.30	0.24	-0.96
-1.56	0.19	-1.29	0.02	0.53
1.38	0.79	-0.96	-0.85	-1.87
-1.58	0.19	1.19	-0.50	-0.27
1.97	-0.26	0.41	0.44	-0.04

Wir interpretieren X als stetige Zufallsgröße und unterteilen die Daten in Klassen. Die empirische Verteilungsfunktion $W_n(x)$ gibt für jedes x die relative Häufigkeit der Werte $x_i \leq x$ an und schätzt damit die Verteilungsfunktion $F_X(x)$.

Als Faustregel für die Anzahl der Klassen K gilt:

$$K \leq \log_{10}(n) \cdot 5$$

mit $n = 50$ die Klassenanzahl ergibt sich $K = 8$. Die Klassen sind:

$$[a_0, a_1), [a_1, a_2), \ldots, [a_7, a_8]$$

mit $a_0 = x_{\min} = -2.52$, $a_8 = x_{\max} = 3.56$, sowie

$$a_i = a_0 + i \cdot d, \qquad \text{mit } d = \frac{x_{\max} - x_{\min}}{K} = \frac{6.08}{8} = 0.76$$

Für jede Klasse $[a_i, a_{i+1})$ bestimmen wir die relative Häufigkeit

$$h_n^{(i)} = \frac{1}{n} \cdot \text{Anzahl der } x_j \text{ mit } x_j \in [a_i, a_{i+1})$$

und schätzen damit die Dichte in diesem Intervall durch

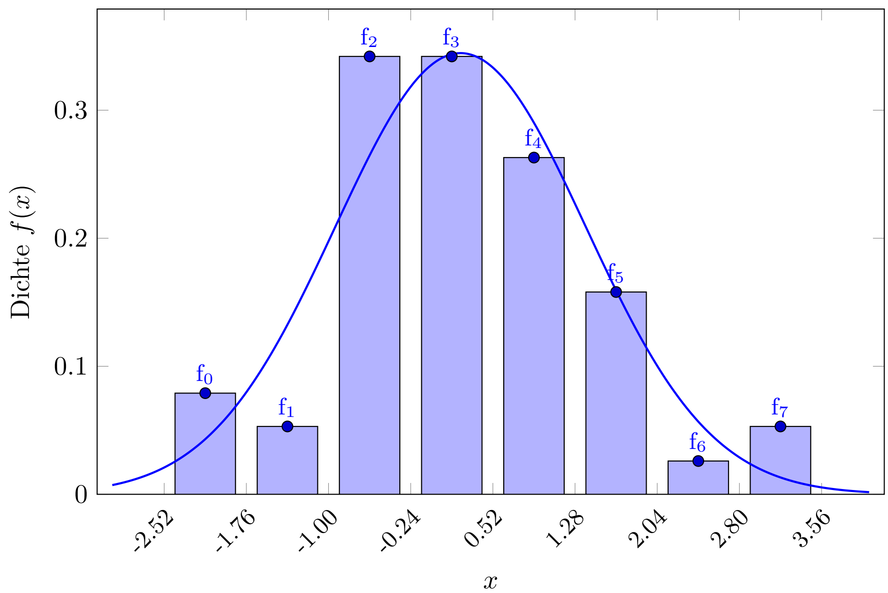
$$\mathbb{P}(X \in [a_i, a_{i+1})) = \int_{a_i}^{a_{i+1}} f_X(x)\,\mathrm{d}x = f_i \cdot d$$

Das heißt $f_i = \frac{h_n^{(i)}}{d}$ ist eine Approximation der Dichte, genannt empirische Dichte.

Unsere Werte ergeben:

i	Klasse	Klassenhäufigkeit	rel. Klassenhäufigkeit $h_n^{(i)}$	$f_i = \frac{h_n^{(i)}}{d}$
0	$[-2.52,\ -1.76)$	3	0.06	≈ 0.079
1	$[-1.76,\ -1.00)$	2	0.04	≈ 0.053
2	$[-1.00,\ -0.24)$	13	0.26	≈ 0.342
3	$[-0.24,\ +0.52)$	13	0.26	≈ 0.342
4	$[+0.52,\ +1.28)$	10	0.20	≈ 0.263
5	$[+1.28,\ +2.04)$	6	0.12	≈ 0.158
6	$[+2.04,\ +2.80)$	1	0.02	≈ 0.026
7	$[+2.80,\ +3.56]$	2	0.04	≈ 0.053

Die empirische Dichte lässt sich wie folgt skizzieren:



Der Graph scheint grob mit der Dichte Normalverteilung übereinzustimmen, wenn diese die Parameter $\mu = \overline{x}_n$ und $\sigma^2 = s_n^2$ hat.

Aufgaben:

Aufgabe 2A.
Als Lagemaß wird eine Abbildung $\ell : \mathbb{R}^n \to \mathbb{R}$ bezeichnet, welche für jedes $c \in \mathbb{R}$ die Bedingung

$$\ell(x_1 + c, \dots, x_n + c) \; = \; \ell(x_1, \dots, x_n) + c$$

erfüllt. Als Streuungsmaß hingegen bezeichnen wir eine Abbildung $s : \mathbb{R}^n \to \mathbb{R}$ mit

$$s(x_1 + c, \dots, x_n + c) \; = \; s(x_1, \dots, x_n)$$

für jedes $c \in \mathbb{R}$. Entscheiden Sie jeweils, ob es sich bei dem Stichprobenmittel und der Stichprobenvarianz um ein Lagemaß oder ein Streuungsmaß handelt.

Aufgabe 2B.
Seien $X_1, \dots, X_n$ quadratintegrierbare, unabhängige und identisch verteilte Zufallsvariablen mit Erwartungswert μ und Varianz ν. Wir betrachten für reelle Konstanten $\alpha_1, \dots, \alpha_n$ die Stichprobenfunktion

$$\varphi : \mathbb{R}^n \to \mathbb{R}, \qquad \varphi(x_1, \dots, x_n) = \sum_{j=1}^{n} \alpha_j x_j.$$

Bestimmen Sie die Koeffizienten $\alpha_1, \dots, \alpha_n$ derart, dass $\mathrm{E}[\varphi(X_1, \dots, X_n)] = \mu$ gilt und $\mathrm{D}^2[\varphi(X_1, \dots, X_n)]$ minimal wird.

Aufgabe 2C.

1. Für die Zufallsvariable $Y : \Omega \to \mathbb{R}$ gelte $Y \sim \Gamma(b, p)$ mit den Parametern $b, p > 0$, das heißt Y besitze die Dichte
$$f_Y(y) = \frac{b^p}{\Gamma(p)} e^{-by} y^{p-1} \mathbf{1}_{(0,\infty)}(y).$$
Bestimmen Sie die charakteristische Funktion von Y.

2. Seien Y_1, Y_2 unabhängige Zufallsvariablen mit $Y_1 \sim \Gamma(b, p_1)$ und $Y_2 \sim \Gamma(b, p_2)$. Zeigen Sie, dass dann $Y_1 + Y_2 \sim \Gamma(b, p_1 + p_2)$ gilt.

Aufgabe 2D.
Sei $r \in \mathbb{N}$ und seien $X_1, \dots, X_r$ unabhängige und standardnormalverteilte Zufallsvariablen. Zeigen Sie, dass die Zufallsvariable

$$U_r = \sum_{i=1}^{r} X_i^2$$

chi-quadratisch mit r Freiheitsgraden ist, das heißt $U_r \sim \chi_r^2$ und die Dichte

$$f_{U_r}(x) = \frac{1}{2^{r/2}\, \Gamma(r/2)}\, x^{\frac{r-2}{2}}\, e^{-x/2}\, \mathbf{1}_{(0,\infty)}(x)$$

hat.

Aufgabe 2E.
Führen Sie zur näherungsweisen Berechnung des Integrals

$$\int_0^{\frac{\pi}{2}} \sin x \, dx$$

eine Monte-Carlo-Simulation mit Hilfe von MATLAB durch. Nutzen Sie dabei $n = 10^i$ Samples für $i \in \{1, \dots, 6\}$. Stellen Sie anschließend den absoluten Fehler der numerisch bestimmten Werte von dem korrekten Integralwert in Abhängigkeit von der Anzahl der Samples in einem doppelt-logarithmischen Diagramm dar.

Aufgabe 2F.
Gegeben sei eine Stichprobe $(X_1, \dots, X_n)$ von der mit Parameter $p \in (0,1)$ Bernoulli-verteilten Zufallsgröße X. Wir wollen nun p^2 mit Hilfe der zufälligen Stichprobenfunktion

$$\varphi_1(X_1, \dots, X_n) \;=\; X_1 \cdot X_2$$

schätzen.

1. Bestimmen Sie den Erwartungswert und die Varianz von $\varphi_1(X_1, \dots, X_n)$. Geben Sie zudem die kleinste Konstante c an, sodass für alle $p \in (0,1)$ die Bedingung

   $$\mathrm{D}^2\big(\varphi_1(X_1, \dots, X_n)\big) \;\le\; c$$

   erfüllt ist.
2. Wir definieren eine zweite zufällige Stichprobenfunktion durch

   $$\varphi_2(X_1, \dots, X_n) = \mathrm{E}\big[\varphi_1(X_1, \dots, X_n) \;\big|\; S\big], \quad S = \sum_{j=1}^{n} X_j.$$

   Bestimmen Sie den Erwartungswert von $\varphi_2(X_1, \dots, X_n)$.
3. Zeigen Sie mit Hilfe der Jensen-Ungleichung, dass

   $$\mathrm{D}^2\big[\varphi_1(X_1, \dots, X_n)\big] \;\ge\; \mathrm{D}^2\big[\varphi_2(X_1, \dots, X_n)\big].$$

Aufgabe 2G.

1. Die Stichprobenvariablen $X_1, \dots, X_n$ seien absolut stetig mit der stückweise stetigen Dichte $f\colon \mathbb{R} \to [0,\infty)$ und $X_{(1)}, \dots, X_{(n)}$ die zugehörigen Ordnungsstatistiken. Zeigen Sie, dass die gemeinsame Dichte $f_{X_{(i)},X_{(j)}}\colon \mathbb{R}^2 \to [0,\infty)$ für $1 \le i < j \le n$ durch

   $$f_{X_{(i)},X_{(j)}}(x_i, x_j) = \begin{cases} \dfrac{n!\, f(x_i)\, f(x_j)\, \big(F(x_i)\big)^{i-1} \big(F(x_j) - F(x_i)\big)^{j-1-i} \big(1 - F(x_j)\big)^{n-j}}{(i-1)!\,(j-1-i)!\,(n-j)!}, & -\infty < x_i < x_j < \infty, \\ 0, & \text{sonst} \end{cases}$$

   gegeben ist.
2. Die Stichprobenvariablen $X_1, \dots, X_n$ seien gleichverteilt im Intervall $(0,b)$ für ein $b > 0$. Bestimmen Sie die Dichte der Spannweite $R_n = X_{(n)} - X_{(1)}$.

Aufgabe 2H.
Sei $(X_1, \dots, X_n)$ eine Stichprobe für X, dessen k-tes Moment existiert. Zeigen Sie, dass

$$\hat{\mu}_k = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - X_n)^k \quad \text{für } n \to \infty$$

fast sicher gegen $\mathrm{E}\big[(X - \mathrm{E}[X])^k\big]$ konvergiert.

Aufgabe 2I.
Für zwei Verteilungsfunktionen F und G ist der Kolmogorov-Abstand durch

$$d_K(F,G) \;=\; \sup_{x \in \mathbb{R}} \big|F(x) - G(x)\big|$$

definiert. Zeigen Sie:

1. Der Kolmogorov-Abstand ist eine Metrik auf dem Raum der Verteilungsfunktionen.

2. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit Verteilungsfunktionen (F_n) und X eine Zufallsvariable mit Verteilungsfunktion F. Zeigen Sie, dass aus

$$d_K(F_n, F) \xrightarrow{n\to\infty} 0$$

die Konvergenz in Verteilung von (X_n) gegen X folgt, und dass die Umkehrung im Allgemeinen nicht gilt.

Aufgabe 2J.
Sei $X_1, \ldots, X_n$ eine Stichprobe mit Verteilungsfunktion F. Als Schätzer für F verwenden wir die empirische Verteilungsfunktion W_n. Zeigen Sie,

$$\int_{\mathbb{R}} \frac{\left|W_n(t) - F(t)\right|^2}{1+t^2}\, dt \xrightarrow{P} 0.$$

Aufgabe 2K.
Sie führen eine Studie durch, welche den Zusammenhang zwischen Studiendauer und Einstiegsgehalt von Universitätsabsolventen der Wirtschaftswissenschaften untersucht. Das Ergebnis ist in der folgenden Tabelle angegeben:

Absolvent i	1	2	3	4	5	6	7	8	9	10	11
Studiendauer x_i (Sem.)	7	10	8	10	6	11	7	8	7	11	12
Einkommen y_i (10^3£)	34	58	29	61	40	51	39	64	39	56	59

1. Berechnen Sie das konkrete Stichprobenmittel und die konkrete Stichprobenvarianz der Stichprobe für X (Studiendauer) sowie Y (Einkommen).
2. Berechnen Sie die Kovarianz sowie den Korrelationskoeffizienten zwischen den beiden Merkmalen. Welche Schlussfolgerung ergibt sich daraus?
3. Plotten Sie ein Streudiagramm von (x_i, y_i). Bewerten Sie Ihre Schlussfolgerung aus (ii) vor dem Hintergrund des Diagramms.
4. Markieren Sie im Streudiagramm die Bachelor-Absolventen (i=1,3,5,7,9) und die Master-Absolventen (i=2,4,6,8,10,11) unterschiedlich.
5. Wie ändert sich die Interpretation des Zusammenhangs zwischen Studiendauer und Einstiegsgehalt?

Aufgabe 2L.
Sei $X_1, \ldots, X_n$ eine Stichprobe mit $X_1 \sim \mathrm{Exp}(1/\theta)$, $\theta > 0$. Gegeben seien die Schätzer

$$T_1 = n \cdot \min\{X_1, \ldots, X_n\}, \quad T_2 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. Bestimmen Sie Erwartungswert und Varianz beider Schätzer. Welchen würden Sie bevorzugen und warum?
2. Entscheiden Sie, ob T_1 in Wahrscheinlichkeit gegen θ konvergiert, und begründen Sie Ihre Entscheidung.

3 Parameterschätzer/Punktschätzer

4 Signifikanztests

4.1 Konfidenzschätzungen

5 Lineare Regression

6 Lösungen

Lösung 1A.
Für die Verteilung von Y_n gilt:

$$\begin{aligned} F_{Y_n}(y) &= \mathbb{P}(Y_n \leq y) = \mathbb{P}(\max\{X_1, \dots, X_n\} \leq y + \ln n) = \mathbb{P}(X_i \leq y + \ln n, \quad \forall i = 1, \dots, n) \\ &= \mathbb{P}(X_1 \leq y + \ln n)^n = \mathbb{1}_{(0,\infty)}(y) \cdot (1 - e^{-(y+\ln n)})^n = \mathbb{1}_{(0,\infty)}(y) \cdot \left(1 - \frac{e^{-y}}{n}\right)^n \end{aligned}$$

Diese Verteilung ist überall stetig und für den Grenzwert gilt

$$\lim_{n\to\infty} F_{Y_n}(y) = \mathbb{1}_{(0,\infty)}(y) \cdot \exp(-e^{-y})$$

Lösung 1B.

a) Aus der Tschebyscheffschen Ungleichung folgt für $p = 2$, dass

$$\begin{aligned} \mathbb{P}(|X_n - a| \geq \varepsilon) &\leq \frac{\mathbb{E}(X_n - a)^2}{\varepsilon^2} \\ &= \frac{\mathbb{E}(X_n)^2 - 2a\mathbb{E}X_n + a^2}{\varepsilon^2} \\ &= \frac{(\mathbb{E}X_n)^2 - \mathbb{V}\mathrm{ar}X_n - 2a\mathbb{E}X_n + a^2}{\varepsilon^2} \\ &\to \frac{a^2 - 0 - 2a^2 + a^2}{\varepsilon^2} = 0 \end{aligned}$$

und damit folgt $X_n \xrightarrow{\mathbb{P}} a$.

b) Erneute Anwendung der Tschebyscheffschen Ungleichung liefert nun

$$\mathbb{P}(|X_n - \mathbb{E}X_n| \geq \tfrac{\varepsilon}{2}) \leq \frac{4\mathbb{V}\mathrm{ar}X_n}{\varepsilon^2} \implies \sum_{n=1}^{\infty} \mathbb{P}(|X_n - \mathbb{E}X_n| \geq \varepsilon) \leq \frac{4}{\varepsilon^2} \sum_{n=1}^{\infty} \mathbb{V}\mathrm{ar}X_n < \infty$$

Das Lemma von Borel-Cantelli sichert nun, dass

$$\mathbb{P}(\underbrace{\limsup_{n\to\infty}\{\omega \,|\, |X_n(\omega) - \mathbb{E}X_n| \geq \tfrac{\varepsilon}{2}\}}_{=:A}) = 0$$

Das heißt es die Menge aller $\omega \in \Omega$ für welche $|X_n(\omega) - \mathbb{E}X_n| \geq \frac{\varepsilon}{2}$ unendlich oft auftritt, ist eine Nullmenge. Demnach gilt für alle anderen ω, dass $|X_n(\omega) - \mathbb{E}X_n| \geq \frac{\varepsilon}{2}$ nur endlich oft auftritt, d.h.

$$\forall \omega \in \Omega \backslash A : \exists N_\omega : \forall n \geq N_\omega : |X_n(\omega) - \mathbb{E}X_n| < \tfrac{\varepsilon}{2}$$

Außerdem existiert ein globales N, ab welchem $|\mathbb{E}X_n - a| < \frac{\varepsilon}{2}$. Für ein festes $\omega \in \Omega\backslash A$ gilt demnach für alle $n \geq \max(N, N_\omega)$:

$$|X_n(\omega) - a| \leq |X_n(\omega) - \mathbb{E}X_n| + |\mathbb{E}X_n - a| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon$$

Also ist $\mathbb{P}(\omega \,|\, X_n(\omega) \to a) = \mathbb{P}(\Omega\backslash A) = 1$

Lösung 1C.

a) Für die neuen Zufallsgrößen X_n gilt:

$$\mathbb{P}(|X_n| \geq \varepsilon) = \mathbb{P}(|Y_n \,/\, \ln n| \geq \varepsilon) = \mathbb{P}(Y_n \leq -\varepsilon \ln n) + \mathbb{P}(Y_n \geq \varepsilon \ln n) = 0 + e^{-\lambda\varepsilon \ln n} = \frac{e^{-\lambda\varepsilon}}{n}$$

Damit folgt

$$\mathbb{P}(|X_n| \geq \varepsilon) \xrightarrow{n\to\infty} 0, \quad \text{aber} \quad \sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq \varepsilon) = e^{-\lambda\varepsilon} \cdot \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Also $X_n \xrightarrow{\mathbb{P}} 0$ und $X_n \not\xrightarrow{f.s.} 0$ (Lemma von Borel-Cantelli).

b) Analog zu a) gilt

$$\mathbb{P}(|X_n| \geq \varepsilon) = \frac{e^{-\lambda\varepsilon}}{n^{\ln n}} \leq \frac{e^{-\lambda\varepsilon}}{n^2} \quad \text{für } n \geq 8$$

und so folgt

$$\mathbb{P}(|X_n| \geq \varepsilon) \xrightarrow{n\to\infty} 0 \quad \text{und} \quad \sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq \varepsilon) \leq e^{-\lambda\varepsilon} \cdot \sum_{n=1}^{7} \frac{1}{n^{\ln n}} + e^{-\lambda\varepsilon} \cdot \sum_{n=8}^{\infty} \frac{1}{n^2} < \infty$$

Also $X_n \xrightarrow{\mathbb{P}} 0$ und $X_n \xrightarrow{f.s.} 0$.

Lösung 1D.

a) Aus dem Lemma von Borel-Cantelli folgt:

$$X_n \xrightarrow{f.s.} X \implies \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty \implies \mathbb{P}(|X_n - X| \geq \varepsilon) < 0 \implies X_n \xrightarrow{\mathbb{P}} X$$

b) Die Verteilungsfunktion der konstanten Zufallsgröße c lautet:

$$F_c(x) = \mathbb{P}(c \leq x) = \mathbb{1}_{[c,\infty)}(x)$$

Wenn also punktweise $F_{X_n} \xrightarrow{n\to\infty} F_c$, dann gilt für $\varepsilon > 0$:

$$\begin{aligned}\mathbb{P}(|X_n - c| \geq \varepsilon) &= \mathbb{P}(X_n \leq -\varepsilon + c) + \mathbb{P}(X_n \geq \varepsilon + c) = F_{X_n}(c-\varepsilon) + 1 - F_{X_n}(c+\varepsilon)\\ &\xrightarrow{n\to\infty} \mathbb{1}_{[c,\infty)}(c-\varepsilon) + 1 - \mathbb{1}_{[c,\infty)}(c+\varepsilon) = 0 + 1 - 1 = 0\end{aligned}$$

Lösung 1E.
Analog zum Beweis von Lemma 1.1 folgt für f nichtnegativ und nichtfallend, dass

$$\begin{aligned}\mathbb{E}[f(Y)] &= \int_{\mathbb{R}} f(y)\,\mathrm{d}F_Y(y) \geq \int_{\{y\geq\varepsilon\}} f(y)\,\mathrm{d}F_Y(y)\\ &\geq \int_{\{y\geq\varepsilon\}} f(\varepsilon)\,\mathrm{d}F_Y(y) = f(\varepsilon)\cdot\mathbb{P}(Y \geq \varepsilon)\\ \implies \mathbb{P}(Y \geq \varepsilon) &\leq \frac{\mathbb{E}[f(Y)]}{f(\varepsilon)}\end{aligned}$$

wenn $\mathbb{E}[f(Y)]$ existiert.

a) Für $f(x) = e^{ax}$ folgt

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[e^{aX}]}{e^{a\varepsilon}}$$

b) Für $Y = |X - \mathbb{E}X|$ folgt

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathbb{E}[f(|X - \mathbb{E}X|)]}{f(\varepsilon)}$$

Lösung 1F.
Wir zeigen, dass $\log X_n \xrightarrow{f.s.} 1$. Da die fast sichere Konvergenz unter stetigen Abbildungen (hier $x \mapsto \exp x$) invariant ist, folgt daraus die Behauptung.

Es gilt

$$\log X_n = \frac{1}{n} \cdot \sum_{i=1}^{n} -\log Y_i$$

Für die Hilfsgröße $Z_i = -\log Y_i$ ergibt sich

$$\mathbb{E}Z_1 = \int_0^1 \log y \, \mathrm{d}y = -y \log y + y \Big|_{y=0}^{1} = 1$$

und somit liefert das Starke Gesetz der großen Zahlen, dass

$$\log X_n = \frac{1}{n} \cdot \sum_{i=1}^{n} -\log Y_i = \frac{1}{n} \cdot \sum_{i=1}^{n} Z_i \xrightarrow{f.s.} \mathbb{E}Z_1 = 1$$

Lösung 1G.
Die skalare Multiplikativität folgt daraus, dass die stochastische und fast sichere Konvergenz unter stetigen Abbildungen invariant sind.

Für die Additivität seien (X_n) und (Y_n) Folgen von Zufallsgrößen mit $X_n \xrightarrow{\mathbb{P}} X$ und $Y_n \xrightarrow{\mathbb{P}} Y$. Wir verwenden, dass

$$P(Z_1 + Z_2 \geq z) \leq P(Z_1 \geq \tfrac{z}{2} \text{ oder } Z_2 \geq \tfrac{z}{2}) \leq P(Z_1 \geq \tfrac{z}{2}) + P(Z_2 \geq \tfrac{z}{2})$$

und folgern damit

$$\begin{aligned}\mathbb{P}(|X_n + Y_n - (X + Y)| \geq \varepsilon) &= \mathbb{P}(|X_n - X| + |Y_n - Y| \geq \varepsilon) \\ &\leq \underbrace{\mathbb{P}(|X_n - X| \geq \tfrac{\varepsilon}{2})}_{\to 0} + \underbrace{\mathbb{P}(|Y_n - Y| \geq \tfrac{\varepsilon}{2})}_{\to 0} \xrightarrow{n\to\infty} 0\end{aligned}$$

Lösung 1H.
Sei $\varepsilon > 0$. Wegen $x_n \to x$ existiert ein $N_1 \in \mathbb{N}$ so dass $|x_n - x| < \frac{\varepsilon}{2}$ für $n \geq N_1$.

Weiter gilt für genau dieses N_1, dass $\frac{1}{n}\sum_{i=1}^{N_1}(x_n - x) \to 0$ und damit existiert ein weiteres $N_2 \in \mathbb{N}$ mit $|\frac{1}{n}\sum_{i=1}^{N_1}(x_n - x)| \leq \frac{\varepsilon}{2}$ für $n \geq N_2$.

Für alle $n \geq \max(N_1, N_2)$ gilt nun

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^{n} x_n - x\right| &= \left|\frac{1}{n}\sum_{k=1}^{n}(x_n - x)\right| = \left|\frac{1}{n}\sum_{k=1}^{N_1}(x_n - x) + \frac{1}{n}\sum_{k=N_1+1}^{n}(x_n - x)\right| \\ &\leq \left|\frac{1}{n}\sum_{k=1}^{N_1}(x_n - x)\right| + \frac{1}{n}\sum_{k=N_1+1}^{n}|x_n - x| < \frac{\varepsilon}{2} + \frac{n-N}{n} \cdot \frac{\varepsilon}{2} \leq \varepsilon\end{aligned}$$

Lösung 1I.

a) Die Dichte der Gleichverteilung ist gegeben durch:

$$f_X(x) = \frac{1}{2a} \cdot \mathbb{1}_{[-a,a]}(x)$$

Damit ergibt sich für $t \neq 0$:

$$\Psi_X(t) = \int_{-a}^{a} e^{itx} \cdot \frac{1}{2a} \,\mathrm{d}x = \frac{1}{2a}\frac{1}{it} \cdot e^{itx}\Big|_{x=-a}^{a} = \frac{e^{-iat} - e^{iat}}{2iat} = \sin(at)$$

und für $t = 0$ ist $\Phi_X(0) = 1$.

b) Die Zweipunktverteilung ist diskret mit $x_1 = 1, x_2 = -1$ und $p_1 = p_2 = 0.5$. Somit ergibt sich

$$\Psi_Y(t) = \frac{e^{it} + e^{-it}}{2} = \cos(t)$$

c) Die Binomialverteilung ist die Summe von unabhängigen Bernoulliverteilungen $B_1, \ldots, B_n \sim$ Ber(). Für diese ergibt sich analog zu b):

$$\Psi_{B_i}(t) = (1-p) + pe^{it}$$

Damit folgt für die Summe:

$$\Psi_Z(t) = \prod_{i=1}^{n} \Psi_{B_i}(t) = \left(1 - p + pe^{it}\right)^n$$

Lösung 1J.

Ein Poissonverteilung P mit Parameter $\lambda > 0$ hat die Verteilungsfunktion $e^{-\lambda} \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k}{k!}$:

Demnach lässt sich der gesuchte Term mit einer Poissonverteilung S_n mit Parameter n darstellen durch:

$$e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!} = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} - e^{-n} = \mathbb{P}(S_n \leq n) - e^{-n}$$

Zur Verwendung des Zentralen Grenzwertsatzes muss S_n als Summe von anderen Verteilungen geschrieben werden. Da die Summe von Poissonverteilungen wieder poissonverteilt ist, d.h. für $X \sim$ Pois(λ_1) und $Y \sim$ Pois(λ_2) gilt $X + Y \sim$ Pois($\lambda_1 + \lambda_2$), ergibt sich die Wahl $X_1, X_2, \cdots \sim$ Pois(1) mit $S_n = \sum_{i=1}^{n} X_i$. Damit folgt:

$$\lim_{n\to\infty} e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!} = \lim_{n\to\infty} \mathbb{P}(S_n \leq n) - e^{-n} = \Psi\left(\frac{n - \mathbb{E}[S_n]}{\sqrt{D^2[S_n]}}\right) = \Psi\left(\frac{n-n}{\sqrt{n}}\right) = \Psi(0) = 0.5$$

Lösung 2A.

Offensichtlich schließen sich die Begriffe Lagemaß und Streuungsmaß gegenseitig aus. Für das Stichprobenmittel ϕ_1 gilt:

$$\varphi_1(x_1 + c, \ldots, x_n + c) = \frac{1}{n}\sum_{i=1}^{n}(x_i + c) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n} \cdot nc = \varphi_1(x_1, \ldots, x_n) + c$$

Es ist also ein Lagemaß.

Für die Stichprobenvarianz gilt:

$$\begin{aligned}
\varphi_2(x_1 + c, \ldots, x_n + c) &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i + c - \varphi_1(x_1 + c, \ldots, x_n + c))^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i + c - \varphi_1(x_1, \ldots, x_n) - c)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \varphi_1(x_1, \ldots, x_n))^2 = \varphi_2(x_1, \ldots, x_n)
\end{aligned}$$

Sie ist also ein Streumaß

Lösung 2B.
Für den Erwartungswert gilt:

$$\mathbb{E}[\varphi(X_1,\dots,X_n)] = \sum_{j=1}^{n} \alpha_j \mathbb{E}[X_i] = \mu \cdot \sum_{j=1}^{n} \alpha_j$$

Das heißt, damit $\mathbb{E}[\varphi(X_1,\dots,X_n)] = \mu$, muss $\sum_{j=1}^{n} \alpha_j = 1$ gelten.

Zur Minimierung der Varianz betrachten wir ein Optimierungsproblem mit Nebenbedingung durch die Methode des Lagrange-Multiplikators. Dabei gilt wegen der Unabhängigkeit der X_i, dass

$$D^2[\varphi(X_1,\dots,X_n)] = \sum_{j=1}^{n} \alpha_j^2 D^2[X_i] = \nu \cdot \sum_{j=1}^{n} \alpha_j^2$$

und somit folgt

$$\Lambda(\alpha_1,\dots,\alpha_n,\lambda) = \nu \cdot \sum_{j=1}^{n} \alpha_j^2 - \lambda \cdot \left(\sum_{j=1}^{n} \alpha_j - 1 \right)$$

Eine Notwendige Bedingung eines Minimum ist $\nabla\Lambda = 0$, d.h. wir erhalten folgendes Gleichungssystem:

$$\frac{\partial\Lambda}{\partial\alpha_j} = 2\alpha_j - \lambda \overset{!}{=} 0 \quad \forall j = 1,\dots,n \tag{1}$$

$$\frac{\partial\Lambda}{\partial\lambda} = -\sum_{j=1}^{n} \alpha_j + 1 \overset{!}{=} 0 \tag{2}$$

Aus (1) folgt, dass alle $\alpha_j = \frac{\lambda}{2}$ konstant sind. Das eingesetzt in (2) liefert

$$0 = -\sum_{j=1}^{n} \alpha_j + 1 = -\sum_{j=1}^{n} \frac{\lambda}{2} + 1 = -\frac{n}{2} \cdot \lambda + 1 \implies \lambda = \frac{2}{n}$$

Demnach muss $\alpha_j = \frac{1}{n}$ und wir erhalten

$$\phi(X_1,\dots,X_n) = \sum_{j=1}^{n} \frac{1}{n} \cdot X_j = \overline{X}_n$$

als Extremwert. Dieser ist auch offensichtlich ein Minimum, da die Varianz strikt konvex ist:

$$\nabla^2 D^2[\phi(x_1,\dots,x_n)] = 2\nu I_n$$

Lösung 2C.

1. Für $Y \sim \Gamma(b,p)$ gilt:

$$\Psi_Y(t) = \mathbb{E}[e^{itY}] = \int_0^\infty e^{ity} \frac{b^p}{\Gamma(p)} e^{-by} y^{p-1}\,\mathrm{d}y = \frac{b^p}{\Gamma(p)} \int_0^\infty e^{-(b-it)y} y^{p-1}\,\mathrm{d}y$$

Wir substituieren $z = (b-it)y$, also $y = \frac{z}{b-it}$ und $\mathrm{d}y = \frac{1}{b-it}\,\mathrm{d}z$:

$$\begin{aligned}\Psi_Y(t) &= \frac{b^p}{\Gamma(p)} \int_0^\infty e^{-(b-it)y} y^{p-1}\,\mathrm{d}y = \frac{b^p}{\Gamma(p)} \int_{z=0}^\infty e^{-z} \left(\frac{z}{b-it}\right)^{p-1} \frac{1}{b-it}\,\mathrm{d}z \\ &= \frac{b^p}{\Gamma(p)} \cdot \frac{1}{(b-it)^p} \int_0^\infty e^{-z} z^{p-1}\,\mathrm{d}z = \frac{b^p}{\Gamma(p)} \cdot \frac{1}{(b-it)^p} \cdot \Gamma(p) = \left(1 - \frac{b}{it}\right)^p\end{aligned}$$

2. Seien $Y_1 \sim \Gamma(b, p1)$ und $Y_2 \sim \Gamma(b, p_2)$ unabhängig so gilt:

$$\Psi_{Y_1+Y_2}(t) = \Psi_{Y_1}(t) \cdot \Psi_{Y_2}(t) = \left(1 - \frac{b}{it}\right)^{p_1} \cdot \left(1 - \frac{b}{it}\right)^{p_2} = \left(1 - \frac{b}{it}\right)^{p_1+p_2}$$

Dies ist wieder die charakteristische Funktion einer Gammaverteilung, also folgt $Y_1{+}Y_2 \sim \Gamma(b, p_1{+}p_2)$.

Lösung 2D.
Für die charakteristische Funktion der Zufallsgröße X_i^2 gilt:

$$\Psi_{X_i^2}(t) = \mathbb{E}[e^{itX}] = \int_0^\infty e^{itx^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \cdot \int_0^\infty e^{-(x(\frac{1}{2}-it))^2} \,\mathrm{d}x$$

Substitution von $z = x(\frac{1}{2} - it)$ und Verwendung des Gaußschen Integrals $\int_0^\infty e^{-x^2} \,\mathrm{d}x = \frac{\sqrt{\pi}}{2}$ folgt

$$\Psi_{X_i^2}(t) = \frac{1}{\sqrt{2\pi}}.$$