Kapitel W

Der zentrale Grenzwertsatz

Zweck des vorliegenden Heftes ist eine axiomatische Begründung der Wahrscheinlichkeitsrechnung. […] Vor Entstehung der Lebesgueschen Maß- und Integrationstheorie war diese Aufgabe ziemlich hoffnungslos. Nach den Lebesgueschen Untersuchungen lag die Analogie zwischen dem Maße einer Menge und der Wahrscheinlichkeit eines Ereignisses sowie zwischen dem Integral einer Funktion und der mathematischen Erwartung einer zufälligen Größe auf der Hand.

Andrei Kolmogorov (1902–1987),
Grundbegriffe der Wahrscheinlichkeitsrechnung (1933)

Inhalt dieses Kapitels W

1. Der zentrale Grenzwertsatz (ZGS)
   - Stochastisch unabhängige Zufallsvariablen
   - Beispiele und erste Beobachtungen
   - Der zentrale Grenzwertsatz
2. Statistische Anwendungen
   - Konfidenzintervalle
   - Fehlerfortpflanzung
   - Regressionsanalyse
3. Fazit: der zentrale Grenzwertsatz
   - Zusammenfassung und Verständnisfragen
   - Summen von Zufallsvariablen und Grenzwertsätze
   - Weitere Aufgaben und Anwendungsbeispiele
4. Analytische Methoden der Wahrscheinlichkeitsrechnung
   - Laplace– und Fourier–Transformation von WMaßen
   - Beweisidee des zentralen Grenzwertsatzes
   - Entropie verstehen und anwenden

Motivation und Zielsetzung

Wir kennen bereits zwei nützliche und wichtige Grenzwertsätze:

🙂 Bei seltenen Ereignissen können wir die **Binomialverteilung** $\mathrm{B}(n,t)$ durch die **Poisson–Verteilung** $\mathrm{P}(nt)$ nähern mit guter Fehlerschranke. Das **Gesetz der kleinen Zahlen** besagt $\mathrm{B}(n,\lambda/n) \to \mathrm{P}(\lambda)$ für $n \to \infty$.

🙂 Der **lokale Grenzwertsatz** hingegen nützt für beliebige $0 < t < 1$ und approximiert Binomialverteilungen durch **Normalverteilungen**:

$$\mathbf{P}(a \le S \le b) \overset{\text{exakt}}{=} \sum_{k=a}^{b} \binom{n}{k} t^k (1-t)^{n-k} \overset{\text{LGS}}{=} \int_{\xi=\alpha}^{\beta} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \;+\; \delta$$

🙂 Die Normalverteilung ist zudem **universell**: Sie entsteht immer, wenn sich viele unabhängige, ähnlich große Beiträge aufsummieren:

$$S = X_1 + \ldots + X_n \implies \mathbf{P}(a \le S \le b) \overset{\text{ZGS}}{=} \int_{\xi=\alpha}^{\beta} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \;+\; \delta$$

Entscheidend ist hier, wie immer, den Fehler δ geeignet zu beschränken. Die genaue Aussage ist Inhalt des **zentralen Grenzwertsatzes** W1D.

Vorgehensweise

Unser Ziel sind **rationale Entscheidungen unter Unsicherheit**.
Hierzu müssen wir stochastische Gesetzmäßigkeiten klären und nutzen.

Eine Messung betrachten wir als **Zufallsvariable** $X : (\Omega, \mathbf{P}) \to \mathbb{R}$; sie definiert eine WVerteilung auf $\mathbb{R}$, Erwartung $\mathbf{E}(X)$ und Varianz $\mathbf{V}(X)$.

Wir können damit die **Chebychev–Ungleichung** T3F formulieren und hieraus das (schwache) **Gesetz der großen Zahlen** T3J ableiten.

Entscheidend ist der Begriff der **Unabhängigkeit** von Zufallsvariablen.
Die intuitive Anschauung präzisieren wir durch Definitionen und Sätze.

Höhepunkt dieses Kapitels ist der **zentrale Grenzwertsatz** W1D.
Diesen formuliere ich mit einer allgemeinen, expliziten Fehlerschranke.

Die vom ZGS vorhergesagte **Normalverteilung** ist ein universelles Werkzeug und bietet für viele Situationen ein brauchbares Modell.

Sie dient insbesondere zur Bestimmung von **Konfidenzintervallen** und wird daher in nahezu allen praktischen Anwendungen benötigt.

Hierzu gebe ich abschließend erste einfache Anwendungsbeispiele.
Mit diesen Grundlagen gewappnet beginnt die mathematische **Statistik**.

Erinnerung: Unabhängigkeit von Ereignissen

Wir erinnern an die Definition T2E der Unabhängigkeit: Im WRaum $(\Omega, \mathscr{A}, \mathbf{P})$ heißt eine Familie von Ereignissen $A_1, A_2, \dots, A_n \in \mathscr{A}$ **(stochastisch) unabhängig** wenn folgende Produktformel gilt

$$\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_\ell}) = \mathbf{P}(A_{i_1})\,\mathbf{P}(A_{i_2}) \cdots \mathbf{P}(A_{i_\ell})$$

für jede Auswahl von Indizes $1 \le i_1 < i_2 < \dots < i_\ell \le n$.

Zwei Ereignisse A, B sind unabhängig, wenn gilt:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$$

Drei Ereignisse A, B, C sind unabhängig, wenn gilt:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$$
$$\mathbf{P}(A \cap C) = \mathbf{P}(A)\,\mathbf{P}(C)$$
$$\mathbf{P}(B \cap C) = \mathbf{P}(B)\,\mathbf{P}(C)$$
$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C)$$

⚠ Die Tripelbedingung folgt nicht aus den drei Paarbedingungen! Sie kennen dieses Problem von linearer Unabhängigkeit von Vektoren.

Erinnerung: Produktwahrscheinlichkeit

Definition W1A: Produkt von Wahrscheinlichkeitsräumen

Wir beschreiben die Durchführung unabhängiger Experimente durch Wahrscheinlichkeitsräume $(\Omega_1, \mathscr{A}_1, \mathbf{P}_1), \dots, (\Omega_n, \mathscr{A}_n, \mathbf{P}_n)$.

Der **Produktraum** $(\Omega, \mathscr{A}, \mathbf{P})$ basiert auf dem kartesischen Produkt

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \big\{\, (\omega_1, \dots, \omega_n) \;\big|\; \omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n \,\big\}.$$

Die **Ereignismenge** $\mathscr{A} = \mathscr{A}_1 \otimes \dots \otimes \mathscr{A}_n$ wird erzeugt von Produkten

$$A = A_1 \times \dots \times A_n \quad \text{mit} \quad A_1 \in \mathscr{A}_1, \dots, A_n \in \mathscr{A}_n.$$

Das **Produktmaß** $\mathbf{P} = \mathbf{P}_1 \otimes \dots \otimes \mathbf{P}_n : \mathscr{A} \to [0,1]$ ist definiert durch

$$\mathbf{P}(A_1 \times \dots \times A_n) = \mathbf{P}_1(A_1) \cdots \mathbf{P}_n(A_n).$$

Wir erhalten den WRaum $(\Omega, \mathscr{A}, \mathbf{P})$ durch eindeutige Fortsetzung.

Für diskrete WRäume kennen und nutzen wir dies bereits, siehe U1A
Sind $\mathbf{P}_1, \dots, \mathbf{P}_n$ kontinuierlich auf $\mathbb{R}$ mit Dichten $f_1, \dots, f_n$, so ist $\mathbf{P}$ auf $\mathbb{R}^n$ kontinuierlich, mit Produktdichte $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$.

Unabhängige Zufallsvariablen

Zwei Zufallsvariablen X_1, X_2 sind unabhängig, wenn das Ergebnis von X_1 nicht die Wkten von X_2 beeinflusst. Dies präzisieren wir wie folgt:

Definition W1B: Unabhängigkeit von Zufallsvariablen

Eine Familie $X = (X_1, \dots, X_n)$ von Zufallsvariablen $X_k : \Omega \to \mathbb{R}$ heißt **(stochastisch) unabhängig**, wenn für alle Intervalle $I_1, \dots, I_n \subseteq \mathbb{R}$ gilt

$$\mathbf{P}(\; X_1 \in I_1,\; \dots,\; X_n \in I_n \;) = \mathbf{P}(X_1 \in I_1) \cdots \mathbf{P}(X_n \in I_n).$$

Ausführlich heißt das: Im WRaum $(\Omega, \mathbf{P})$ ist jede Familie $(A_1, \dots, A_n)$ von Ereignissen $A_k = X_k^{-1}(I_k) = \{\, \omega \in \Omega \mid X_k(\omega) \in I_k \,\}$ unabhängig.

Die gemeinsame Verteilung von $X_1, \dots, X_n$ ist dann das Produktmaß

$$\mathbf{P}_X = \mathbf{P}_{(X_1,\dots,X_n)} = \mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_n}.$$

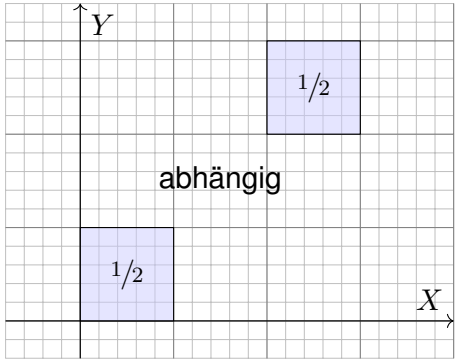
Sind $\mathbf{P}_{X_1}, \dots, \mathbf{P}_{X_n}$ auf $\mathbb{R}$ kontinuierliche WMaße mit Dichten $f_1, \dots, f_n$, so ist auch das Produktmaß $\mathbf{P}_X$ auf $\mathbb{R}^n$ kontinuierlich, mit Produktdichte

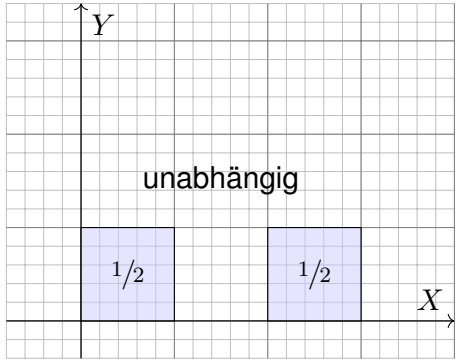
$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

Unabhängige Zufallsvariablen

Anschaulich: Zwei Zufallsvariablen X, Y sind unabhängig, wenn das Ergebnis von X nicht die Wkten von Y beeinflusst, und umgekehrt.

Beispiel: Die gemeinsame Verteilung von $(X, Y) : \Omega \to \mathbb{R}^2$ sei wie folgt:





Die Skizze zeigt die Gleichverteilung auf den Mengen $Q, R \subseteq \mathbb{R}^2$ mit $Q = ([0,1] \times [0,1]) \cup ([2,3] \times [2,3])$, $R = ([0,1] \times [0,1]) \cup ([2,3] \times [0,1])$.
Im ersten Fall ist das Paar (X, Y) abhängig, im zweiten unabhängig.

Erwartungswert und Varianz bei Unabhängigkeit

Satz W1c: Fubini für unabhängige Zufallsvariablen

(1) Erwartungen unabhängiger Zufallsvariablen X, Y multiplizieren sich:

$$\mathbf{E}(X \cdot Y) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$$

Ihre Kovarianz $\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ verschwindet somit.

(2) Varianzen unabhängiger Zufallsvariablen X, Y addieren sich:

$$\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\mathbf{Cov}(X,Y)$$

Aufgabe: Rechnen Sie dies nach! **Lösung:** (1) Dank Unabhängigkeit:

$$\mathbf{E}(X \cdot Y) \overset{\text{Def}}{=} \int_{\mathbb{R}^2} xy \,\mathrm{d}\mathbf{P}_{(X,Y)} \overset{\substack{\text{Fub}\\ \text{Un}}}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} xy \,\mathrm{d}\mathbf{P}_X \,\mathrm{d}\mathbf{P}_Y$$
$$\overset{\text{Lin}}{=} \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} x \,\mathrm{d}\mathbf{P}_X \Big) y \,\mathrm{d}\mathbf{P}_Y \overset{\text{Lin}}{=} \int_{\mathbb{R}} x \,\mathrm{d}\mathbf{P}_X \cdot \int_{\mathbb{R}} y \,\mathrm{d}\mathbf{P}_Y \overset{\text{Def}}{=} \mathbf{E}(X) \cdot \mathbf{E}(Y)$$

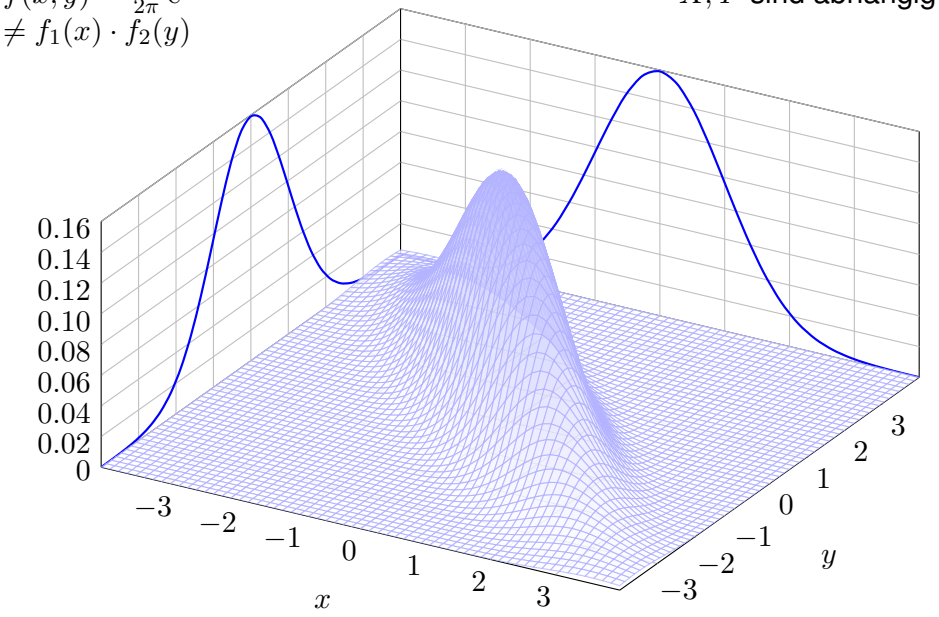
(2) Wir setzen die Definition ein und vereinfachen geduldig: T331

$$\mathbf{V}(X+Y) \overset{\substack{\text{Def}\\ \text{Lin}}}{=} \mathbf{E}\big[(X - \mathbf{E}(X) + Y - \mathbf{E}(Y))^2\big]$$
$$\overset{\text{Bin}}{=} \mathbf{E}\big[(X - \mathbf{E}(X))^2 + (Y - \mathbf{E}(Y))^2 + 2(\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y))\big]$$

Beispiel einer zweidimensionalen Normalverteilung

$f(x,y) = \frac{1}{2\pi} \mathrm{e}^{-(x+y)^2-(x-y)^2/16}$
$\neq f_1(x) \cdot f_2(y)$

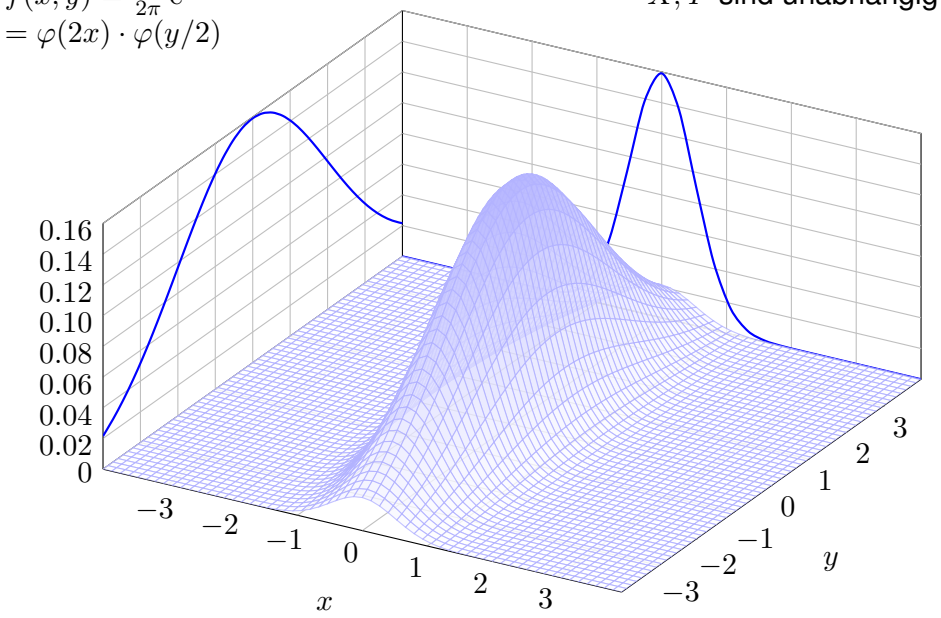
X, Y sind abhängig!



Beispiel einer zweidimensionalen Normalverteilung

$f(x,y) = \frac{1}{2\pi} \mathrm{e}^{-2x^2+y^2/8}$
$= \varphi(2x) \cdot \varphi(y/2)$

X, Y sind unabhängig!



Mehrdimensionale Normalverteilung

Die n–dimensionale Standard-Normalverteilung hat die WDichte

$$f : \mathbb{R}^n \to \mathbb{R} : x \mapsto f(x) = \frac{\mathrm{e}^{-x^2/2}}{\sqrt{(2\pi)^n}} = \varphi(x_1) \cdots \varphi(x_n).$$

Die Zufallsvariablen $Z_k : \mathbb{R}^n \to \mathbb{R} : z \mapsto z_k$ sind unabhängig und identisch verteilt gemäß der eindim. Standard-Normalverteilung, $Z_k \sim \mathrm{N}(0,1)$.

Sei nun $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ eine Familie von Zufallsvariablen. Dies definiert den Vektor $\mu \in \mathbb{R}^n$ mit $\mu_k = \mathbf{E}(X_k)$ und die symmetrische Matrix $\Sigma \in \mathbb{R}^{n \times n}$ mit $\Sigma_{k\ell} = \mathbf{Cov}(X_k, X_\ell)$, insbesondere $\Sigma_{kk} = \mathbf{V}(X_k)$. Wir nehmen Σ als positiv definit an, insbesondere $\det \Sigma > 0$. Wir nennen X normalverteilt, kurz $X \sim \mathrm{N}(\mu, \Sigma)$, falls $\mathbf{P}_X$ folgende WDichte hat:

$$f(x) = \frac{\exp[-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)]}{\sqrt{(2\pi)^n \det \Sigma}}$$

Übung: (1) Für jede affin-lineare Transformation $Y = AX + v$ mit Matrix $A \in \mathrm{GL}_n \mathbb{R}$ und Vektor $v \in \mathbb{R}^n$ gilt $Y \sim \mathrm{N}(A\mu + v, A\Sigma A^{-1})$.

(2) Sind die Komponenten von X paarweise unkorreliert, so sind sie sogar stochastisch unabhängig (analog zu Indikatorfunktionen T350).

Motivation zum zentralen Grenzwertsatz

Anwendung: Jede Messung X_k ist zufälligen Störungen unterworfen. Sie habe Erwartung $\mu_k = \mathbf{E}(X_k)$ und Varianz $\sigma_k^2 = \mathbf{E}(|X_k - \mu_k|^2) > 0$.

Die Summe $S = X_1 + \ldots + X_n$ hat die Erwartung $\mu = \mu_1 + \ldots + \mu_n$. Bei Unabhängigkeit addieren sich die Varianzen $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$.

Das sind soweit nur grobe Kenngrößen. Jetzt geschieht das Wunder: **Typischerweise** ähnelt die Verteilung von S der Normalverteilung:

$$S \sim \mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2), \quad \text{das heißt ausgeschrieben}$$

$$\mathbf{P}(a \le S \le b) \approx \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \overset{\text{Tabelle}}{=} \Big[\Phi\Big]_{(a-\mu)/\sigma}^{(b-\mu)/\sigma}$$

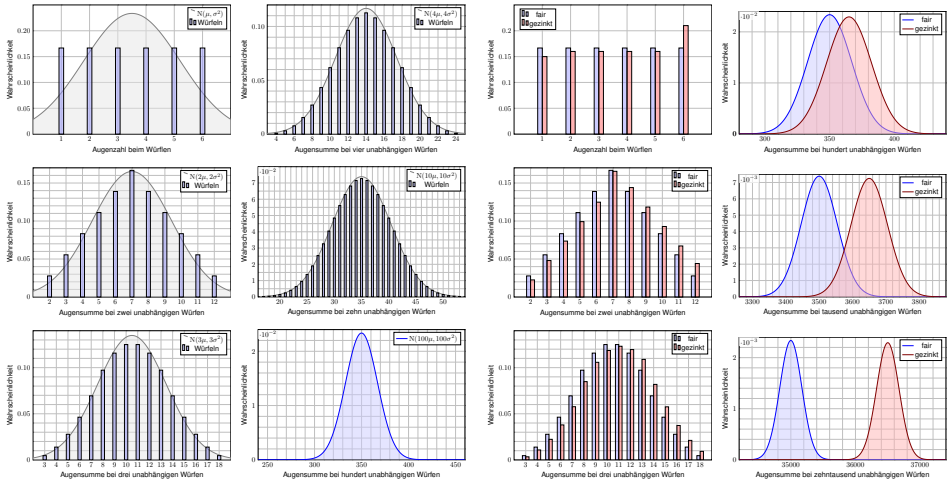
Das ist eine enorme Vereinfachung: Die Verteilung von S ist kompliziert und meist unbekannt, die Normalverteilung hingegen ist explizit & leicht! Die Normalverteilung entsteht, egal welche Verteilungen die X_k haben!

Wie gut ist die Approximation für festes n? Wie groß ist der Fehler δ? Gilt $\delta \to 0$ für $n \to \infty$? Das ist Inhalt des zentralen Grenzwertsatzes!

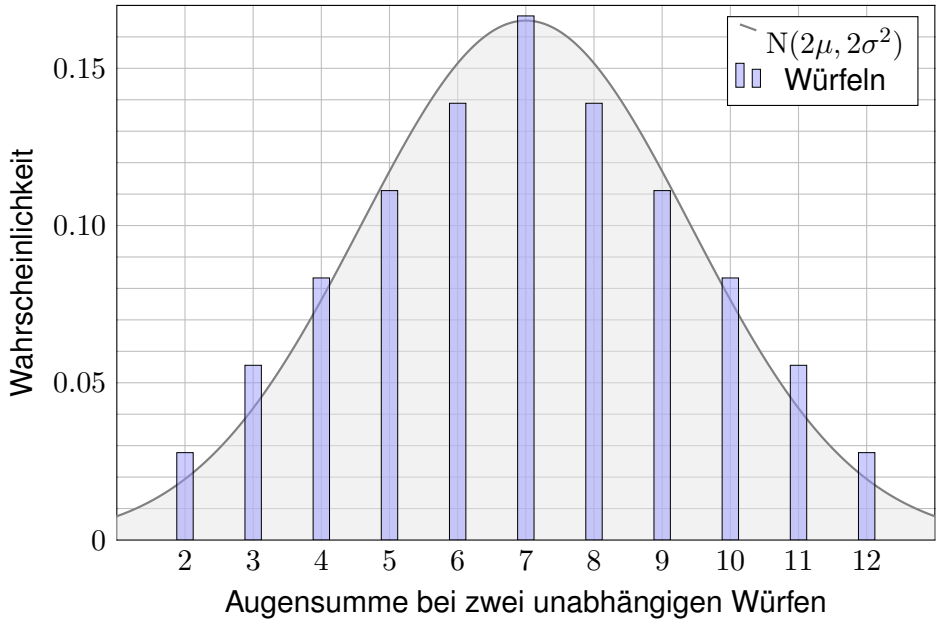
Motivation zum zentralen Grenzwertsatz

Aufgabe: Illustrieren Sie diese Näherung numerisch für Würfelsummen mit einem fairen Würfel (T105) und unserem gezinkten Würfel (T107).

Anschließend erklärt Ihnen der zentrale Grenzwertsatz W1D, was von diesen konkreten Beobachtungen allgemein garantiert werden kann.

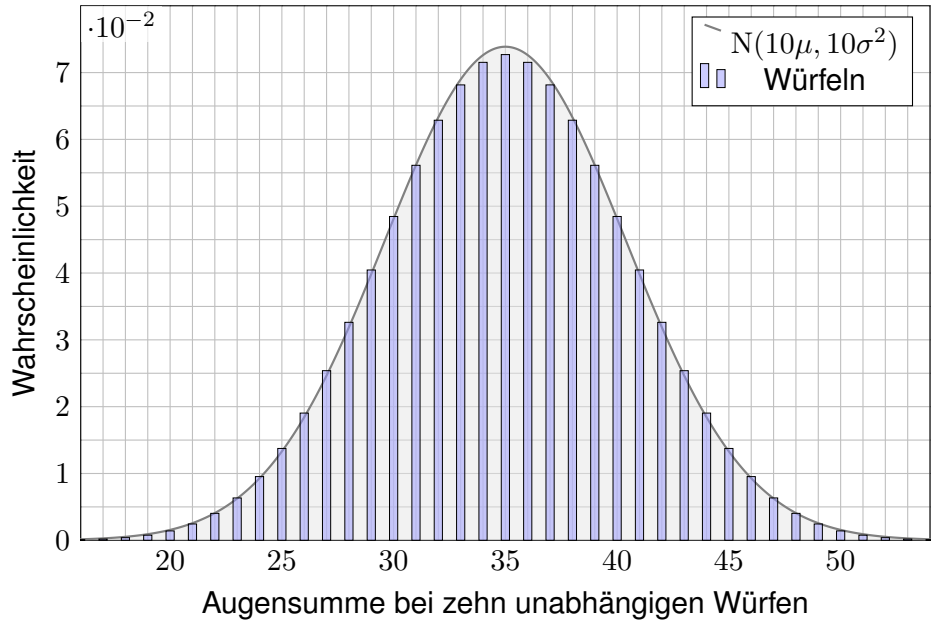


Augensumme bei zweimaligem Würfeln



☺ Die Augensumme hat Erwartung $2\mu = 7$ und Varianz $2\sigma^2 = 35/6$.

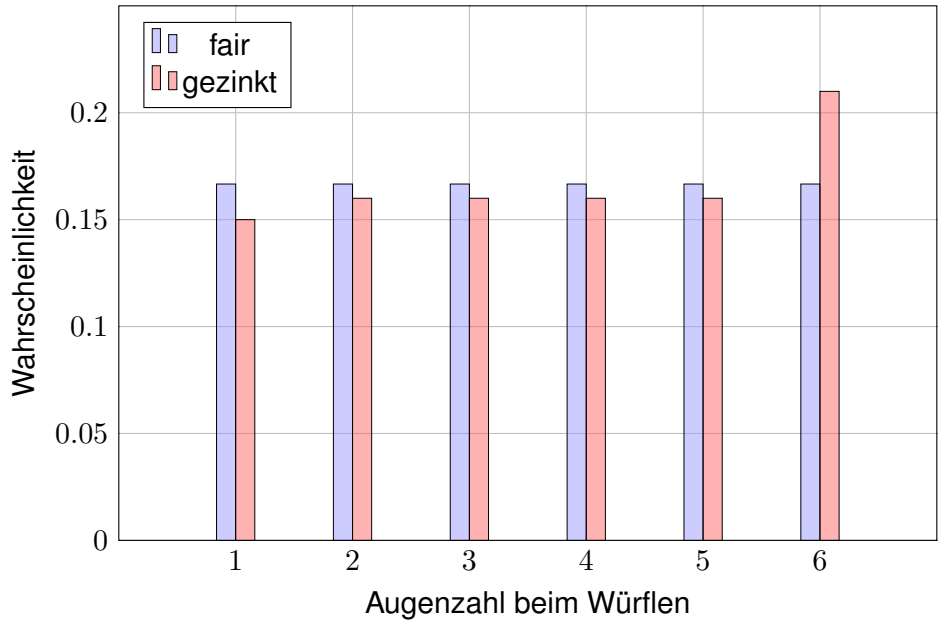
Augensumme bei zehnmaligem Würfeln



☺ Große Zahlen: Die Verteilung von $\hat{X}$ konzentriert sich um μ.

Augenzahl bei einmaligem Würfeln

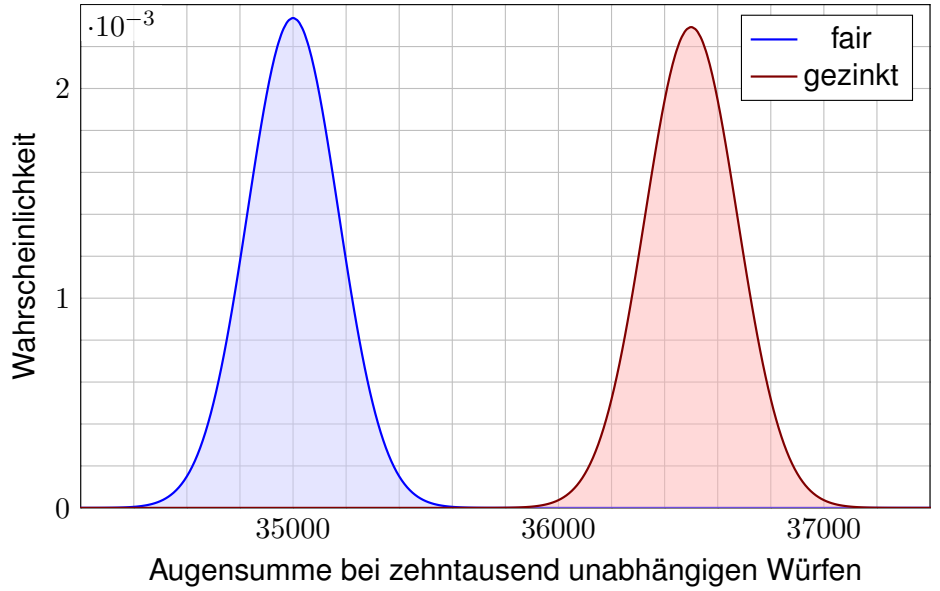
W113
Ausführung



🙂 Gezinkt ist die Erwartung $\mu_2 = 3.65$ etwas größer als fair $\mu_1 = 3.5$.

Augensumme bei zehntausendmaligem Würfeln

W114
Ausführung



🙂 Bei zehntausend Würfen unterscheiden sich die Ergebnisse sicher.
So können wir den gezinkten Würfel experimentell erkennen! W329

Was sehen wir in diesen Beispielen?

W115
Erläuterung

Die Wahrscheinlichkeitsverteilung des Experiments X_k, hier Würfeln, bestimmt die Erwartung $\mu = \mathbf{E}(X_k)$ und die Varianz $\sigma^2 = \mathbf{V}(X_k)$.
Das sind, wie immer so auch hier, die wichtigsten Kenngrößen.

🙂 Die Summe $S = X_1 + \ldots + X_n$ von n unabhängigen Wiederholungen hat dann Erwartung $n\mu$ und Varianz $n\sigma^2$. ☹ Sie ist nicht binomialverteilt, also können wir den lokalen Grenzwertsatz V3A hier nicht anwenden.

🙂 Nach vielen Wiederholungen spielt die Ausgangsverteilung keine Rolle: Es entsteht immer annähernd die Normalverteilung $\mathrm{N}(n\mu, n\sigma^2)$.
Ihre universelle Rolle ist ein phantastisch nützliche Eigenschaft!

Man kann dies (etwas vage aber poetisch) auch wie folgt formulieren:
Im Kleinen, bei einzelnen Experimenten regiert der Zufall ungezügelt;
im Großen, bei vielen Experimenten muss er sich Gesetzen beugen.

🙂 Diese Gesetzmäßigkeit müssen Sie kennen und sollten Sie nutzen:
Solide Grundlagen ermöglichen Ihnen den sicheren Umgang mit Wkten.
„Nul n'est censé ignorer la loi.“ [Niemand darf das Gesetz ignorieren.]

Was sagt der zentrale Grenzwertsatz?

W116
Erläuterung

🙂 Der ZGS besagt: Die Summe $S = X_1 + \ldots + X_n$ vieler unabhängiger, aber ähnlich großer Zufallsvariablen ist annähernd normalverteilt.

Beispiel: Für unabhängige und identisch verteilte 0–1–Zufallsvariablen $X_k : \Omega \to \{0, 1\}$ ist $S = X_1 + \ldots + X_n$ exakt binomialverteilt, $S \sim \mathrm{B}(n, t)$.
Hier gilt $\mathrm{B}(n, t) \approx \mathrm{N}(\mu, \sigma^2)$ mit expliziten Fehlerschranken (LGS, V3A).

🙂 Im allgemeineren ZGS dürfen $X_1, X_2, \ldots, X_n$ beliebig verteilt sein.
Wir benötigen hierzu allerdings einige Vorsichtsmaßnahmen W309:

1. Die Beiträge $X_1, X_2, \ldots, X_n$ sollen unabhängig sein.
   Für $S = X + X + \ldots + X$ gilt der ZGS sicher nicht.
2. Die Beiträge $X_1, X_2, \ldots, X_n$ sollen alle etwas streuen.
   Für $S = 1 + 1 + \ldots + 1$ gilt der ZGS sicher nicht.
3. Die Beiträge $X_1, X_2, \ldots, X_n$ sollen „ähnlich groß“ sein.

Folgender Satz, formuliert nach A.C. Berry 1941 und C.G. Esséen 1944, präzisiert diese Voraussetzungen in Form des 2. und 3. Moments und quantifiziert den Approximationsfehler durch eine explizite Schranke.

🙂 Allgemein gültige Näherung mit verlässlicher Fehlerabschätzung!

Der zentrale Grenzwertsatz: allgemein und verlässlich

Satz W1D: zentraler Grenzwertsatz, ZGS

Sei $(\Omega, \mathcal{A}, \mathbf{P})$ ein WRaum und $X_1, X_2, X_3, \ldots : \Omega \to \mathbb{R}$ unabhängig mit

1. endlichen Erwartungen $\mu_k = \mathbf{E}(X_k)$ dank $\mathbf{E}(|X_k|) < \infty$,
2. strikt positiven Varianzen $\sigma_k^2 = \mathbf{E}\big(|X_k - \mu_k|^2\big) \geq \sigma_0^2 > 0$,
3. beschränkten dritten Momenten $\rho_k^3 = \mathbf{E}\big(|X_k - \mu_k|^3\big) \leq \rho_0^3 < \infty$.

Die Summe $S = X_1 + \ldots + X_n$ hat die Erwartung $\mu = \mu_1 + \ldots + \mu_n$ und die Varianz $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Es gilt $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$, genauer

$$\mathbf{P}(a \leq S \leq b) = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi + \delta,$$

$$|\delta| \leq \frac{\rho_1^3 + \ldots + \rho_n^3}{(\sigma_1^2 + \ldots + \sigma_n^2)^{3/2}} \leq \frac{\rho_0^3}{\sigma_0^3\sqrt{n}} \searrow 0.$$

Zentriert und normiert zu $S^* := (S - \mu)/\sigma$ erhalten wir $\mathbf{P}_{S^*} \approx \mathrm{N}(0, 1)$.

⚠ Alle drei Voraussetzungen (1–3) werden wirklich benötigt. W309

Vergleich zwischen ZGS und LGS

Aufgabe: Wer ist für $X_1, \ldots, X_n \sim \mathrm{B}(1, t)$ genauer: ZGS oder LGS?

Lösung: Zu $X_k \sim \mathrm{B}(1, t)$ finden wir (T310) $\mu_k = t$ und $\sigma_k^2 = t(1-t)$ sowie $\rho_k^3 = t(1-t)(t^2 + (1-t)^2)$. Die Summe $S = X_1 + \ldots + X_n$ ist $\mathrm{B}(n, t)$–verteilt mit $\mu = nt$ und $\sigma^2 = nt(1-t)$, also exakt / genähert:

$$\mathbf{P}(a \leq S \leq b) = \sum_{k=a}^{b} \binom{n}{k} t^k (1-t)^{n-k} = \int_{(a-1/2-\mu)/\sigma}^{(b+1/2-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi + \delta$$

Für $\sigma \geq 5$ ist der Approximationsfehler δ im LGS/ZGS beschränkt durch

$$|\delta| \overset{\text{LGS}}{<} \frac{1}{6\sigma} = \frac{1}{6\sqrt{nt(1-t)}} \overset{\text{ZGS}}{<} \frac{\rho_0^3}{\sigma_0^3\sqrt{n}} = \frac{t^2 + (1-t)^2}{\sqrt{nt(1-t)}}$$

Der LGS liefert die bessere Konstante: $1 \geq t^2 + (1-t)^2 \geq 1/2 > 1/6$.
Die Fehlerschranke ist also um einen Faktor 3 bis 6 kleiner als im ZGS.
Mit Korrekturterm κ sinkt der Fehler auf $|\varepsilon| < 1/(3\sigma^2) = 1/[3nt(1-t)]$.

🙂 Der lokale Grenzwertsatz ist spezieller, dabei aber auch präziser.

🙂 Der ZGS gilt nicht nur für $S \sim \mathrm{B}(n, t)$, sondern viel allgemeiner!

Vergleich zwischen ZGS und LGS

Der LGS nutzt die Stetigkeitskorrektur $\pm 1/2$ in den Integrationsgrenzen zur genaueren Approximation. Der ZGS kennt diese Korrektur nicht: Er gilt für kontinuierlich verteilte Zufallsvariablen ebenso wie für diskrete, und im kontinuierlichen Fall hat die Stetigkeitskorrektur keinen Sinn.

Aufgabe: Welcher Fehler entsteht hierdurch schlimmstenfalls?

Lösung: Für $S \sim \mathrm{B}(n, t)$ und $k = 0, 1, \ldots, n$ gilt exakt und nach LGS:

$$\mathbf{P}\big(k \leq S \leq k\big) = \binom{n}{k} t^k (1-t)^{n-k} \overset{\text{LGS}}{\approx} \int_{(k-1/2-\mu)/\sigma}^{(k+1/2-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi$$

$$\approx \frac{\mathrm{e}^{-(k-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \leq \frac{1}{\sqrt{2\pi nt(1-t)}}$$

Beachten Sie die Stetigkeitskorrektur $\pm 1/2$ in den Integrationsgrenzen! Andernfalls wäre die Approximation $\mathbf{P}\big(k \leq S \leq k\big) \approx 0$ völlig nutzlos.

Die obere Schranke erhalten wir für $k = \mu$ und nutzen $\sigma^2 = nt(1-t)$. Hier nimmt die Normalverteilung (Glockenkurve) ihr Maximum an.

Vergleich zwischen ZGS und LGS

Der ZGS hingegen liefert ohne Stetigkeitskorrektur leider nur

$$\mathbf{P}\big(k \leq S \leq k\big) \overset{\text{ZGS}}{=} \int_{(k-\mu)/\sigma}^{(k-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi + \delta = 0 + \delta$$

Dieses Integral verschwindet! Belügt uns der ZGS in diesem Fall? Nein, er nützt nur nicht mehr viel, denn er besagt jetzt lediglich:

$$0 \leq \mathbf{P}\big(k \leq S \leq k\big) = \delta \leq \frac{\rho_0^3}{\sigma_0^3\sqrt{n}}$$

Alles ist gut. Der Fehler δ bleibt unter der Fehlerschranke des ZGS:

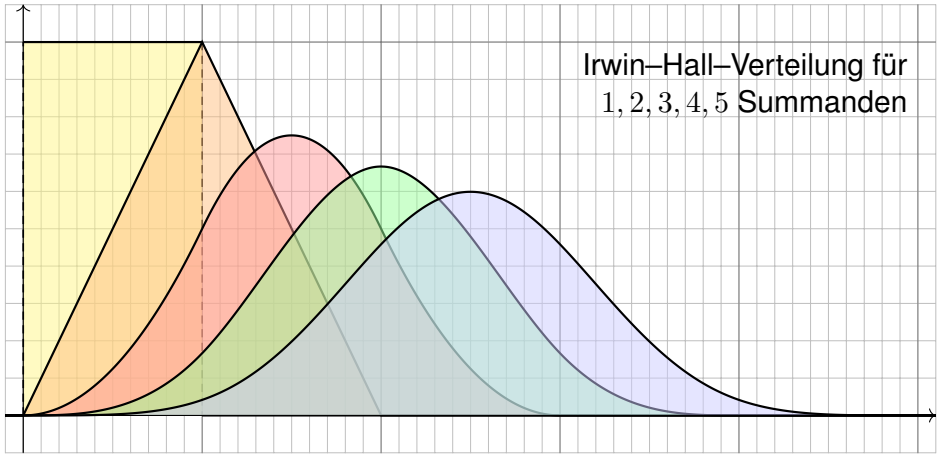
$$\frac{\rho_0^3}{\sigma_0^3\sqrt{n}} = \frac{t^2 + (1-t)^2}{\sqrt{nt(1-t)}} \geq \frac{1}{2\sqrt{nt(1-t)}} \geq \frac{1}{\sqrt{2\pi nt(1-t)}} \geq \delta \geq 0$$

In solch extremen Fällen wird man den ZGS wohl kaum nutzen wollen, doch sie illustrieren eindrücklich, was schlimmstenfalls passieren kann.

🙂 Hier wird der ZGS zwar missbraucht, aber dank δ bleibt alles richtig: Die Approximation kann schlecht sein, die Fehlerschranke stimmt immer.

🙂 Der LGS mit Stetigkeitskorrektur ist hier wesentlich genauer.

Summen gleichverteilter Zufallsvariablen



Aufgabe: Wir betrachten unabhängige gleichverteilte Zufallsvariablen $X_1, X_2, X_3, \ldots : \Omega \to [0,1]$, das heißt $X_k \sim \mathbf{P}_1$ mit WDichte $f_1 = \mathbf{I}_{[0,1]}$. Welche Verteilung hat $S_n = X_1 + \ldots + X_n$? Bestimmen Sie die Dichte:

$$f_n(x) = (f_{n-1} * f_1)(x) = \int_{-\infty}^{\infty} f_{n-1}(x-u) f_1(u)\,\mathrm{d}u = \int_{u=0}^{1} f_{n-1}(x-u)\,\mathrm{d}u$$

Summen gleichverteilter Zufallsvariablen

Lösung: Für $x \in [0,n]$ gilt $f_n(x) = \sum_{k=0}^{\lfloor x \rfloor} (-1)^k \binom{n}{k} (x-k)^{n-1}/(n-1)!$. Die Dichte f_n ist stückweise polynomiell vom Grad n. Für $t \in [0,1]$ gilt:

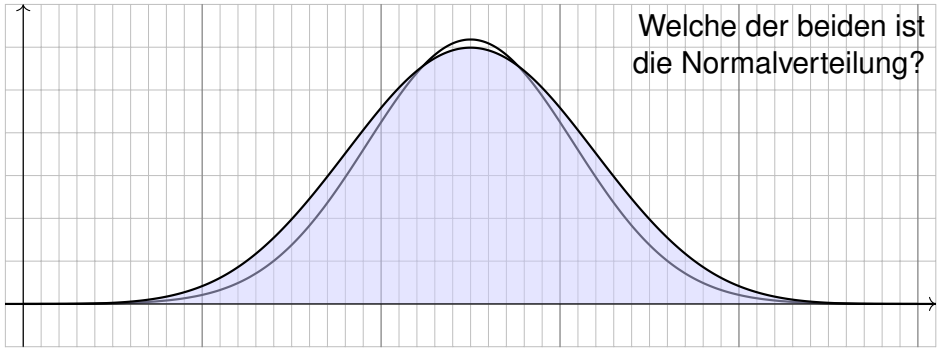
$$f_2(k+t) = \begin{cases} t & \text{für } k=0, \\ 1-t & \text{für } k=1. \end{cases}$$

$$f_3(k+t) = \begin{cases} \frac{1}{2}t^2 & \text{für } k=0, \\ \frac{1}{2}+t-t^2 & \text{für } k=1, \\ \frac{1}{2}-t+\frac{1}{2}t^2 & \text{für } k=2. \end{cases}$$

$$f_4(k+t) = \begin{cases} \frac{1}{6}t^3 & \text{für } k=0, \\ \frac{1}{6}(1+3t+3t^2-3t^3) & \text{für } k=1, \\ \frac{1}{6}(4-6t^2+3t^3) & \text{für } k=2, \\ \frac{1}{6}(1-3t+3t^2-t^3) & \text{für } k=3. \end{cases}$$

☺ Die ersten Terme $f_1, f_2, \ldots, f_5$ sind in obiger Graphik dargestellt.

Summen gleichverteilter Zufallsvariablen

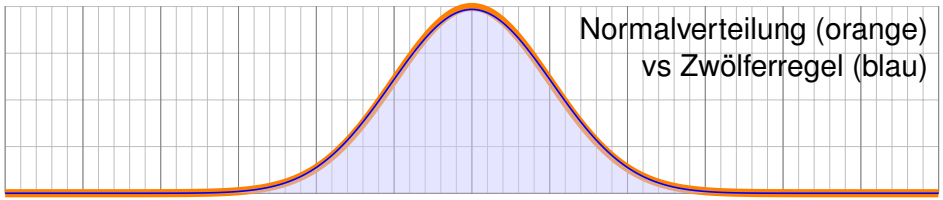


Die obige Verteilung der fünffachen Summe S_5 ähnelt augenscheinlich der Normalverteilung $\mathrm{N}(5/2, 5/12)$. Der Vergleich zeigt Abweichungen.

⚠ Nicht jede glockenähnliche Kurve ist eine Normalverteilung!
Die obigen Illustrationen zeigen eindrückliche Gegenbeispiele.

Aufgabe: Untersuchen Sie die **Zwölferregel**: „Wählt man Zufallszahlen $X_1, \ldots, X_{12}$ unabhängig und gleichverteilt im Intervall $[-1/2, 1/2]$, so ist ihre Summe $S = X_1 + \ldots + X_{12}$ näherungsweise std-normalverteilt.“

Summen gleichverteilter Zufallsvariablen



Lösung: Es gilt $\mu_k = 0$. Wir berechnen das n-te absolute Moment:

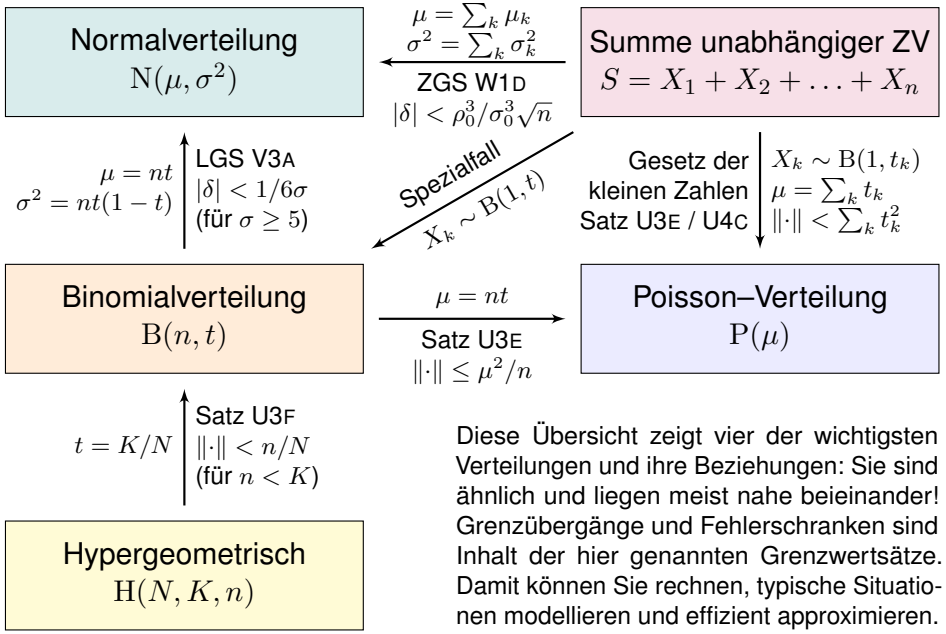
$$\int_{-1/2}^{1/2} |x|^n\,\mathrm{d}x = 2\int_{x=0}^{1/2} x^n\,\mathrm{d}x = \frac{2}{n+1}\Big[x^{n+1}\Big]_{x=0}^{1/2} = \frac{2^{-n}}{n+1}$$

Demnach gilt $\sigma_k^2 = 1/12$ und $\rho_k^3 = 1/32$. Die Summe S hat Erwartung $\mathbf{E}(S) = 0$ und Varianz $\mathbf{V}(S) = 1$. Die Näherung ist recht brauchbar, doch die pessimistische Fehlerschranke des ZGS garantiert nur $|\delta| \le 12/32$.

☺ Die **Polarmethode** ist wesentlich genauer und effizienter, siehe `en.wikipedia.org/wiki/Marsaglia_polar_method`: Sie erzeugt *zwei* exakt normalverteilte Zufallszahlen durch den Gaußschen Kunstgriff. Mathematische Sachkenntnis ermöglicht effiziente Lösungen.

Wichtige Verteilungen und Grenzwertsätze

W125



Diese Übersicht zeigt vier der wichtigsten Verteilungen und ihre Beziehungen: Sie sind ähnlich und liegen meist nahe beieinander! Grenzübergänge und Fehlerschranken sind Inhalt der hier genannten Grenzwertsätze. Damit können Sie rechnen, typische Situationen modellieren und effizient approximieren.

Wichtige Verteilungen und Grenzwertsätze

W126
Erläuterung

Viele Zufallsexperimente beschreiben wir durch **Standardmodelle**.
Hierzu nutzen wir vier besonders häufige und wichtige WVerteilungen:

- die hypergeometrische Verteilung $\mathrm{H}(N,K,n)$ für Stichproben oZ,
- die Binomialverteilung $\mathrm{B}(n,t)$ für Trefferzahlen und Stichproben mZ,
- die Poisson–Verteilung $\mathrm{P}(\mu)$ als Näherung für kleine Trefferwkt t,
- die Normalverteilung $\mathrm{N}(\mu,\sigma^2)$ als Näherung für beliebige $t \in]0,1[$.

Viele Rechnungen vereinfachen wir durch asymptotische **Näherungen**:

- Für große Stichproben ($N \to \infty$) gilt $\mathrm{H}(N,K,n) \approx \mathrm{B}(n,K/N)$.
- Gesetz der kleinen Zahlen: $\mathrm{B}(n,t) \approx \mathrm{P}(nt)$ für kleines t.
- Lokaler Grenzwertsatz: $\mathrm{B}(n,t) \approx \mathrm{N}(\mu,\sigma^2)$ für großes σ.

Insbesondere für eine **Summe S unabhängiger Zufallsvariablen** gilt:

- Als Spezialfall erhalten wir exakt die Binomialverteilung $\mathrm{B}(n,t)$.
- Gesetz der kleinen Zahlen: $\mathbf{P}_S \approx \mathrm{P}(\mu)$ für kleine t_k, siehe U4C.
- Zentraler Grenzwertsatz: $\mathbf{P}_S \approx \mathrm{N}(\mu,\sigma^2)$, siehe W1D.

🙂 Unsere Sätze garantieren bequeme, allgemeine Fehlerschranken.

Wichtige Verteilungen und Grenzwertsätze

W127
Ausführung

Dieser abschließende Rückblick ist auch ein guter Moment für eine grundsätzliche, strategische Frage: Wozu tun wir das alles? Warum geben wir uns soviel Mühe mit diesen Grenzwertsätzen? Machen heutige Computer nicht solche Näherungen überflüssig?

🙂 Tatsächlich sind Computer eine enorme Hilfe; sie können dank brutaler Rechengewalt (engl. *brute force*) viele Probleme berechnen, die vor wenigen Jahrzehnten noch außerhalb unserer Reichweite lagen.

🙂 Dazu helfen nicht nur schnellere Prozessoren, sondern mehr noch bessere mathematische Werkzeuge in Form effizienter Algorithmen. Dazu gehören insbesondere gute Näherungen und Grenzwertsätze!

Moores Regel prognostiziert, dass Computer etwa alle 18 Monate ihre Leistung verdoppeln; bislang beschreibt dies recht gut die Entwicklung.

⚠ Auch wenn diese Faustregel weiterhin zutreffen sollte, fallen selbst einfache Berechnungen wie W321 leicht so groß aus, dass sie weit außer Reichweite liegen. Das Grundproblem bleibt somit immer bestehen.

⚠ Effiziente Näherungen bleiben daher weiterhin numerisch wichtig!

Wichtige Verteilungen und Grenzwertsätze

W128
Ausführung

🙂 Die Grenzwertsätze erklären Ihnen die logischen und numerischen **Beziehungen** zwischen den verschiedenen stochastischen Modellen.

Die Struktur des vorgelegten Problems diktiert meist ein Modell, die gegebenen Daten kalibrieren es und bestimmen die Parameter. Anschließend können Sie das Modell modifizieren und vereinfachen, und innerhalb der Fehlergrenzen gegen ein einfacheres austauschen.

⚠ Diese Beziehungen helfen Ihnen, in praktischen Anwendungen das passende mathematische Werkzeug zu wählen und sicher zu rechnen.

🙂 Dasselbe gilt für die Nutzung und **Programmierung** von Computern: Sie müssen dem Computer genau sagen, *was* er zu rechnen hat, und in vielen Fällen sogar, *wie* er zu rechnen hat. Denn nur Sie wissen, welche Genauigkeit Sie am Ende verlangen und wieviel Rechenzeit Sie dafür zur Verfügung stellen. Hierzu dient die Wahl der Methode, der Näherung.

Schließlich müssen Sie computer-berechnete Ergebnisse verstehen, auf Korrektheit bzw. Plausibilität prüfen, interpretieren und anwenden.

⚠ Gerade hierzu dienen Gesetzmäßigkeiten und Rechenregeln!

Lobeshymne auf den zentralen Grenzwertsatz

W129
Ausführung

In Natur- und Ingenieurwissenschaften, in Wirtschaft und Gesellschaft treten **zufällige Größen** X auf. Sie sind meist unvermeidbar, genauere Informationen sind oft zu teuer oder manchmal gar nicht zugänglich.

Die Stochastik liefert Werkzeuge zum sicheren Umgang mit unsicheren Ereignissen, allgemein zum rationalen Entscheiden unter Unsicherheit. Hierzu ist es von zentraler Bedeutung, die **Verteilung** $\mathbf{P}_X$ zu kennen, oder zumindest so gut abzuschätzen wie es eben geht oder nötig ist.

Oft ist unsere Zufallsgröße S eine Summe zahlreicher kleiner Einflüsse, die unabhängig voneinander wirken. Der zentrale Grenzwertsatz erklärt, dass wir dann eine **Normalverteilung** erwarten dürfen, $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$.

Der Name des Satzes geht zurück auf George Pólyas Arbeit *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung* von 1920.

Wie in der Mathematik üblich, geht man zum **Grenzwert** (Limes) über, um die Ergebnisse bequem und übersichtlich formulieren zu können. Für praktische Anwendungen ist eine **Fehlerschranke** notwendig für den Wechsel von $\mathbf{P}_S$ zu $\mathrm{N}(\mu, \sigma^2)$. Satz W1D stellt hierzu alles bereit.

Lobeshymne auf den zentralen Grenzwertsatz

W130
Ausführung

🙂 Der zentrale Genzwertsatz ist ein phantastisches Ergebnis.

Er ist allgemein gültig und daher vielseitig anwendbar. Das ermöglicht die Modellierung, Berechnung und Erklärung vieler stochastischer Phänomene, die uns täglich umgeben. Implizit oder explizit wird er in praktisch allen naturwissenschaftlich-technischen Modellen oder industriellen Prozessen benutzt, in denen zufällige Schwankungen vorkommen.

🙂 Die Voraussetzungen des ZGS sind sehr allgemein und oft erfüllt. Daher ist der ZGS nahezu überall anwendbar (anders als der LGS).

Zunächst setzen wir die Unabhängigkeit der Zufallsvariablen $X_1, X_2, X_3, \dots$ voraus, was für viele Anwendungen realistisch ist. (Man kann dies weiter abschwächen, und ein kontrolliertes Maß an schwacher Abhängigkeit erlauben, aber dies würde uns hier zu weit vom Weg führen.)

Die vorausgesetzten Schranken $\sigma_k \geq \sigma_0 > 0$ und $\rho_k \leq \rho_0 < \infty$ des zentralen Grenzwertsatzes W1D stellen sicher, dass alle Beiträge $X_1, X_2, X_3, \dots$ ähnlich stark streuen. Andernfalls könnte zum Beispiel X_1 alle anderen Summanden $X_2, X_3, \dots$ überwiegen, und in der Summe würde keine Normalverteilung entstehen, sondern im Wesentlichen die Verteilung von X_1. W309

🙂 Eine allgegenwärtige Anwendung sind Konfidenzintervalle. W209

Diese werden überall in Naturwissenschaft und Technik verwendet, um die Genauigkeit einer Messung, Rechnung oder Spezifikation zu quantifizieren (Fehlerrechnung). Diese wichtige Information muss man mindestens passiv intuitiv verstehen, meist auch aktive präzise nutzen.

Lobeshymne auf den zentralen Grenzwertsatz

W131
Ausführung

🙂 Sehr wenige Informationen reichen für den ZGS bereits aus!

Für die Näherung genügen Erwartungen μ_k und Varianzen $\sigma_k^2 \geq \sigma_0^2 > 0$, für die Fehlerschranke die dritten Momente $\rho_k^3 \leq \rho_0^3 < \infty$, notfalls nach oben abgeschätzt. Das gelingt oft problemlos.

Ein besonders wichtiger Spezialfall ist die unabhängige Wiederholung von Messungen bzw. Experimenten: Hier sind die Zufallsvariablen $X_1, \dots, X_n$ unabhängig und identisch verteilt, insbesondere gilt für ihre Kenngrößen $\mu_k = \mu_0$, $\sigma_k = \sigma_0$, $\rho_k = \rho_0$ für alle $k = 1, 2, \dots, n$. Zur Existenz dieser Integrale benötigen wir lediglich $X_k \in L^3(\Omega; \mathbb{R})$, denn $L^3 \supseteq L^2 \supseteq L^1$.

Der lokale Grenzwertsatz V3A ist bei Bedarf genauer: Dort geht es nur um Binomialverteilungen, dafür liefert der LGS den nützlichen Korrekturterm κ und wesentlich schärfere Fehlerschranken. Diese Verbesserung ist auch für den ZGS möglich und führt zur Edgeworth–Entwicklung mittels höherer Momente. Der obige zentrale Grenzwertsatz ist die erste Näherung dieser Reihe.

🙂 Wir können den ZGS durch Fourier–Transformation nachrechnen.

Da wir die Fourier–Transformation aus Kapitel K bereits gut kennen, skizziere ich die Rechnung im Anhang zu diesem Kapitel, ab Seite W406. Die Idee ist einfach-genial und leicht zu verstehen!

⚠ Ohne die geforderten Momente gilt der ZGS im Allgemeinen nicht!

Wir finden hierzu einfache Gegenbeispiele, in denen der ZGS nicht anwendbar ist. W309
Die Cauchy–Verteilungen W413 zum Beispiel erfüllen nicht die Voraussetzungen des ZGS und auch nicht seine Schlussfolgerung, denn sie konvergieren nicht gegen die Normalverteilung.

Lobeshymne auf den zentralen Grenzwertsatz

W132
Ausführung

🙂 Der ZGS liefert eine explizite Fehlerschranke $|\delta| \leq \rho_0^3/\sigma_0^3\sqrt{n}$.

Der Grenzwertsatz in der Form $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$ oder besser $\mathbf{P}_S \to \mathrm{N}(\mu, \sigma^2)$ für $n \to \infty$ ist eine qualitative Aussage: Er garantiert die Konvergenz der Verteilungen, sagt aber noch nichts über die Konvergenzgeschwindigkeit oder den Approximationsfehler für ein vorgegebenes n.

Oft verwendet man den ZGS nicht für $n \to \infty$, sondern als $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$ für ein festes n. Für die konkrete Anwendung ist es dann wichtig zu wissen: Wie gut ist die Approximation? Obige Formulierung von Berry–Esséen ist eine quantitative Präzisierung: Sie besagt, wie schnell die Konvergenz ist (nämlich wie $1/\sqrt{n} \to 0$) und garantiert sogar explizite Fehlerschranken!

Der LGS bietet eine genauere Schranke; symmetrisch oder mit Korrekturterm sinkt der Fehler sogar von $|\delta| \leq \mathrm{const}/\sqrt{n}$ auf $|\varepsilon| \leq \mathrm{const}/n$. Wenn im ZGS eine ähnliche Garantie nötig sein sollte, müsste man genauere und umfangreichere Rechnungen anstrengen. Auch das kommt in sicherheitskritischen Anwendungen vor. Hierzu existiert eine ausgedehnte Spezialliteratur.

🙂 Oft ist die Näherung besser als die pessimistische Fehlerschranke!

Die im zentralen Grenzwertsatz angegebene Fehlerschranke $|\delta| \leq \rho_0^3/\sigma_0^3\sqrt{n}$ ist ein bequemer erster Schritt. In günstigen Fällen ist die Approximation tatsächlich besser, wie Beispiele zeigen. Die Fehlerschranke gilt allgemein, sie muss daher pessimistisch vom schlechtesten Fall ausgehen.

⚠ Beispiele zeigen den asymptotischen Fehler $|\delta| \sim \mathrm{const}/\sqrt{n}$. W317
Diese Fehlerschranke lässt sich also allgemein nicht weiter verbessern.

Anwendung auf Konfidenzintervalle

Anwendung: Jede Messung X ist zufälligen Störungen unterworfen, Messwerte schwanken um den Erwartungswert μ mit Streuung $\sigma > 0$.

Beispiel: Sie führen $n = 400$ unabhängige Messungen $X_1, X_2, \ldots, X_{400}$ durch und erhalten Messwerte $x_1, x_2, \ldots, x_{400}$. Sie schätzen μ durch den

Stichprobenmittelwert $\hat{x} := \frac{1}{n}\sum_{k=1}^{n} x_k \overset{\text{Messung (Beispiel)}}{=} 57.5$

Die Streuung von X schätzen Sie (dank der Stichprobe) auf $\sigma \lessapprox 10$.
Wie genau ist $\hat{x}$? Wo liegt vermutlich der wahre Erwartungswert μ?

$$I_1 = [57, 58], \quad I_2 = [56.5, 58.5], \quad I_3 = [56, 59], \quad I_k = [57.5 \pm k \cdot 0.5]$$

Aufgabe: Welches Intervall I_k überdeckt den Erwartungswert μ mit Sicherheit $\alpha_k \geq 95\%$ bzw. $\alpha_k \geq 99\%$? Was sagt Chebychev? der ZGS?

Der Erwartungswert μ ist eine Konstante. Leider ist sie uns unbekannt. Das Ziel der Messung ist, diese Konstante μ möglichst genau zu bestimmen. Als Schätzung für den Wert μ dient uns $\hat{x}$. Wie die Messwerte $x_1, \ldots, x_n$ ist auch $\hat{x}$ zufälligen Schwankungen unterworfen. Wenn Sie eine neue unabhängige Stichprobe durchführen, erhalten Sie einen neuen Mittelwert $\hat{x}$, und mit ihm verschieben sich obige Intervalle. Wir wollen sichergehen und verlangen, dass in 95% aller Stichproben unser Intervall den wahren Wert μ überdecken soll. Welchem Intervall vertrauen wir?

Anwendung auf Konfidenzintervalle

Lösung: Wir nutzen das Gesetz der großen Zahlen und den ZGS!
Der Mittelwert $\hat{X} := \frac{1}{n}(X_1 + \ldots + X_n)$ unabhängiger Messungen hat Erwartung $\mathbf{E}(\hat{X}) = \mathbf{E}(X) = \mu$ und Varianz $\mathbf{V}(\hat{X}) = \mathbf{V}(X)/n$.
Also $\sigma(\hat{X}) = \sigma(X)/\sqrt{n}$. Im Beispiel gilt $\sigma(\hat{X}) = 10/\sqrt{400} = 0.5$.
Chebychev liefert eine einfache, aber leider nur grobe Abschätzung:

$$\alpha_1 \geq 1 - 1/1^2 = 0, \qquad \alpha_2 \geq 1 - 1/2^2 \geq 75\%,$$
$$\alpha_3 \geq 1 - 1/3^2 > 88\%, \qquad \alpha_4 \geq 1 - 1/4^2 > 93\%,$$
$$\alpha_5 \geq 1 - 1/5^2 \geq 96\%, \qquad \alpha_{10} \geq 1 - 1/10^2 \geq 99\%,$$

Dank ZGS gilt annähernd $\hat{X} \sim \mathrm{N}(\mu, \sigma^2/n)$ mit besseren Schranken:

$$\alpha_1 > 68.26\% - \delta, \qquad \alpha_2 > 95.44\% - \delta,$$
$$\alpha_3 > 99.73\% - \delta, \qquad \alpha_4 > 99.993\% - \delta,$$
$$\alpha_5 > 99.99994\% - \delta, \qquad \alpha_6 > 99.9999998\% - \delta.$$

🙂 Für die Normalverteilung gilt die 68–95–99–Regel V120:
Das 2σ–Intervall genügt für 95%ige Sicherheit, und 3σ für 99%.

Anwendungsbeispiel: Meinungsforschung

Eine Wahl mit über $5\,000\,000$ Stimmberechtigten steht bevor.
Der Stimmanteil $p \in [0.02, 0.98]$ einer Partei soll geschätzt werden.
Hierzu werden n zufällige Personen befragt. Die Schätzung soll mit Wahrscheinlichkeit $\geq 95\%$ bis auf einen Fehler von ≤ 0.01 genau sein.

Aufgabe: Wie viele Personen sollten hierfür befragt werden?

Lösung: Sicherheit 95% garantieren wir mit einer 2σ–Umgebung.
Der empirische Mittelwert $\hat{X} = (X_1 + \ldots + X_n)/n$ hat die Varianz

$$\mathbf{V}(\hat{X}) = \frac{1}{n}\mathbf{V}(X) = \frac{p(1-p)}{n} \leq \frac{1}{4n}, \qquad \sigma(\hat{X}) \leq \frac{1}{2\sqrt{n}}.$$

Wir wollen eine Genauigkeit von $2\sigma(\hat{X}) \leq 0.01$. Hierzu genügt

$$2\sigma(\hat{X}) \leq \frac{1}{\sqrt{n}} \overset{!}{\leq} 0.01, \quad \text{also} \quad n \geq 10\,000.$$

🙂 Die detaillierte Rechnung haben wir bereits im letzten Kapitel V mit dem LGS und allen Fehlerabschätzungen ausgeführt. V415
Mit der Approximation dieses Kapitels wird's ein Zweizeiler.

Anwendungsbeispiel: Wareneingangsprüfung

Ihr Zulieferer schickt Ihnen $N = 6000$ Bauteile (z.B. Temperaturfühler).
Als Eingangsprüfung nehmen Sie eine Stichprobe von $n = 200$ Teilen und prüfen die vereinbarten Anforderungen; $k = 10$ Stück fallen durch.

Aufgabe: Wie viele Teile von den 6000 sind fehlerhaft? Geben Sie ein Intervall an, das den wahren Wert mit 95%iger Sicherheit überdeckt.
(Für einen analogen, aber einfacheren Hypothesentest siehe U331.)

Lösung: Wir schätzen den Anteil fehlerhafter Teile auf $t = k/n = 5\%$.
Das entspräche $K = Nt = 300$ Teilen in der gesamten Lieferung.
Wie gut ist diese Schätzung? Einzelprüfung: $\sigma^2(X) = t(1-t) \leq 0.08$.
Stichprobe: $\sigma^2(\hat{X}) \leq 0.08/200 = 0.0004$, Streuung $\sigma(\hat{X}) \leq 0.02$.
Das 2σ–Intervall um unsere Schätzung ist demnach:

$$\text{Anteil} \quad t \in [0.01, 0.09], \qquad \text{absolut} \quad K \in [60, 540]$$

Die Wareneingangsprüfung obliegt dem Käufer (§377 HGB). Hierzu gibt es normierte Methoden. Die Norm ISO2859 etwa beschreibt ein statistisches Qualitätsprüfungsverfahren für attributive Merkmale („in Ordnung“ / „nicht in Ordnung“). Sie wurde 1974 eingeführt und basiert auf dem Standard der amerikanischen Streitkräfte zur Qualitätsprüfung der anzukaufenden Ausrüstung.

Wie liest / schreibt / berechnet man Konfidenzintervalle?

Beispiel: Hilft Melatonin beim Einschlafen? Hierzu lesen wir *Effects of exogenous melatonin on sleep: a meta-analysis.* Sleep Medicine Reviews 9 (2005) 41–50. doi.org/10.1016/j.smrv.2004.06.004

> *The study effect size was taken to be the difference between the response on placebo and the mean response on melatonin for each outcome measured. Melatonin [...] reduced sleep onset latency by 4.0 min (95% CI 2.5, 5.4) [...] and increased total sleep duration by 12.8 min (95% CI 2.9, 22.8).*

Übung: Lesen Sie laut vor und übersetzen Sie in Ihre eigenen Worte. Was wurde gemessen? Wie gerechnet? Was bedeuten diese Intervalle? (Zur umsichtigen Entscheidung müssen weitere Fragen geklärt werden; mehr zu Melatonin siehe ▶ Doktor Whatson, youtu.be/oQ-o0d60NuI.)

Problem: Jede Messung X unterliegt zufälligen Störungen, die wir nicht beseitigen können. Wir betrachten sie als Zufallsvariable $X : \Omega \to \mathbb{R}$. Messwerte schwanken um den Erwartungswert μ mit Streuung $\sigma > 0$.

☹ Die Erwartung μ und die Verteilung von X sind jedoch unbekannt. Das Ziel unserer Messung ist, μ möglichst genau zu bestimmen.

Wie liest / schreibt / berechnet man Konfidenzintervalle?

Annahme: Die Messung lässt sich beliebig, unabhängig wiederholen. Wir erhalten unabhängige Zufallsvariablen $X_1, X_2, \dots$ verteilt wie X, und

$$\text{die Schätzfunktion} \quad \hat{X} := \frac{1}{n}\sum_{k=1}^{n} X_k \;:\; \Omega \to \mathbb{R}$$

☺ Der **ZGS** hilft: $\hat{X}$ ist annähernd normalverteilt, kurz $\hat{X} \sim \mathrm{N}(\mu, \sigma^2/n)$. Für große n spielt die genaue Verteilung von X gar keine Rolle mehr!
Anwendung: Aus Messwerten $x_1, \dots, x_n \in \mathbb{R}$ bilden wir

$$\text{den Stichprobenmittelwert} \quad \hat{x} := \frac{1}{n}\sum_{k=1}^{n} x_k.$$

Dieser wird meist vom Erwartungswert $\mu = \mathbf{E}(X)$ abweichen!
Als 2σ–Konfidenzintervall dieser Stichprobe bezeichnet man

$$I_2 = \big[\ \hat{x} - 2\sigma/\sqrt{n},\ \ \hat{x} + 2\sigma/\sqrt{n}\ \big].$$

Wie $\hat{X}$ ist auch das Intervall I_2 zufälligen Schwankungen unterworfen.
☺ Mit 95% Wkt überdeckt dieses Intervall den Erwartungswert μ.

Bessel–Korrektur zum erwartungstreuen Schätzer

Problem: Wie die Erwartung μ ist auch die Streuung σ unbekannt.
Lösung: Wir schätzen die Varianz σ^2 durch die (Bessel–korrigierte)

$$\text{Stichprobenvarianz} \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{x})^2.$$

⚠ Der korrigierte Nenner $n-1$ sichert Erwartungstreue: $\mathbf{E}(S^2) = \sigma^2$.
Für große n ist der Unterschied zwischen n und $n-1$ unerheblich.
Wir rechnen nach: Der Schätzer $\hat{X}$ für μ ist erwartungstreu, denn

$$\mathbf{E}\big[\hat{X}\big] \;=\; \mathbf{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] \;=\; \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\big[X_k\big] \;=\; \frac{1}{n}n\mu \;=\; \mu$$

Ist der Erwartungswert μ bekannt, so schätzen wir die Varianz σ durch

$$S_0^2 := \frac{1}{n}\sum_{k=1}^{n}\big(X_k - \mu\big)^2, \quad \text{denn in Erwartung gilt tatsächlich}$$

$$\mathbf{E}\big[S_0^2\big] \;=\; \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\big[(X_k-\mu)^2\big] \;=\; \frac{1}{n}\sum_{k=1}^{n}\mathbf{V}\big[X_k\big] \;=\; \frac{1}{n}n\sigma^2 \;=\; \sigma^2.$$

Bessel–Korrektur zum erwartungstreuen Schätzer

Ist μ unbekannt, so würde man naiv zunächst folgendes versuchen:

$$\begin{aligned} S_1^2 \;&:=\; \frac{1}{n}\sum_{k=1}^{n}(X_k - \hat{X})^2 \;=\; \frac{1}{n}\sum_{k=1}^{n}\big[(X_k-\mu) - (\hat{X}-\mu)\big]^2 \\ &=\; \left[\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right] - 2\left[\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)(\hat{X}-\mu)\right] + (\hat{X}-\mu)^2 \\ &=\; \left[\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right] - (\hat{X}-\mu)^2 \end{aligned}$$

Somit gilt $\mathbf{E}(S_1^2) = \mathbf{V}(X) - \mathbf{V}(\hat{X}) = \sigma^2 - \sigma^2/n = \sigma^2 \cdot (n-1)/n$.
Als Schätzer nimmt man daher die korrigierte Stichprobenvarianz

$$S^2 := \frac{1}{n-1}\sum_{k=1}^{n}\big(X_k - \hat{X}\big)^2.$$

Dieser Schätzer ist nun erwartungstreu, das heißt $\mathbf{E}(S^2) = \sigma^2$.
Das Gesetz der großen Zahlen T3J garantiert die Konvergenz

$$\hat{X} \to \mathbf{E}(\hat{X}) = \mu \quad \text{und} \quad S^2 \to \mathbf{E}(S^2) = \sigma^2 \quad \text{für } n \to \infty.$$

Was sind und was bedeuten Konfidenzintervalle?

Satz W2A: Konfidenzintervalle

Als Stichprobe für X führen wir n **unabhängige Messungen** aus.
Aus den so gewonnenen Messwerten $x_1, \dots, x_n \in \mathbb{R}$ berechnen wir

den **Stichprobenmittelwert** $\quad \hat{x} := \frac{1}{n} \sum_{k=1}^{n} x_k,$

die **Stichprobenvarianz** $\quad \hat{\sigma}^2 := \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \hat{x})^2,$

das 2σ**–Konfidenzintervall** $\quad I_2 := \left[\hat{x} - 2\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{x} + 2\frac{\hat{\sigma}}{\sqrt{n}} \right].$

Für große n gilt: Bei 95% aller Stichproben überdeckt das Intervall I_2 den (konstanten, aber uns unbekannten) Erwartungswert $\mu = \mathbf{E}(X)$.

🙂 Ein wunderbar praktischer Satz! Was genau bedeutet hier „große n“? Als Faustregel wird oft $n \geq 30$ gefordert. Satz V4A: Bei Normalverteilung genügt $n \geq 60$. Bei ungünstiger Verteilung von X ist größeres n nötig.

Was sind und was bedeuten Konfidenzintervalle?

🙂 Das 2σ–Intervall liefert 95% Sicherheit, das 3σ–Intervall über 99%. Bei sicherheitskritischen Anwendungen wird man noch mehr verlangen. Als ein extremes Sicherheitsniveau wurde etwa 6σ propagiert, was bei exakter Normalverteilung über 99.9999998% Sicherheit bedeutet. W201

Kosten-Nutzen-Abwägung: Kleines n ist billig, großes n ist präzise: Mit wachsendem n wird der Intervallradius $2\hat{\sigma}/\sqrt{n}$ beliebig klein. Der Nenner $\sqrt{n}$ bedeutet: Für doppelte Genauigkeit braucht man eine viermal größere Stichprobe, für zehnfache Genauigkeit eine 100mal größere Stichprobe. Das ist Fluch und Segen der großen Zahlen!

⚠ Sicherheit $\geq 95\%$ für das 2σ–Intervall gilt für „hinreichend große n“. Solche Faustregeln ganz ohne Begründung sind Autoritätsargumente; manchmal beruhen sie auf etwas Erfahrung, meist jedoch auf Unwissen. Kleine Stichproben erfordern Korrekturen. Den Fall $n = 2$ skizziere ich ab V420. Allgemein antwortet der Satz V4A von Gosset alias „Student“.

Ich belasse es bei diesem Überblick. Für die ernsthafte Anwendung statistischer Methoden verweise ich auf die umfangreiche Literatur.

Was sind und was bedeuten Konfidenzintervalle?

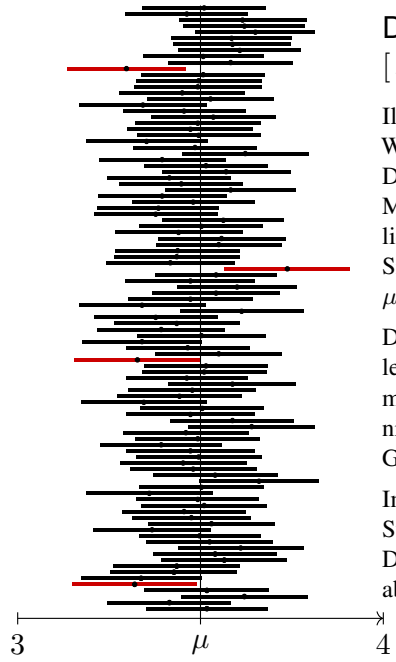
⚠ Ich warne vorsorglich vor einem weitverbreiteten **Missverständnis**: Man sagt „Mit Wkt $\geq 95\%$ überdeckt das Intervall I den wahren Wert μ.“ Das ist bequem, aber irreführend; die Interpretation müssen wir klären.

Der Mittelwert $\mu = \mathbf{E}(X) \in \mathbb{R}$ ist eine **Zahl**, unbekannt aber konstant. Aus den Messdaten $x_1, \dots, x_n \in \mathbb{R}$ berechnen wir ein Intervall $I \subseteq \mathbb{R}$. Entweder liegt μ in I oder nicht: Das ist eine Aussage über eine feste Zahl μ und ein festes Intervall I, mit Wkt hat das nichts mehr zu tun.

⚠ Die Aussage „mit Wkt $\geq 95\%$“ betrifft genau genommen also nicht ein einzelnes berechnetes Intervall, sondern das **Verfahren** insgesamt: Bei unabhängiger Wiederholung können wir jedesmal dieses Verfahren anwenden, und in $\geq 95\%$ aller Fälle werden wir damit richtig liegen.

⚠ Das Verfahren ist einfach, doch seine Interpretation erfordert Übung! Am besten spielen Sie es gegeneinander: `eiserm.de/lehre/Konfidenz`
Zur Illustration gebe ich zwei Beispiele: zunächst ein etwas künstliches, aber besonders einfaches zum Würfeln, sodann ein sehr realistisches, aber auch komplizierteres mit echten Daten zum Venus-Express.

Was sind und was bedeuten Konfidenzintervalle?



Die Graphik zeigt 100 Konfidenzintervalle $[\hat{x} \pm 2\hat{\sigma}/\sqrt{n}]$, Sicherheitsniveau $\geq 95\%$.

Illustration: Wir werfen 400 mal einen fairen Würfel. Wir wissen $\mu = 3.5$ und $\sigma = \sqrt{35/12} \approx 1.70783$. Die Messwerte schwanken, und mit ihnen Stichproben-Mittelwert $\hat{x} \approx \mu$ und -Streuung $\hat{\sigma} \approx \sigma$. Jede Stichprobe liefert daher ihr eigenes Konfidenzintervall. Bei unserer Simulation überdecken 96 Intervalle den Erwartungswert $\mu = 3.5$, die übrigen 4 Intervalle (in rot) tun dies nicht.

Dieses Beispiel ist etwas künstlich, dafür aber besonders leicht zu verstehen. In realistischen Anwendungen kennt man den wahren Mittelwert μ und die wahre Streuung σ nicht, sondern will / muss sie durch Stichprobe schätzen. Genau dieses Verfahren wenden wir hier 100 mal an.

Im Allgemeinen weiß man nicht, ob man eine typische Stichprobe erwischt (schwarz) oder viele Ausreißer (rot). Das Verfahren kann keine absolute Sicherheit garantieren, aber immerhin 95%. Die 3σ–Umgebung liefert 99%, etc.

Ausprobieren: `eiserm.de/lehre/Konfidenz`

Stochastisches Modell des Messprozesses

Das Ergebnis einer Messung X hängt ab vom wahren Wert x (dieser ist meist unbekannt und wird gesucht) und einem Fehler F_X als Störung:

$$X = x + F_X \quad \text{etwa mit} \quad F_X \sim \mathrm{N}(\mu_X, \sigma_X^2)$$

In der Praxis nehmen wir zur Vereinfachung meist an, dass der Fehler F_X die Summe von vielen unabhängigen Fehlerquellen ist. Der zentrale Grenzwertsatz W1D sagt dann für F_X eine Normalverteilung voraus.

Der Fehler F_X wird beschrieben durch zwei Kenngrößen:

1. einem systematischen (konstanten) Fehler μ_X,
2. einem stochastischen Fehler mit Streuung σ_X.

🙂 Für normalverteilte Fehler genügen bereits diese zwei Parameter.

Wiederholte Messung ergibt immer denselben systematischen Fehler. Mittelwertbildung reduziert nur den stochastischen Fehler, siehe oben!

Systematische Fehler entstehen durch falsche Justierung oder Nutzung der Messinstrumente, oder falsche Theorie, siehe etwa GPS (W217).

Idealfall: Nach Behebung aller systematischen Fehler gilt $\mu_X = 0$.

Fehlerfortpflanzung: Vereinfachung durch Linearisierung

Die Messwerte X, Y verarbeiten wir zu einer Zielgröße $Z = f(X, Y)$.
Im einfachsten Falle ist die hier genutzte Funktion f affin-linear:

$$\begin{aligned} f(X,Y) &= f(x,y) + c_1 \cdot \underbrace{(X-x)}_{=F_X} + c_2 \cdot \underbrace{(Y-y)}_{=F_Y} \\ z &= f(x,y) \\ \mu_Z &= c_1\mu_X + c_2\mu_Y \\ \sigma_Z^2 &= c_1^2\sigma_X^2 + c_2^2\sigma_Y^2 + 2c_1c_2\mathbf{Cov}(X,Y) \end{aligned}$$

Wir berechnen so $Z = z + F_Z$ als den wahren Wert $z = f(x, y)$ plus den Fehler $F_Z \sim \mathrm{N}(\mu_Z, \sigma_Z^2)$. Der systematische Fehler pflanzt sich linear fort. Bei unabhängigen stochastischen Fehlern addieren sich die Varianzen.

Im allgemeinen Falle linearisieren wir um (x, y) nach Taylor:

$$\begin{aligned} f(X,Y) &\approx f(x,y) + \underbrace{\partial_1 f(x,y)}_{=c_1} \cdot \underbrace{(X-x)}_{=F_X} + \underbrace{\partial_2 f(x,y)}_{=c_2} \cdot \underbrace{(Y-y)}_{=F_Y} \\ z &= f(x,y) \\ \mu_Z &\approx \partial_1 f(x,y)\,\mu_X + \partial_2 f(x,y)\,\mu_Y \qquad \text{(systematisch)} \\ \sigma_Z^2 &\approx [\partial_1 f(x,y)]^2\,\sigma_X^2 + [\partial_2 f(x,y)]^2\,\sigma_Y^2 \qquad \text{(stochastisch)} \end{aligned}$$

Fehlerfortpflanzung am Beispiel des Ohmschen Gesetzes

Aufgabe: Wir betrachten das Ohmsche Gesetz $R = U/I$. Wie pflanzen sich Messfehler von U und I auf R fort? systematisch? stochastisch?

Lösung: Für $R = U/I$ gilt $\partial R/\partial U = 1/I$ und $\partial R/\partial I = -U/I^2$.
Für hinreichend **kleine Fehler** gilt demnach näherungsweise:

$$\text{systematischer Fehler} \qquad \mu_R \approx \frac{1}{I}\,\mu_U - \frac{U}{I^2}\,\mu_I$$

$$\text{stochastischer Fehler} \qquad \sigma_R^2 \approx \frac{1}{I^2}\,\sigma_U^2 + \frac{U^2}{I^4}\,\sigma_I^2$$

Nach Behebung aller systematischen Fehler gilt $\mu_U = 0$ und $\mu_I = 0$.

Wir messen die Spannung $U = \bar{U} \pm \sigma_U$ und den Strom $I = \bar{I} \pm \sigma_I$ als Mittelwerte $\bar{U}$ und $\bar{I}$ mit stochastischen Fehlern $\pm\sigma_U$ und $\pm\sigma_I$.

Daraus berechnen wir den Widerstand $R = \bar{R} \pm \sigma_R$ als den Mittelwert $\bar{R} = \bar{U}/\bar{I}$ mit stochastischem Fehler σ_R gemäß der Fehlerfortpflanzung

$$\sigma_R = \sqrt{(1/\bar{I}^2)\,\sigma_U^2 + (\bar{U}^2/\bar{I}^4)\,\sigma_I^2}.$$

🙂 Dank Unabhängigkeit heben sich stochastische Fehler teilweise auf.

Fehlerfortpflanzung und gefährliche Auslöschung

Erste Annahme: Die Fehler F_U und F_I sind stochastisch unabhängig. In diesem Falle addieren sich die Varianzen, nicht die Streuungen! Das ist anschaulich plausibel: Dank Unabhängigkeit heben sich stochastische Fehler teilweise auf. Wir nutzen dies im Gesetz der großen Zahlen T3J.

Zweite Annahme: Die Fehler F_U und F_I sind klein genug, sodass wir die Linearisierung als Näherung nutzen können. Genau genommen müssten wir den Satz B4C von Taylor nutzen und das Restglied abschätzen. Das wird in der hastigen Praxis meist ignoriert. Wird schon gut gehen?

Im folgenden Beispiel ist die Rechnung tatsächlich linear, auch ohne Taylor–Approximation.

Aufgabe: Sie wollen Ihre Hauskatze wiegen ($\approx 4\mathrm{kg}$). Dazu haben Sie eine Personenwaage ($\pm 1\%$), auf der Ihre Katze aber nicht stehenbleibt. Sie halten daher Ihre Katze auf dem Arm und wiegen sich gemeinsam (Messwert X), davon substrahieren Sie Ihr eigenes Gewicht ($Y \approx 75\mathrm{kg}$). Wie genau können Sie damit das Gewicht Ihrer Katze bestimmen?

Lösung: Wir berechnen $Z = X - Y$ und erwarten einen Wert $z < 5\mathrm{kg}$. Dabei gilt $X = x \pm \sigma_X$ mit $\sigma_X \approx 790\mathrm{g}$ und $Y = y \pm \sigma_Y$ mit $\sigma_Y \approx 750\mathrm{g}$.

Die Differenz $Z = X - Y$ hat den Mittelwert $z = x - z$ und die Streuung $\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2} \approx 1.1\mathrm{kg}$. Der relative Fehler ist demnach über 25%.

🙂 Wiederholung und Mittelwert reduzieren den stochastischen Fehler!

Genauigkeit und Messfehler bei der Satellitennavigation

W217
Erläuterung

Einsteins Relativität wird verehrt als Inbegriff einer genialen Theorie, sie genießt Kultstatus in der Popkultur. Wie jede mathematische Theorie gilt sie vielen jedoch als nutzlos-abstrakt und fern alltäglichen Lebens.

Das Gegenteil ist der Fall: Wir alle nutzen die Relativitätstheorie täglich! Sie ist nicht nur Grundlage der modernen Astronomie, von schwarzen Löchern bis Gravitationswellen, sondern mitten im Alltag angekommen, insbesondere in den Milliardenprojekten rund um Satellitennavigation.

Flugzeuge, Schiffe, Autos, Mobiltelephone nutzen Satellitennavigation, genauer das **Global Navigation Satellite System** (GNSS) und speziell das US-amerikanische **NAVSTAR Global Positioning System** (GPS).

Relativität ist dabei nicht bloß eine unwichtige Spitzfindigkeit, sondern ihr genaues Verständnis und ihre praktische Anwendung sind essentiell für das Funktionieren der Navigation. Ohne präzise Theorie geht es nicht!

Die GPS-Konfiguration besteht aus (mindestens) 24 Satelliten, sodass von jedem Punkt der Erde zu jeder Zeit mindestens vier sichtbar sind. Jeder Satellit S_i sendet ständig seine aktuelle Position und Systemzeit. Aus der Signallaufzeit berechnet der Empfänger die Entfernung r_i zu S_i.

Genauigkeit und Messfehler bei der Satellitennavigation

W218
Erläuterung

Die Entfernung r_1 zu einem ersten Satelliten S_1 genügt noch nicht: Der Empfänger findet sich irgendwo auf der Sphäre um S_1 mit Radius r_1. Die Entfernung r_2 zu einem zweiten Satelliten S_2 schränkt weiter ein: Schnittmenge zweier Sphären (in allgemeiner Lage) ist eine Kreislinie. Die Entfernung r_3 zu einem dritten Satelliten S_3 lässt als Schnittmenge nur noch zwei Punkte. Eine der beiden Lösungen liegt weit über der Erde, somit ist die andere als richtige erkennbar. Diese quadratischen Gleichungen werden in kartesischen Koordinaten gelöst und geozentrisch in Kugelkoordinaten (C237) umgerechnet zu geographischer Länge φ, Breite ϑ, Höhe $h = r - r_E$.

Wir benötigen Präzision! Jeder Satellit trägt eine Atomuhr mit einer wöchentlichen Genauigkeit von $\theta \approx 1\,\mathrm{ns} = 10^{-9}\,\mathrm{s}$. Im Vergleich zur Lichtgeschwindigkeit $c \approx 3 \cdot 10^8\,\mathrm{m/s}$ entspricht das der räumlichen Entfernung $c\,\theta = 30\,\mathrm{cm}$. Handelsübliche mobile GPS-Empfänger haben jedoch keine langzeitstabile Atomuhr, sondern nur eine lokale Messgenauigkeit um $10\,\mathrm{ns}$, also etwa $3\,\mathrm{m}$. Hinzu kommen zufällige Störungen und Messfehler. Die erreichte Genauigkeit ist bemerkenswert und liegt bei etwa $10\,\mathrm{m}$! Im Auto genügt dies gerade noch für Position und Geschwindigkeit.

Räumlich genügen die Abstände zu drei Satelliten. GPS-Empfänger führen keine ausreichend genaue absolute Uhrzeit, um die Laufzeiten präzise genug zu messen. Deshalb wird zusätzlich das Signal eines vierten Satelliten genutzt, und so die obige Rechnung statt im euklidischen $\mathbb{R}^3$ in der vierdimensionalen Minkowski–Raumzeit $\mathbb{R}^{3+1}$ durchgeführt. Das Prinzip ist dasselbe.

📖 Siehe G. Strang, K. Borre: *Linear Algebra, Geodesy, and GPS* (1997) sowie für offizielle technische Spezifikationen `www.navcen.uscg.gov/pubs/gps/sigspec/gpssps1.pdf`. Damit können Sie die konkreten Daten einsetzen und selbständig alle Rechnungen prüfen. Die Mühe lohnt sich, denn GPS ist ein mathematisch-physikalisch-technisches Meisterwerk.

Genauigkeit und Messfehler bei der Satellitennavigation

W219
Erläuterung

Um die nötige extreme Präzision zu gewährleisten, müssen vermeidbare Fehler als solche erkannt und möglichst behoben werden. Dazu zählen insbesondere relativistische Effekte, diese sind hier deutlich spürbar!

Ein Objekt mit Geschwindigkeit v im Erdgravitationsfeld GM/r hat eine Eigenzeit $\Delta\tau$ im Vergleich zur Referenzzeit Δt weit außerhalb der Erde:

$$\frac{\Delta\tau}{\Delta t} \approx 1 - \left(\frac{v^2}{2c^2} + \frac{GM}{rc^2}\right)$$

Im Vergleich zwischen Satellit S und Empfänger E erhalten wir somit:

$$\frac{\Delta\tau_S}{\Delta\tau_E} \approx 1 - \left(\frac{v_S^2}{2c^2} + \frac{GM}{r_S\,c^2}\right) + \left(\frac{v_E^2}{2c^2} + \frac{GM}{r_E\,c^2}\right)$$

Für GPS-Satelliten gilt $r_S \approx 26580\,\mathrm{km}$ und $v_S \approx 3872\,\mathrm{m/s}$. Am Äquator haben wir $r_E \approx 6380\,\mathrm{km}$ und $v_E \approx 464\,\mathrm{m/s}$. Die Korrekturen sind also:

$$\text{SRT:}\quad \frac{v_E^2}{2c^2} - \frac{v_S^2}{2c^2} \approx -8.3 \cdot 10^{-11} \approx -7.2\,\mu\mathrm{s/day}$$

$$\text{GRT:}\quad \frac{GM}{r_E c^2} - \frac{GM}{r_S c^2} \approx 5.31 \cdot 10^{-10} \approx 45.9\,\mu\mathrm{s/day}$$

Genauigkeit und Messfehler bei der Satellitennavigation

W220
Erläuterung

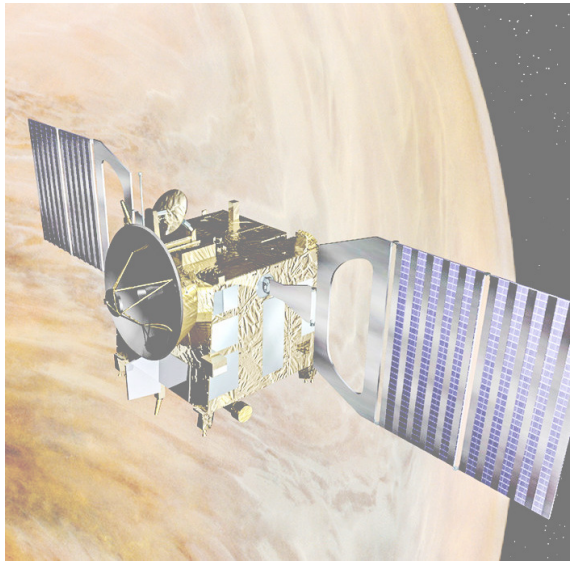
Aufgrund spezieller Relativität (SRT) gehen Satellitenuhren um $7200\,\mathrm{ns}$ pro Tag nach, aufgrund allgemeiner Relativität (GRT) um $45900\,\mathrm{ns}$ vor, in der Summe also $38600\,\mathrm{ns}$ vor. Wir benötigen aber eine Genauigkeit im Bereich von wenigen Nanosekunden! Dazu wird ein genial-einfacher Trick angewendet: Die Satellitenuhren werden nicht auf ihre nominelle Frequenz $10.23\,\mathrm{Mhz}$ geeicht, sondern auf $10.229999995453\,\mathrm{Mhz}$. Das erledigt die relativistisch nötige Korrektur! Von der Erde aus gesehen ticken die Satelliten mit der gewünschten Frequenz von $10.23\,\mathrm{Mhz}$.

Die Physik arbeitet seit über hundert Jahren mit Einsteins spezieller Relativitätstheorie (1905) und seiner allgemeinen Relativitätstheorie (1915). Beide sind in vielen Situationen mit großer Genauigkeit experimentell überprüft worden, zum Beispiel durch astronomische Beobachtungen, beim Müon-Zerfall in der Erdatmosphäre, in Teilchenbeschleunigern, und vielen mehr. Die sehr genauen Zeitmessungen des GPSystems sind eine weitere beeindruckende Illustration. Hier wirkt die Theorie auch umgekehrt auf ingenieurstechnische Fragen der Konstruktion: Um die nötige Zeitgenauigkeit zu garantieren, wurden die Satelliten sorgfältig relativistisch korrekt konstruiert. Damit ist Einsteins Relativitätstheorie innerhalb eines Jahrhunderts alltagsrelevant geworden.

Fake science? In Diskussionsforen tobt die Debatte, ob die relativistischen Korrekturen für das Funktionieren des GPSystems *absolut notwendig* sind, oder ob es auch irgendwie anders ginge: Speziell für die Positionsbestimmung könnte der Empfänger die Absolutzeit überflüssig machen, indem er nur Laufzeit*differenzen* misst und verrechnet. Das ist eine alternative Herangehensweise, zur Ausführung siehe etwa `www.alternativephysics.org/book/GPSmythology.htm`. Klar ist und bleibt dabei: Die Zeitdilatation ist real, sie wird im GPSystem tatsächlich gemessen, und ihre Korrektur ist zur *globalen* Kalibrierung der GPSystemzeit notwendig. *Science works!*

Mars & Venus Express

W221



Missionen der ESA

Start Jun. 2003 in Baikonur
Mars-Orbit ab Jan. 2004
→ Suche nach Wasser

Start Nov. 2005 in Baikonur
Venus-Orbit ab Apr. 2006
→ Atmosphäre der Venus

Orbiter: Masse 633kg leer
plus Treibstoff (MMH+NTO)

Acht Steuertriebwerke mit
je 10N Schub (im Labor)

Fortsetzung oder Ende:
Wie lange reicht der Sprit?

Siehe `en.wikipedia.org/wiki/Mars_Express` und `/Venus_Express`

Mars & Venus Express

W222
Ausführung

Design erfordert Entscheidungen und meist Kompromisse. Die ESA hat auf Messgeräte für die Tankfüllung verzichtet. Zum geplanten Ende der Mission stellt sich die Frage: Kann die Mission verlängert werden? Ist dazu noch genug Treibstoff im Tank? Wie kann man das mit ausreichender Sicherheit herausfinden? Idee: Die ESA verfügt über alle bisherigen Positions- und Steuerdaten. Hieraus könnte man jeweils die Trägheit berechnen und indirekt die Gesamtmasse! Geht das?

Realistische Analogie: Heutige Autos haben eine Masse von etwa 1.2 bis 1.6 Tonnen bei einem Tankvolumen von 50 bis 70 Litern. Sie fahren ein Auto ohne Tankanzeige, die Tankfüllung erspüren Sie in Kurven, beim Gasgeben und Abbremsen. Ist das verrückt? Ja. Ist das möglich? Schwierig! Man muss genau und sehr häufig messen – und dann die Messfehler rausfiltern!

Vereinfachte Analogie: Sie haben eine Gasflasche für einen Herd oder Grill: Leergewicht 6kg, Inhalt 0-5kg Propan, ebenfalls ohne Anzeige des Füllstandes. Sie können durch Schütteln recht gut erspüren, wie voll die Gasflasche ist. Probieren Sie es bei Gelegenheit mal aus!

Jede Messung ist mit Fehlern behaftet: systematische müssen wir erkennen und dann korrigieren, bei zufälligen können wir dies nicht direkt! Wir müssen lernen, sie geschickt rauszufiltern, um ein möglichst verlässliches Ergebnis zu erhalten. Das ist das Ziel der mathematischen Statistik.

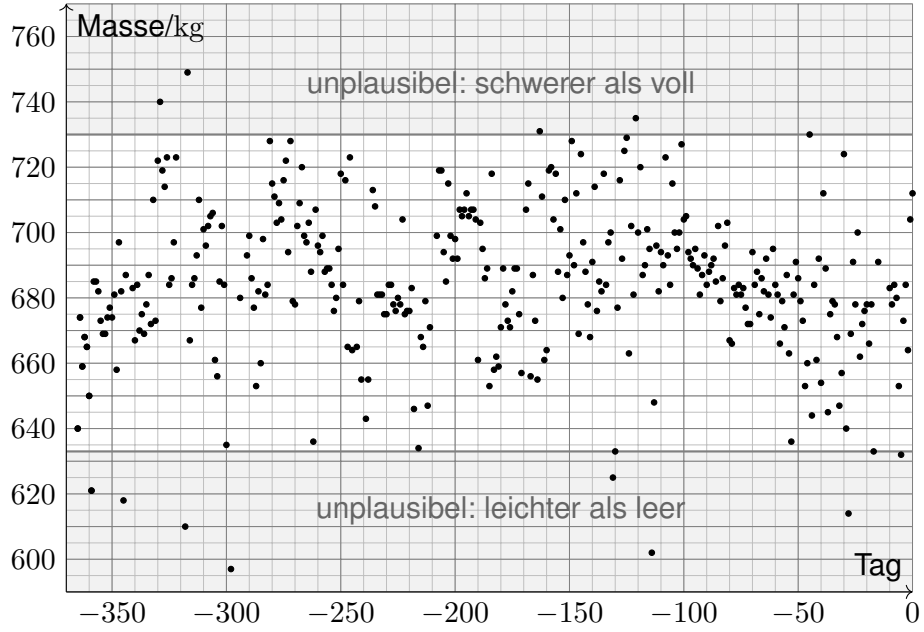
Es geht um rationale Entscheidungen unter Unsicherheit. Dafür genügt es nicht, eine willkürliche Schätzung auszuspucken! Die Anwendung ist ernst, es geht um Geld, wir müssen überzeugen. Wir wollen nicht nur eine gute Schätzung, sondern auch die Güte der Schätzung berechnen!

Hierzu will ich Sie mit meiner Einführung anleiten. Wer darüber hinaus ernsthafte statistische Analysen betreiben will/muss, kann sich darauf aufbauend in die Spezialliteratur einarbeiten.

Venus Express: Wieviel Treibstoff ist noch im Tank?

W223

Aus Steuermanövern errechnete Masse für 366 Tage bis 31.12.2012.



Venus Express: Wieviel Treibstoff ist noch im Tank?

W224
Ausführung

Impuls-Messungs-Methode (Impulse Measurement Method, IMM):
Aus Positions- und Steuerdaten errechnet man Werte für die Masse;
folgende Daten finden Sie unter `eiserm.de/lehre/HM3/VEX2012.txt`.

Tag	Datum	Masse
1	01.01.2012	640
2	02.01.2012	674
3	03.01.2012	659
...	...	...

Tag	Datum	Masse
...	...	...
364	29.12.2012	664
365	30.12.2012	704
366	31.12.2012	712

Datengewinnung ist mühsam, erfordert Umsicht und Sachkenntnis! Diese Daten stellt uns freundlicherweise Herr Caglayan Gürbüz zur Verfügung. Er hat 2012 die Höhere Mathematik 3 gehört und am IRS / ESOC seine Bachelor-Arbeit zum Mars & Venus Express geschrieben.

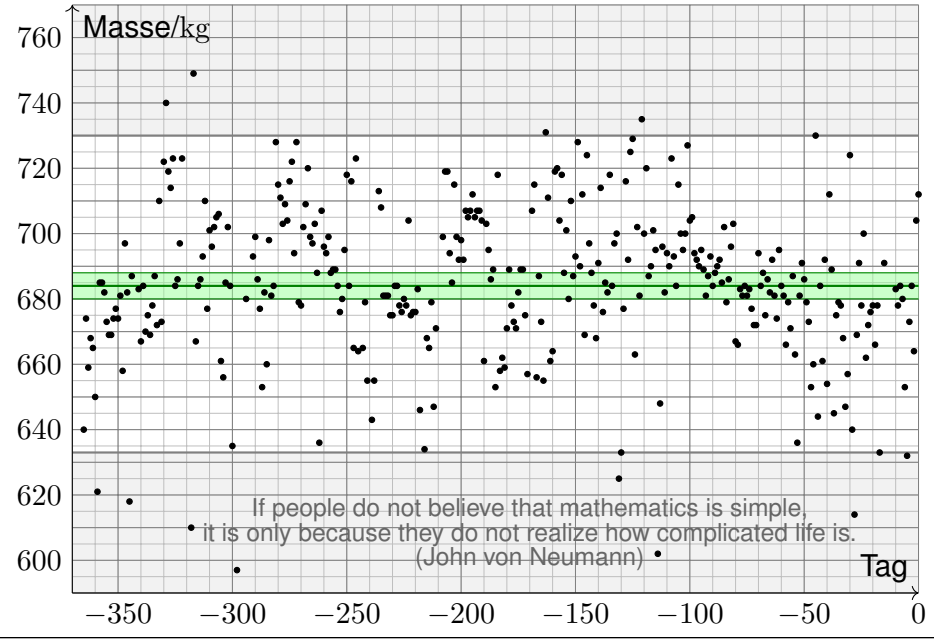
Rationale Entscheidung: Kann / sollte man die Mission verlängern? Grundlose Schätzungen sind sinnlos, wir brauchen Sicherheit!

Techniken: Erwartung und Streuung, Stichproben und Schätzung, Gesetz der großen Zahlen, zentraler Grenzwertsatz, Konfidenzintervall

🙂 *This is rocket science.* Mathematische Statistik zeigt, wie es geht.

Mathematische Statistik: Konfidenzintervalle

Jahresmittelwert der Gesamtmasse mit 3σ–Konfidenzintervall.



Mathematische Statistik: Konfidenzintervalle

Aufgabe: Wie groß war der Tankinhalt im Jahresmittel 2012?
Geben Sie das 3σ–Konfidenzintervall an (für 99%ige Sicherheit).

Lösung: Unser Datensatz liefert Mittelwert und Streuung:
Der Stichprobenmittelwert ist $\hat{m} = 684\text{kg}$. Dies schätzt die Masse m.
Die Stichprobenstreuung ist $\hat{\sigma}_m = 24.6\text{kg}$. Dies schätzt die Streuung.

Bei $n = 345$ Datenpunkten gilt $\hat{\sigma}_{\hat{m}} = \hat{\sigma}_m/\sqrt{n} = 1.33\text{kg}$. Großes n hilft!
Das 3σ–Intervall $[\hat{m} \pm 3\hat{\sigma}_{\hat{m}}] = [680; 688]\text{kg}$ ist erstaunlich schmal!
Der Tankinhalt betrug demnach $[47; 55]\text{kg}$ mit 99%iger Sicherheit.

☹ Messdaten sind voller Messfehler und anderer Ungenauigkeiten.

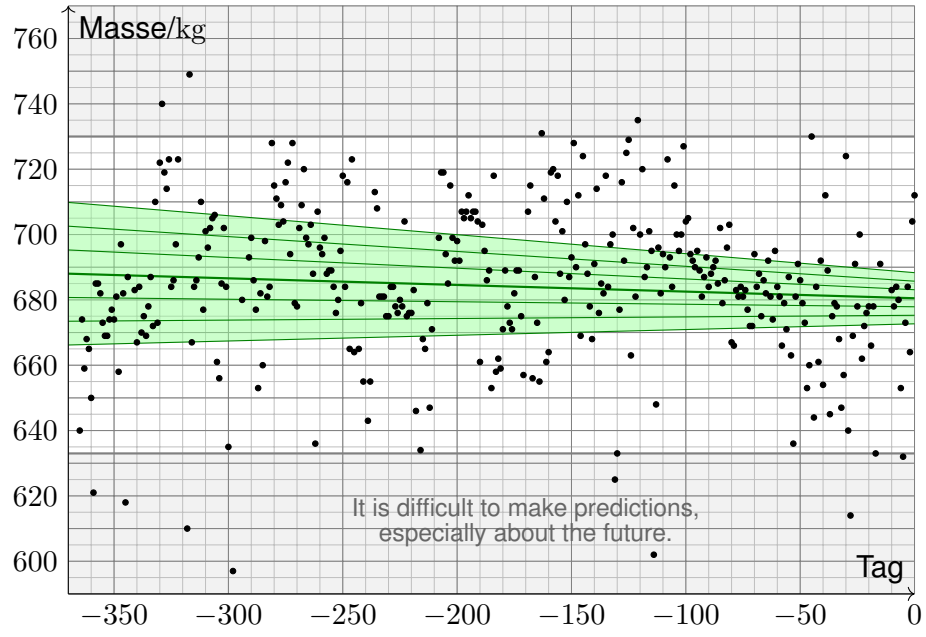
🙂 Unsere Daten sind schmutzig doch glücklicherweise sehr zahlreich:
Wir können Gesetze der großen Zahlen und Grenzwertsätze nutzen!

🙂 Dank passender Technik können Sie so präzise Schlüsse ziehen!
Dank Statistik fördern Sie Information zutage, die zuvor verborgen war.

⚠ Jede Prognose ist höchstens so gut wie die zugrundeliegenden Daten! Zufällige Fehler können wir durch große Messreihen rausfiltern, systematische Fehler nicht! Hier helfen nur Umsicht und Sachkenntnis!

Mathematische Statistik: lineare Regression

Regressionsgerade mit Konfidenzintervallen: 1σ, 2σ, 3σ.



Mathematische Statistik: lineare Regression

Aufgabe: Wie groß war der Tankinhalt Ende 2012?
Wie lange reicht es noch? Wie sicher ist Ihre Prognose?

Lösung: Unser Datensatz liefert die Regressionsgerade $y = \hat{a} + \hat{b}\,x$:
Ende 2012 ist die Gesamtmasse $\hat{a} = 680.5\text{kg}$ mit Streuung $\hat{\sigma}_{\hat{a}} = 2.6\text{kg}$.
Im Tank verbleiben $\hat{t} = 47.5\text{kg}$ Treibstoff mit Streuung $\hat{\sigma}_{\hat{t}} = 2.6\text{kg}$.
Der Verbrauch ist $\hat{b} = -20.3\text{g/Tag}$ mit Streuung $\hat{\sigma}_{\hat{b}} = 12.6\text{g/Tag}$.

Sehr grobe Prognose der verbleibenden Missionsdauer: 2300 Tage.
Wie verlässlich ist diese Prognose aufgrund unserer Messdaten?

Mit 95% Sicherheit reicht der Treibstoff noch für über 930 Tage:
Tank $[\hat{t} \pm 2\hat{\sigma}_{\hat{t}}] \subseteq [42.3; 52.7]\text{kg}$, Verbrauch $[\hat{b} \pm 2\hat{\sigma}_{\hat{b}}] \subseteq [-45.5; 4.9]\text{g/Tag}$.

Mit 99% Sicherheit reicht der Treibstoff noch für über 680 Tage:
Tank $[\hat{t} \pm 3\hat{\sigma}_{\hat{t}}] \subseteq [39.7; 55.3]\text{kg}$, Verbrauch $[\hat{b} \pm 3\hat{\sigma}_{\hat{b}}] \subseteq [-58.1; 17.5]\text{g/Tag}$.

🙂 Wir extrahieren sorgsam, was die vorliegenden Daten hergeben!
⚠ Systematische Fehler? physikalische Annahmen? Alternativen?
War 2012 ein typisches Jahr? Wenn man Sprit spart, reicht er länger!
VEX wird 2013/14 tiefer in Atmosphäre abgesenkt, der Verbrauch steigt.
Die Mission endet im Dezember 2014 wegen Treibstoffmangels. Wow!

Mathematische Statistik: Konfidenzintervalle

Das Ergebnis X einer Messung hängt von zufälligen Störungen F ab:

$$X = x + F \quad \text{etwa mit} \quad F \sim \mathrm{N}(0, \sigma_F^2)$$

Das heißt: Die Messung $X = x + F$ schwankt um den Wert $x = \mathbf{E}(X)$, der Fehler $F = X - x$ hat Erwartung $\mathbf{E}(F) = 0$ und Varianz $\mathbf{V}(F) = \sigma_F^2$.

Problem: Erwartung $\mathbf{E}(X) = x$ und Varianz $\mathbf{V}(X) = \sigma_F^2$ sind uns nicht bekannt. Wir wollen sie durch wiederholte Messungen ermitteln.

Unabhängige Wiederholung ergibt die Messwerte $x_1, x_2, \ldots, x_n$.
Erwartung $\mu_X = x$ und Varianz $\sigma_X^2 = \sigma_F^2$ schätzen wir hieraus wie folgt:

$$x \approx \hat{x} := \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \sigma_X^2 \approx \hat{\sigma}_X^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x})^2, \quad \sigma_{\hat{X}} \approx \hat{\sigma}_{\hat{X}} := \frac{\hat{\sigma}_X}{\sqrt{n}}$$

Hieraus erhalten wir Konfidenzintervalle für den gesuchten Wert x:

$$I_2 = [\hat{x} \pm 2\hat{\sigma}_{\hat{X}}] = \left[\, \hat{x} - 2\hat{\sigma}_{\hat{X}} \;;\; \hat{x} + 2\hat{\sigma}_{\hat{X}} \,\right] \quad \ni x \text{ mit } 95\% \text{ Sicherheit}$$
$$I_3 = [\hat{x} \pm 3\hat{\sigma}_{\hat{X}}] = \left[\, \hat{x} - 3\hat{\sigma}_{\hat{X}} \;;\; \hat{x} + 3\hat{\sigma}_{\hat{X}} \,\right] \quad \ni x \text{ mit } 99\% \text{ Sicherheit}$$

Mathematische Statistik: Konfidenzintervalle

Gesucht ist der wahre Wert x, doch dieser ist uns leider unbekannt.
Wir schätzen ihn durch $\hat{x}$ anhand der Messdaten $(x_1, x_2, \ldots, x_n)$.

Die Güte dieser Schätzung wird bestimmt durch die Varianz σ_X^2.
Wir schätzen sie wiederum durch die Stichprobenvarianz $\hat{\sigma}_X^2$.

Grundlegende Annahmen unseres **stochastischen Modells**:
Der Messfehler F ist zufällig und hat **Mittelwert** $\mathbf{E}(F) = 0$.
Andernfalls machen wir einen **systematischen Fehler**!

Den Fehler F können wir nicht direkt beobachtet (und korrigieren).
Je mehr Messwerte wir haben, desto besser können wir ihn rausfiltern:

Mit Wkt 95% überdeckt das Konfidenzintervall I_2 den wahren Wert x.
Mit Wkt 99% überdeckt das Konfidenzintervall I_3 den wahren Wert x.

Dies sind die Konfidenzniveaus 95% bzw. 99% der **Normalverteilung**!
Sie gilt falls $F \sim \mathrm{N}(0, \sigma_F^2)$ normalverteilt ist oder n ausreichend groß,
denn $\hat{X} \sim \mathrm{N}(x, \sigma_F^2/n)$ gilt exakt oder in guter Näherung dank ZGS.

🙂 Mit hinreichend vielen Messdaten bestimmen wir x beliebig genau.
Merke: Doppelte Genauigkeit kostet viermal so viele Messungen.

Mathematische Statistik: lineare Regression

Wir nehmen an, Y hängt linear von X ab plus zufällige Störung F:

$$Y = a + b\,X + F \quad \text{etwa mit} \quad F \sim \mathrm{N}(0, \sigma_F^2)$$

Unabhängige Wiederholung ergibt die Messdaten $(x_1, y_1), \ldots, (x_n, y_n)$.
Die **Regressionsparameter** a und b schätzen wir hieraus wie folgt:

$$b \approx \hat{b} := \frac{\sum(x_i - \hat{x})(y_i - \hat{y})}{\sum(x_i - \hat{x})^2}, \qquad a \approx \hat{a} := \hat{y} - \hat{b}\,\hat{x}$$

Die **Güte dieser Schätzungen** wird bestimmt durch die Varianzen:

$$\sigma_{\hat{b}}^2 \approx \hat{\sigma}_{\hat{b}}^2 := \frac{\sum(\hat{a} + \hat{b}\,x_i - y_i)^2}{(n-2)\sum(x_i - \hat{x})^2}, \qquad \sigma_{\hat{a}}^2 \approx \hat{\sigma}_{\hat{a}}^2 := \frac{\hat{\sigma}_{\hat{b}}^2}{n}\sum_{i=1}^{n} x_i^2$$

Wir erhalten **Konfidenzintervalle** für die gesuchten Werte a bzw. b:

$$\hat{I}_2 = [\hat{a} \pm 2\hat{\sigma}_{\hat{a}}] = \left[\, \hat{a} - 2\hat{\sigma}_{\hat{a}} \;;\; \hat{a} + 2\hat{\sigma}_{\hat{a}} \,\right] \quad \ni a \text{ mit } 95\% \text{ Sicherheit}$$
$$\hat{I}_3 = [\hat{a} \pm 3\hat{\sigma}_{\hat{a}}] = \left[\, \hat{a} - 3\hat{\sigma}_{\hat{a}} \;;\; \hat{a} + 3\hat{\sigma}_{\hat{a}} \,\right] \quad \ni a \text{ mit } 99\% \text{ Sicherheit}$$

Mathematische Statistik: lineare Regression

Gesucht sind die Werte a, b, doch diese sind uns leider unbekannt.
Wir schätzen sie durch $\hat{a}, \hat{b}$ anhand der Messdaten $(x_1, y_1), \ldots, (x_n, y_n)$.

Grundlegende Annahmen unseres **stochastischen Modells**:
Der Messfehler F ist unabhängig von X mit **Mittelwert** $\mathbf{E}(F) = 0$.
Andernfalls machen wir immer denselben **systematischen Fehler**!

Den Fehler F können wir weder direkt beobachten noch korrigieren.
Je mehr Messwerte wir haben, desto besser können wir ihn rausfiltern:

Mit Wkt 95% überdeckt das Konfidenzintervall I_2 den wahren Wert a.
Mit Wkt 99% überdeckt das Konfidenzintervall I_3 den wahren Wert a.

Auch hier nutzen wir die Konfidenzniveaus der **Normalverteilung**!
Sie gilt falls $F \sim \mathrm{N}(0, \sigma_F^2)$ normalverteilt ist oder n ausreichend groß,
denn $\hat{A} \sim \mathrm{N}(a, \sigma_{\hat{a}}^2)$ und $\hat{B} \sim \mathrm{N}(b, \sigma_{\hat{b}}^2)$ gilt exakt oder in guter Näherung.

🙂 Ausreichend viele Messdaten bestimmen a, b beliebig genau.
Zudem erhalten wir eine Schätzung für die Güte der Näherung.

⚠ Jede gegebene Messreihe liefert nur eine begrenzte Genauigkeit.
Doppelte Genauigkeit kostet viermal so viele Messungen.

Bestimmung der Ausgleichsgeraden

Wir nehmen an, y_i hängt linear von x_i ab plus zufällige Störung f_i:

$$y_i = a + b\,x_i + f_i \quad \text{mit Fehler} \quad f_i = y_i - a - b\,x_i$$

Wir suchen nach der Geraden, die diese Fehler insgesamt minimiert:

$$Q(a,b) := \sum_{i=1}^n f_i^2 = \sum_{i=1}^n (y_i - a - b\,x_i)^2 \quad \overset{!}{\to} \quad \min$$

Das ist eine positive quadratische Funktion in a und b. Minimum:

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^n 2(y_i - a - b\,x_i)\cdot(-1) \quad \overset{!}{=} \quad 0$$
$$\frac{\partial Q}{\partial b} = \sum_{i=1}^n 2(y_i - a - b\,x_i)\cdot(-x_i) \quad \overset{!}{=} \quad 0$$

Dieses lineare Gleichungssystem führt zu den obigen Formeln:

$$\hat a = \hat y - \hat b\,\hat x, \qquad \hat b = \frac{\sum(x_i-\hat x)(y_i-\hat y)}{\sum(x_i-\hat x)^2} = \frac{\sum x_i y_i - n\hat x\hat y}{\sum x_i^2 - n\hat x^2}$$

🙂 Geometrisch: Für die zentrierten Vektoren $\vec x_i = x_i - \hat x$ und $\vec y_i = y_i - \hat y$ berechnet sich $\hat b = \langle \vec x \mid \vec y\rangle / \langle \vec x \mid \vec x\rangle$ durch Skalarprodukte!

Bestimmung der Ausgleichsgeraden

Wir nennen $y = \hat a + \hat b\,x$ die **Kleinste-Quadrate-Gerade** (KQ-Gerade).
Der Wert $\hat y_i = \hat a + \hat b\,x_i$ heißt der **KQ-gefittete Wert** und entsprechend die Abweichung $\hat f_i = y_i - \hat y_i = y_i - \hat a + \hat b\,x_i$ der **KQ-gefittete Fehler**.

Eigenschaften: Die Summe aller Fehler $\hat f_i = y_i - a - b\,x_i$ ist Null.
Liegen alle Datenpunkte (x_i, y_i) auf einer Geraden $y = a + bx$, so gilt $\hat a = a$ und $\hat b = b$ mit perfektem Fit: $\hat y_i = y_i$ und $\hat f_i = 0$.

Wie gut beschreibt die Gerade die Daten? **Korrelationskoeffizient**

$$r(x,y) := \frac{\sum(x_i-\hat x)(y_i-\hat y)}{\sqrt{\sum(x_i-\hat x)^2}\sqrt{\sum(y_i-\hat y)^2}} = \frac{\langle \vec x \mid \vec y\rangle}{\sqrt{\langle \vec x \mid \vec x\rangle}\sqrt{\langle \vec y \mid \vec y\rangle}}$$

Es gilt $-1 \le r \le 1$, mit $r = \pm 1$ genau dann, wenn alle Daten auf einer Geraden liegen: gleichsinnig für $r = +1$ und gegensinnig für $r = -1$.

Die Linearisierung ist gut für große Werte von r^2, schlecht für kleine.

Geometrisch bedeutet $r = 0$: Die Vektoren $\vec x, \vec y$ stehen **senkrecht**.
Stochastisch: Die Zufallsvariablen x, y sind **linear unkorreliert**.

Für Skalarprodukte $\langle \vec x \mid \vec y\rangle$ gilt die Cauchy–Schwarz–Ungleichung. I134

📖 Zur Wiederholung siehe Stroppel, Höhere Mathematik 1, §3.5.

Bestimmung der Ausgleichsgeraden

Wir nehmen an, Y hängt linear von X ab plus zufällige Störung F:

$$Y = a + b\,X + F \quad \text{etwa mit} \quad F \sim \mathrm{N}(0,\sigma_F^2)$$

Wie gut ist unsere Schätzung $\hat a, \hat b$ für die unbekannten Parameter a, b?
Wir betrachten unabhängig wiederholte Messungen (X_i, Y_i) und analysieren unsere Schätzungen $\hat a, \hat b$ als Zufallsvariable $\hat A, \hat B$:

$$Y_i = a + b\,X_i + F_i, \qquad F_i \sim \mathrm{N}(0,\sigma_F^2)$$
$$\hat X := \frac{1}{n}\sum_{i=1}^n X_i, \qquad \hat Y := \frac{1}{n}\sum_{i=1}^n Y_i$$
$$\hat B = \frac{\sum(X_i-\hat X)(Y_i-\hat Y)}{\sum(X_i-\hat X)^2}, \qquad \hat A = \hat Y - \hat B\hat X$$

Wir nehmen vereinfachend an, die additiven Fehler $F_i \sim \mathrm{N}(0,\sigma_F^2)$ sind normalverteilt. Dann sind die Zufallsvariablen $\hat A$ und $\hat B$ normalverteilt; allgemein gilt dies für große n in guter Näherung dank ZGS.

Bestimmung der Ausgleichsgeraden

Für die Erwartungen finden wir (nach kurzer Rechnung):

$$\mathbf{E}(\hat B) = b \approx \hat b := \frac{\sum(x_i-\hat x)(y_i-\hat y)}{\sum(x_i-\hat x)^2}, \qquad \mathbf{E}(\hat A) = a \approx \hat a := \hat y - \hat b\,\hat x$$

Für die Varianzen finden wir (nach längerer Rechnung):

$$\mathbf{V}(\hat B) = \frac{\sigma_F^2}{n\,\mathbf{V}(X)} \qquad \approx \hat\sigma_b^2 := \frac{\sum(\hat a + \hat b\,x_i - y_i)^2}{(n-2)\sum(x_i-\hat x)^2}$$
$$\mathbf{V}(\hat A) = \frac{\sigma_F^2\,\mathbf{E}(X^2)}{n\,\mathbf{V}(X)} \qquad \approx \hat\sigma_a^2 := \hat\sigma_b^2 \cdot \frac{1}{n}\sum_{i=1}^n x_i^2$$

Die Form der Varianzschätzer ist plausibel, die kleine Korrektur von $1/n$ zu $1/(n-1)$ oder $1/(n-2)$ sichert die richtige Erwartung. (Für große n, etwa $n \gg 100$, ist diese Korrektur meist unerheblich.)

Hieraus erhalten wir Vertrauensintervalle, etwa 2σ zu 95% Sicherheit.
Der Aufwand lohnt sich: Wir gewinnen ein Maß für die (Un)Sicherheit.

🙂 Wir bekommen nicht nur **Punktschätzer** der gesuchten Werte a, b, sondern sogar **Intervallschätzer**: Das enthält viel mehr Information!

⚠ Grundlose Schätzungen sind sinnlos, wir brauchen Sicherheit!

Zufallsvariablen und Erwartungswerte

Sei $(\Omega, \mathscr{A}, \mathbf{P})$ ein WRaum, zum Beispiel diskret oder kontinuierlich. V112
Eine reelle **Zufallsvariable** ist eine messbare Funktion $X : \Omega \to \mathbb{R}$. V211
Sie ordnet jedem Zufallsergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega) \in \mathbb{R}$ zu.
Im Bildbereich $\mathbb{R}$ definiert sie die **Verteilung** $\mathbf{P}_X : \mathscr{B}(\mathbb{R}) \to [0,1]$ durch

$$\mathbf{P}_X(B) = \mathbf{P}(X \in B) = \mathbf{P}\big(\{\, \omega \in \Omega \;\big|\; X(\omega) \in B \,\}\big).$$

Ihr **Erwartungswert** ist gegeben durch

$$\mu = \mathbf{E}(X) := \int_\Omega X(\omega)\, \mathrm{d}\mathbf{P} = \int_{\mathbb{R}} x\, \mathrm{d}\mathbf{P}_X.$$

Ist die Verteilung $\mathbf{P}_X$ auf $\mathbb{R}$ kontinuierlich mit Dichte $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$, so gilt

$$\mathbf{E}(X) = \int_{\mathbb{R}} x\, f(x)\, \mathrm{d}x.$$

Ist $\mathbf{P}_X$ diskret mit Wkten $p(x) = \mathbf{P}(\{\, \omega \in \Omega \mid X(\omega) = x \,\})$, so gilt

$$\mathbf{E}(X) = \sum\nolimits_{x \in \mathbb{R}} x\, p(x).$$

⚠ Wir setzen hierzu absolute Integrierbarkeit / Summierbarkeit voraus!

Varianz und Streuung

Sei $(\Omega, \mathscr{A}, \mathbf{P})$ ein WRaum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.
Sei $\mathbf{E}(|X|) < \infty$, so dass der Erwartungswert $\mu = \mathbf{E}(X)$ existiert.
Das (absolute, zentrierte) **n-te Moment** von X ist gegeben durch

$$\rho^n(X) := \mathbf{E}\big[|X - \mu|^n\big].$$

Fall $n = 2$: Die **Varianz** von X ist die mittlere quadratische Abweichung

$$\sigma^2(X) = \mathbf{V}(X) := \mathbf{E}\big[(X - \mu)^2\big] = \mathbf{E}(X^2) - \mathbf{E}(X)^2 \geq 0.$$

Die **Streuung** oder **Standardabweichung** ist $\sigma(X) := \sqrt{\mathbf{V}(X)}$.
Nach Chebychev gelten folgende Abschätzungen für alle $h, k > 0$:

$$\mathbf{P}\Big[|X - \mu| \geq k\sigma\Big] \leq \frac{1}{k^2}, \qquad \mathbf{P}\Big[\mu - k\sigma < X < \mu + k\sigma\Big] \geq 1 - \frac{1}{k^2},$$
$$\mathbf{P}\Big[X \geq \mu + k\sigma\Big] \leq \frac{1}{1+k^2}, \quad \mathbf{P}\Big[\mu - h\sigma < X < \mu + k\sigma\Big] \geq \frac{4(hk-1)}{(h+k)^2}$$

🙂 Kennen Sie nur μ und σ, so sind diese Ungleichungen optimal.
Die Wkt großer Abweichungen fällt mindestens quadratisch, mit $1/k^2$.

Unabhängige Zufallsvariablen

Eine Familie $X = (X_1, \ldots, X_n)$ von Zufallsvariablen $X_k : \Omega \to \mathbb{R}$ heißt **(stochastisch) unabhängig**, wenn für alle Intervalle $I_1, \ldots, I_n \subseteq \mathbb{R}$ gilt

$$\mathbf{P}(\; X_1 \in I_1,\; \ldots,\; X_n \in I_n \;) = \mathbf{P}(X_1 \in I_1) \cdots \mathbf{P}(X_n \in I_n).$$

Die gemeinsame Verteilung von $X_1, \ldots, X_n$ ist dann das Produktmaß

$$\mathbf{P}_X = \mathbf{P}_{(X_1,\ldots,X_n)} = \mathbf{P}_{X_1} \otimes \cdots \otimes \mathbf{P}_{X_n}.$$

Sind $\mathbf{P}_{X_1}, \ldots, \mathbf{P}_{X_n}$ auf $\mathbb{R}$ kontinuierliche WMaße mit Dichten $f_1, \ldots, f_n$, so ist auch das Produktmaß $\mathbf{P}_X$ auf $\mathbb{R}^n$ kontinuierlich, mit Produktdichte

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

Erwartungen unabhängiger Zufallsvariablen X, Y multiplizieren sich:

$$\mathbf{E}(X \cdot Y) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$$

Ihre Kovarianz $\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ verschwindet somit.
Varianzen unabhängiger Zufallsvariablen X, Y addieren sich:

$$\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\mathbf{Cov}(X,Y)$$

⚠ Für abhängige Zufallsvariablen gilt dies im Allgemeinen nicht!

Das Gesetz der großen Zahlen

Eine Messung / ein Experiment entspricht einer Zufallsvariablen
$X : \Omega \to \mathbb{R}$ mit Erwartung $\mu = \mathbf{E}(X)$ und Varianz $\sigma^2 = \mathbf{V}(X) < \infty$.
Wir führen unabhängige Wiederholungen $X_1, X_2, X_3, \ldots$ durch.
Aus diesen Messwerten bilden wir den **empirischen Mittelwert**

$$\hat{X} := \frac{1}{n}(X_1 + \ldots + X_n).$$

Es gilt $\mathbf{E}(\hat{X}) = \mu$ und dank Unabhängigkeit zudem $\mathbf{V}(\hat{X}) = \sigma^2/n$.
Mit zunehmendem n streut $\hat{X}$ immer weniger, denn $\hat{\sigma} = \sigma/\sqrt{n} \to 0$.
Der Mittelwert $\hat{X}$ nähert sich dem Erwartungswert μ: Für jedes $\varepsilon > 0$ gilt

$$\mathbf{P}\Big[|\hat{X} - \mu| \geq \varepsilon\Big] \;\leq\; \frac{\sigma^2}{n\,\varepsilon^2} \;\searrow\; 0.$$

Dieser Grenzwert besagt in Worten: Große Abweichungen $|\hat{X} - \mu| \geq \varepsilon$ werden beliebig unwahrscheinlich, wenn wir n hinreichend groß wählen.

In praktischen Anwendungen möchte man n nicht allzu groß wählen, denn wiederholte Messungen sind teuer. Wir brauchen daher bessere Schranken für die Abweichung $\mathbf{P}(|\hat{X} - \mu| \geq \varepsilon)$. Diese liefert der ZGS!

Der zentrale Grenzwertsatz (ZGS)

Der ZGS besagt: Die Summe $S = X_1 + \ldots + X_n$ vieler unabhängiger, aber ähnlich großer Zufallsvariablen ist annähernd normalverteilt.

Sei $(\Omega, \mathscr{A}, \mathbf{P})$ ein WRaum und $X_1, X_2, X_3, \ldots : \Omega \to \mathbb{R}$ unabhängig mit

1. endlichen Erwartungen $\mu_k = \mathbf{E}(X_k)$ dank $\mathbf{E}(|X_k|) < \infty$,
2. strikt positiven Varianzen $\sigma_k^2 = \mathbf{E}\big(|X_k - \mu_k|^2\big) \geq \sigma_0^2 > 0$,
3. beschränkten dritten Momenten $\rho_k^3 = \mathbf{E}\big(|X_k - \mu_k|^3\big) \leq \rho_0^3 < \infty$.

Die Summe $S = X_1 + \ldots + X_n$ hat die Erwartung $\mu = \mu_1 + \ldots + \mu_n$ und die Varianz $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Es gilt $\mathbf{P}_S \approx N(\mu, \sigma^2)$, genauer

$$\mathbf{P}(a \leq S \leq b) = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi + \delta,$$

$$|\delta| \leq \frac{\rho_1^3 + \ldots + \rho_n^3}{(\sigma_1^2 + \ldots + \sigma_n^2)^{3/2}} \leq \frac{\rho_0^3}{\sigma_0^3\sqrt{n}} \searrow 0.$$

Zentriert und normiert zu $S^* := (S - \mu)/\sigma$ erhalten wir $\mathbf{P}_{S^*} \approx N(0, 1)$.
Der lokale Grenzwertsatz V3A ist spezieller, dabei aber auch präziser.

Konfidenzintervalle

In der Praxis ist folgende Anwendung des ZGS besonders wichtig:
Als Stichprobe für X führen wir n **unabhängige Messungen** aus.
Aus den so gewonnenen Messwerten $x_1, \ldots, x_n \in \mathbb{R}$ berechnen wir

den **Stichprobenmittelwert** $\hat{x} := \frac{1}{n}\sum_{k=1}^{n} x_k,$

die **Stichprobenvarianz** $\hat{\sigma}^2 := \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{x})^2,$

das 2σ**–Konfidenzintervall** $I_2 := \Big[\, \hat{x} - 2\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{x} + 2\frac{\hat{\sigma}}{\sqrt{n}} \,\Big].$

Für große n gilt: Bei 95% aller Stichproben überdeckt das 2σ–Intervall I_2 den (konstanten aber uns unbekannten) Erwartungswert $\mu = \mathbf{E}(X)$.
Noch strenger: Bei 99% aller Stichproben überdeckt das 3σ–Intervall I_3 den (konstanten aber uns unbekannten) Erwartungswert $\mu = \mathbf{E}(X)$.
Der Nenner $\sqrt{n}$ bedeutet: Für doppelte Genauigkeit braucht man eine viermal größere Stichprobe, für zehnfache Genauigkeit eine 100mal größere Stichprobe. Das ist Fluch und Segen der großen Zahlen!

Verständnisfragen: Zufallsvariablen

Versuchen Sie, folgende Fragen frei aber genau zu beantworten, etwa so, wie Sie dies einer Kommiliton:in / Kolleg:in erklären wollen.

Was ist ein Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbf{P})$? Was ist hierauf eine (reelle) Zufallsvariable X? Was ist ihre Verteilung $\mathbf{P}_X$ im Bildraum $\mathbb{R}$? Erinnerung: Wann ist $(\Omega, \mathscr{A}, \mathbf{P})$ diskret? endlich? laplacesch?

Wie berechnet man den Erwartungswert $\mu = \mathbf{E}(X)$? bei diskreter Verteilung? bei kontinuierlicher Verteilung? Wie berechnet man die Varianz $\sigma^2 = \mathbf{V}(X)$ und Streuung σ? Was sind die Momente von X?

Nennen Sie wichtige Verteilungen und ihre Anwendungen.
Was sind ihre Kenngrößen μ und σ? Wie berechnet man sie?
Wie verhalten sie sich bei einer affinen Skalierung $Y = aX + b$?

Wie lauten und was besagen die Ungleichungen von Chebychev?
Wann lassen sie sich anwenden, was muss man hierzu wissen?

Gibt es genauere Ungleichungen, wenn man die Verteilung kennt?
Was gilt für die Normalverteilung? Was besagt die 68–95–99–Regel?

Verständnisfragen: zentraler Grenzwertsatz

Wann sind zwei Zufallsvariablen X, Y unabhängig? Wann sind die Indikatorfunktionen $\mathbf{I}_A, \mathbf{I}_B$ zweier Ereignisse $A, B \subseteq \Omega$ unabhängig?
Kennen Sie weitere Beispiele von un/abhängigen Zufallsvariablen?

Gilt stets $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$? und $\mathbf{E}(X \cdot Y) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$? und $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(X)$? Was gilt bei Unabhängigkeit?

Was sind Erwartungswert und Varianz für $\mathrm{B}(1, t)$? Wie kann man hieraus mühelos Erwartungswert und Varianz von $\mathrm{B}(n, t)$ ablesen?

Was besagt das Gesetz der großen Zahlen? Wie kann man Wkten / Erwartungswerte empirisch messen? Wie schnell ist die Konvergenz?

Was besagt der zentrale Grenzwertsatz? Welche drei Voraussetzungen braucht, welche Schlussfolgerung gewinnt man? Wie schnell ist die Konvergenz? Was ist gleich / anders beim lokalen Grenzwertsatz?

Wie berechnet man Stichprobenmittelwert und -Varianz?
Wie bestimmt man hieraus das Konfidenzintervall der Stichprobe?
Wie liegt es zum (gesuchten aber unbekannten) Erwartungswert?

ZGS: Wann erhalten wir nicht die Normalverteilung?

⚠ Die Voraussetzungen $\sigma_k \geq \sigma_0 > 0$ und $\rho_k \leq \rho_0 < \infty$ des zentralen Grenzwertsatzes W1D stellen sicher, dass alle Beiträge $X_1, X_2, X_3, \ldots$ ähnlich stark streuen. Andernfalls können wir nicht garantieren, dass ihre Summen $S_n = X_1 + \ldots + X_n$ sich der Normalverteilung nähern. Warnende Gegenbeispiele illustrieren dies, einfach aber eindrücklich:

Seien $X_1, X_2, X_3 \ldots : \Omega \to \{0, 1\}$ unabhängig und jeweils gleichverteilt.

Aufgabe: (1) Wir betrachten die Summe $S_n = X_1 + \ldots + X_n$ sowie ihre Normierung $S_n^* = (S_n - \mu(S_n))/\sigma(S_n)$, sodass $\mu(S_n^*) = 0$ und $\sigma(S_n^*) = 1$.
(a) Berechnen Sie explizit $\mu(X_k)$, $\sigma(X_k)$, $\rho(X_k)$ sowie $\mu(S_n)$, $\sigma(S_n)$.
(b) Bestimmen Sie die Verteilung $\mathbf{P}_{S_n}$, skizzieren Sie die ersten Fälle.
(c) Welche Voraussetzungen des ZGS sind erfüllt, welche nicht?
(d) Welcher Grenzverteilung nähern sich $\mathbf{P}_{S_n}$ bzw. $\mathbf{P}_{S_n^*}$ für $n \to \infty$?

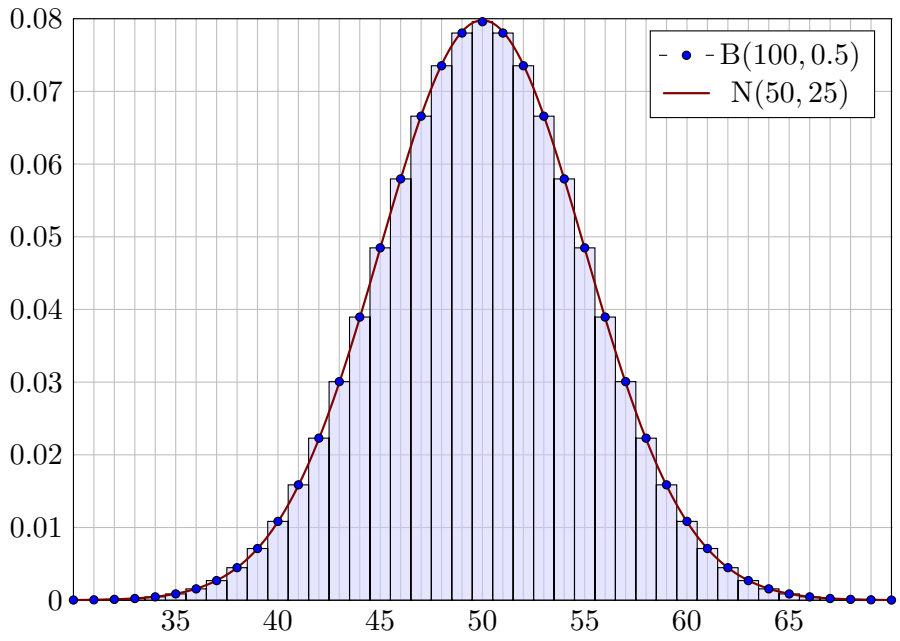
(2) Dieselben Fragen (a–d) für $S_n = Y_1 + \ldots + Y_n$ mit $Y_k = X_1$.

(3) Dieselben Fragen (a–d) für $S_n = Y_1 + \ldots + Y_n$ mit $Y_k = 2^{-k} X_k$.

(4) Dieselben Fragen (a–d) für $S_n = Y_1 + \ldots + Y_n$ mit $Y_k = 2^{k-1} X_k$.

Variante: $X_1, X_2, X_3 \ldots : \Omega \to \{-1, 0, +1\}$ und (3–4) zur Basis 3 statt 2.

ZGS: Wann erhalten wir nicht die Normalverteilung?



ZGS: Wann erhalten wir nicht die Normalverteilung?

Lösung: (1a) Hier gilt $\mu(X_k) = 1/2$ und $\sigma^2(X_k) = 1/4$ und $\rho^3(X_k) = 1/8$, dank Linearität $\mu(S_n) = n/2$, zudem dank Unabhängigkeit $\sigma^2(S_n) = n/4$.

(1b) Wir wissen bereits $\mathbf{P}_{S_n} = \mathrm{B}(n, p)$ mit $p = 1/2$, also ausgeschrieben:

$$\mathbf{P}(S_n = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \;=\; \binom{n}{k} \frac{1}{2^n}$$

(1c) Alle drei Voraussetzungen des Satzes W1D sind hier erfüllt! ✔ ✔ ✔

(1d) Dank ZGS erhalten wir $\mathbf{P}_{S_n^*} \to \mathrm{N}(0, 1)$. Ausführlich und genauer:

$$\mathbf{P}(a \leq S_n \leq b) \;=\; \int_{(a-\mu_n)/\sigma_n}^{(b-\mu_n)/\sigma_n} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \;+\; \delta$$

$$\mathbf{P}(\alpha \leq S_n^* \leq \beta) \;=\; \int_{\alpha}^{\beta} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \;+\; \delta$$

mit der Fehlerschranke $|\delta| \leq 1/\sqrt{n} \to 0$. Der lokale Grenzwertsatz V3A garantiert die noch bessere Fehlerschranke $|\delta| \leq 4/3n$ für $n \geq 100$.

⚠ Ohne Normierung gilt $\mathbf{P}(S_n \in [a, b]) \to 0$ für $n \to \infty$. (Skizze!)

ZGS: Wann erhalten wir nicht die Normalverteilung?

(2a) Wie in (1a) gilt $\mu(Y_k) = 1/2$ und $\sigma^2(Y_k) = 1/4$ und $\rho^3(Y_k) = 1/8$.
Für die Summe $S_n = nX_1$ gilt $\mu(S_n) = n/2$ aber $\sigma^2(S_n) = n^2/4$.

(2b) Die Summe S_n ist gleichverteilt auf $\{0, n\}$, das heißt,
sie nimmt nur zwei Werte an: $\mathbf{P}(S_n{=}0) = \mathbf{P}(S_n{=}n) = 1/2$.
Normiert ist $S_n^* = 2X_1 - 1$ gleichverteilt auf $\{-1, +1\}$.

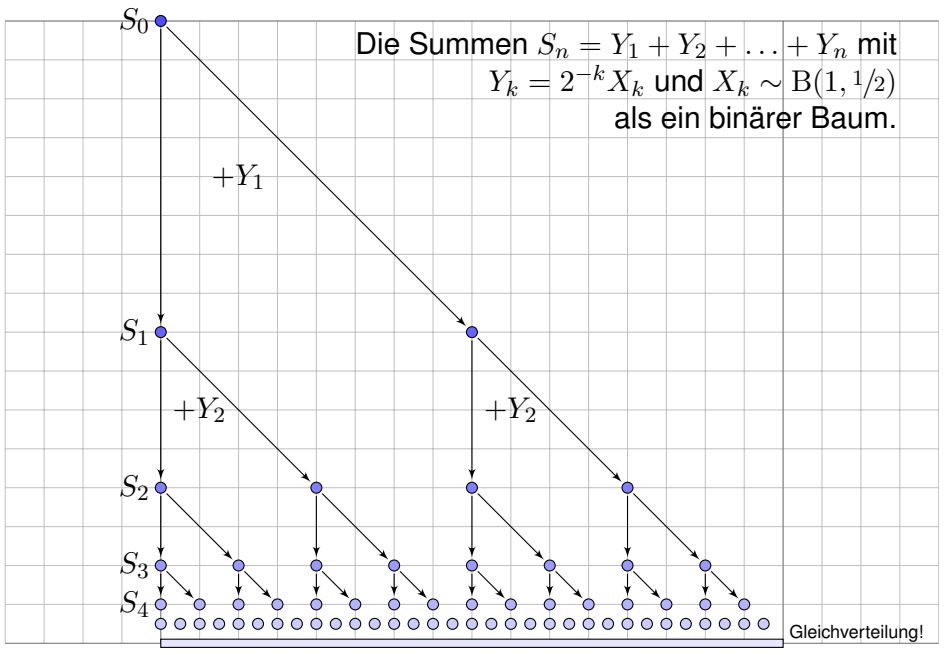
(2c) Nur zwei der drei Voraussetzungen des ZGS sind hier erfüllt:
Die Varianzen $\sigma_k^2 = 1/4 > 0$ sind strikt positiv. ✔
Die dritten Momente $\rho_k^3 = 1/8 < \infty$ sind beschränkt. ✔
Aber die Zufallsvariablen $Y_1, Y_2, \ldots$ sind nicht unabhängig! ✘

⚠ Der zentrale Grenzwertsatz lässt sich daher hier nicht anwenden.
Tatsächlich konvergiert $\mathbf{P}_{S_n^*}$ *nicht* gegen die Normalverteilung $\mathrm{N}(0, 1)$!

(2d) Für alle n ist die normierte Summe S_n^* gleichverteilt auf $\{-1, +1\}$.

⚠ Ohne Normierung gilt für $n \to \infty$ der Grenzwert $\mathbf{P}(S_n \in [a, b]) \to 0$ falls $0 \notin [a, b]$ und $\mathbf{P}(S_n \in [a, b]) \to 1/2$ falls $0 \in [a, b]$. Das ist keine WVerteilung mehr: Wie in (1d) verschwindet Masse nach Unendlich.

ZGS: Wann erhalten wir nicht die Normalverteilung?



ZGS: Wann erhalten wir nicht die Normalverteilung?

(3a) Für $Y_k = 2^{-k} X_k$ gilt dank (1a) nach Skalierung $\mu(Y_k) = 2^{-k}/2$ und $\sigma^2(Y_k) = 2^{-2k}/4$ und $\rho^3(Y_k) = 2^{-3k}/8$. Dank Linearität gilt dann $\mu(S_n) = (1 - 2^{-n})/2$, dank Unabhängigkeit $\sigma^2(S_n) = (1 - 2^{-2n})/12$.

(3b) Im Binärsystem gilt:

$$S_n = 2^{-1} X_1 + 2^{-2} X_2 + \ldots + 2^{-n} X_n = (0.X_1 X_2 \ldots X_n)_{\mathrm{bin}}$$

Demnach ist $\mathbf{P}_{S_n}$ die Gleichverteilung auf der endlichen Menge

$$M_n = \left\{ m/2^n \;\middle|\; m = 0, 1, \ldots, 2^n - 1 \right\} \subseteq [0, 1].$$

(3c) Nur zwei der drei Voraussetzungen des ZGS sind hier erfüllt:
Die Zufallsvariablen $Y_k = 2^{-k} X_k$ sind untereinander unabhängig. ✔
Die dritten Momente $\rho_k^3 = 2^{-3k}/8 \le \rho_1^3 = 1/64 < \infty$ sind beschränkt. ✔
Hingegen gilt $\sigma_k^2 = 2^{-2k}/4 \searrow 0$, also nicht $\sigma_k \ge \sigma_0 > 0$. ✘

⚠ Der zentrale Grenzwertsatz lässt sich daher hier nicht anwenden.
Genauer zeigen wir schließlich, dass die Verteilungen $\mathbf{P}_{S_n}$ bzw. $\mathbf{P}_{S_n^*}$ tatsächlich *nicht* gegen die Normalverteilung konvergieren!

ZGS: Wann erhalten wir nicht die Normalverteilung?

(3d) Für $n \to \infty$ nähert sich $\mathbf{P}_{S_n}$ der Gleichverteilung auf $[0, 1]$.

Dank (3b) scheint dies plausibel. Wir rechnen es ausführlich nach.
Für alle $0 \le a < b \le 1$ und $k \in \mathbb{N}$ gelten folgende Implikationen:

$$\begin{aligned} & \big|[a, b] \cap M_n\big| = k \\ \Longrightarrow \quad & k 2^{-n} \le |b - a| < (k + 1) 2^{-n} \\ \Longrightarrow \quad & \big|[a, b] \cap M_n\big| \in \{k, k + 1\} \end{aligned}$$

Man beweist dies für $k = 0$ und per Induktion für $k = 1, 2, 3, \ldots$. Also:

$$\mathbf{P}\big(S_n \in [a, b]\big) \in \frac{\lfloor (b - a) 2^n \rfloor + \{0, 1\}}{2^n} \to |b - a|$$

🙂 Die Wahrscheinlichkeit für das Ereignis $S_n \in [a, b]$ geht für $n \to \infty$ gegen die Intervalllänge $|b - a|$. Das ist die Gleichverteilung auf $[0, 1]$.

Demnach konvergiert die normierte Zufallsvariable $S_n^* = (S_n - \mu_n)/\sigma_n$ gegen die kontinuierliche Gleichverteilung auf dem Intervall $[-\sqrt{3}, \sqrt{3}]$.
All diese Verteilungen haben Erwartung 0 und Streuung 1. V235

ZGS: Wann erhalten wir nicht die Normalverteilung?

(4a) Für $Y_k = 2^{k-1} X_k$ gilt nach (1a) dank Skalierung $\mu(Y_k) = 2^{k-1}/2$ und $\sigma^2(Y_k) = 2^{2k-2}/4$ und $\rho^3(Y_k) = 2^{3k-3}/8$. Dank Linearität gilt $\mu(S_n) = (2^n - 1)/2$ dank Unabhängigkeit $\sigma^2(S_n) = (2^{2n} - 1)/12$.

(4b) Im Binärsystem gilt:

$$S_n = X_1 + 2X_2 + \ldots + 2^{n-1} X_n = (X_n \ldots X_2 X_1)_{\mathrm{bin}}$$

Demnach ist $\mathbf{P}_{S_n}$ die Gleichverteilung auf der endlichen Menge

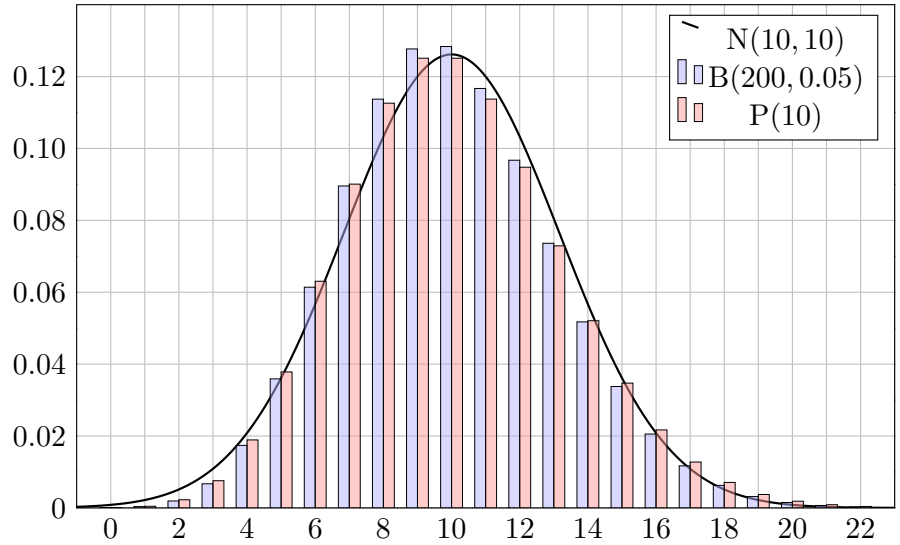
$$M_n = \{0, 1, \ldots, 2^n - 1\} \subseteq \mathbb{N}.$$

(4c) Nur zwei der drei Voraussetzungen des ZGS sind hier erfüllt:
Die Zufallsvariablen $Y_k = 2^{-k} X_k$ sind untereinander unabhängig. ✔
Die Varianzen $\sigma_k^2 = 2^{2k-2}/4 \ge \sigma_1^2 = 1/4 > 0$ sind strikt positiv. ✔
Hingegen gilt $\rho_k^3 = 2^{3k-3}/8 \nearrow \infty$, also nicht $\rho_k \le \rho_0 < \infty$. ✘

⚠ Der zentrale Grenzwertsatz lässt sich daher hier nicht anwenden.
Wie in (3d) folgt, dass die Verteilung von S_n^* tatsächlich nicht gegen $\mathrm{N}(0, 1)$ konvergiert, sondern gegen die Gleichverteilung auf $[-\sqrt{3}, \sqrt{3}]$.

Vergleich: Binomial, Poisson, Normal

W317
Ausführung



Aufgabe: Die Poisson–Näherung $\mathrm{B}(n,\mu/n) \to \mathrm{P}(\mu)$ für $n \to \infty$ und der LGS $\mathrm{B}(n,\mu/n) \approx \mathrm{N}(\mu,\mu)$ zeigen $\mathrm{P}(\mu) \approx \mathrm{N}(\mu,\mu)$ wie skizziert. Welche Fehlerschranke gilt demnach zwischen $\mathrm{P}(\mu)$ und $\mathrm{N}(\mu,\mu)$?

Vergleich: Binomial, Poisson, Normal

W318
Ausführung

Lösung: Für $\mu \geq 0$ ist die Poisson–Verteilung gegeben durch

$$\mathrm{P}(\mu) : \mathbb{N} \to [0,1] \quad \text{mit} \quad \mathrm{P}(\mu)(k) = \mathrm{e}^{-\mu}\,\frac{\mu^k}{k!}.$$

Sie hat Erwartung μ und Varianz $\sigma^2 = \mu$. Für $t = \mu/n$ und $n \to \infty$ gilt

$$\mathrm{B}(n,t) \to \mathrm{P}(\mu) \quad \text{also punktweise} \quad \binom{n}{k} t^k (1-t)^{n-k} \to \mathrm{e}^{-\mu}\,\frac{\mu^k}{k!}.$$

Dank LGS nähert sich also auch die Poisson– der Normalverteilung:

$$\sum_{k=a}^{b} \binom{n}{k} t^k (1-t)^{n-k} \;\to\; \sum_{k=a}^{b} \mathrm{e}^{-\mu}\,\frac{\mu^k}{k!} \;=\; \int_\alpha^\beta \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \;+\; \delta$$
$$= \int_\alpha^\beta \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi + \Big[\frac{1-\xi^2}{6\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\xi^2/2}\Big]_\alpha^\beta \;+\; \varepsilon$$

mit den Grenzen $\alpha = (a - 1/2 - \mu)/\sigma$ und $\beta = (b + 1/2 - \mu)/\sigma$
und Fehlerschranken $|\delta| < 1/(6\sigma)$ bzw. $|\varepsilon| < 1/(3\sigma^2)$ für $\sigma \geq 5$.
☺ Dies zeigt $|\delta| < 1/\sigma \to 0$, aber auch $|\delta| \sim \mathrm{const}/\sqrt{n}$ im ZGS.

Fixpunkte zufälliger Abbildungen

W319
Ausführung

Aufgabe: (1) Wir betrachten zufällige Abbildungen $\omega : I \to I$ der Menge $I = \{1,\ldots,n\}$ in sich. Hierzu sei $(\Omega = I^I, \mathbf{P})$ laplacesch mit $|\Omega| = n^n$.
Zu $\omega \in \Omega$ sei $F(\omega) := \sharp\{\, i \in I \mid \omega(i) = i \,\}$ die Anzahl der Fixpunkte.
(a) Bestimmen Sie die Erwartung $\mathbf{E}(F)$ und die Varianz $\mathbf{V}(F)$.
(b) Ist die Zufallsvariable F binomial- oder Poisson–verteilt?
(2) Dieselben Fragen für zufällige Permutationen $\omega : I \xrightarrow{\sim} I$.

Lösung: Die Rechnung scheint zunächst schwierig, gar hoffnungslos. Unsere mathematischen Werkzeuge helfen: Geschickt zerlegen wir $F = F_1 + \ldots + F_n$ mit $F_i(\omega) = 1$ falls $\omega(i) = i$ und $F_i(\omega) = 0$ sonst.
(1a) Es gilt $\mathbf{E}(F_i) = 1/n$, dank Linearität der Erwartung also $\mathbf{E}(F) = 1$.
Zur Varianz $\mathbf{V}(F) = \mathbf{E}(F^2) - \mathbf{E}(F)^2$ gilt $F^2 = \sum_i F_i^2 + 2\sum_{i<j} F_i F_j$.
Hierbei ist $F_i^2 = F_i$, also $\mathbf{E}(F_i^2) = 1/n$, sowie $\mathbf{E}(F_i F_j) = 1/n^2$.
Somit gilt $\mathbf{V}(F) = \mathbf{E}(F^2) - \mathbf{E}(F)^2 = n(n-1)/n^2 = 1 - 1/n$.
(1b) Die Zufallsvariablen $F_1,\ldots,F_n$ sind unabhängig $\mathrm{B}(1,1/n)$–verteilt. Somit ist $F \sim \mathrm{B}(n,1/n)$ tatsächlich binomialverteilt. Für große n gilt die Näherung $\mathbf{P}_F \approx \mathrm{P}(1)$ dank Poissons Gesetz der kleinen Zahlen U3E.

Fixpunkte zufälliger Permutationen

W320
Ausführung

(2a) Wir nutzen weiterhin die Summenzerlegung $F = F_1 + \ldots + F_n$. Diese Zufallsvariablen auf $\Omega = \{\, \omega : I \to I \,\}$ schränken wir nun jedoch ein auf $\Omega' = \{\, \omega : I \xrightarrow{\sim} I \,\}$. Auch $(\Omega', \mathbf{P}')$ ist laplacesch, wobei $|\Omega'| = n!$. Auch hier gilt $\mathbf{E}(F_i) = 1/n$, dank Linearität der Erwartung also $\mathbf{E}(F) = 1$. Im Unterschied zu (1) gilt jedoch $\mathbf{E}(F_i F_j) = 1/n(n-1)$, somit $\mathbf{V}(F) = 1$.

☺ Einfache Merkregel: Eine zufällige Permutation hat 1 ± 1 Fixpunkte.

⚠ Auf $(\Omega', \mathbf{P}')$ sind die Zufallsvariablen $F_1,\ldots,F_n$ nicht unabhängig! Wegen $\mathbf{Cov}(F_i,F_i) = \mathbf{E}(F_i F_j) - \mathbf{E}(F_i)\mathbf{E}(F_j) = 1/n^2(n-1) > 0$ sind sie positiv korreliert, wie anschaulich zu erwarten, wenn auch nur schwach. Ihre Varianzen $\mathbf{V}(F_i)$ addieren sich nicht zu $\mathbf{V}(F)$, nur dank Korrektur gemäß $\mathbf{V}(F) = \sum_i \mathbf{V}(F_i) + 2\sum_{i<j} \mathbf{Cov}(F_i,F_j) = (1 - 1/n) + 1/n = 1$.

(2b) Die Zufallsvariable $F : (\Omega', \mathbf{P}') \to \mathbb{N}$ ist nicht exakt Poisson–verteilt, da sie nur ein endliches Bild $\{0,1,\ldots,n\}$ hat. Für $n \to \infty$ nimmt die Korrelation ab, und wir vermuten die Näherung $\mathbf{P}_F \approx \mathrm{P}(1)$ wie in (1). Numerisches Experiment oder sorgfältige Rechnung bestätigen dies. Die Normalverteilung $\mathrm{N}(1,1)$ ist offensichtlich eine schlechte Näherung.

Anwendungsbeispiel: radioaktiver Zerfall

Ein Atom Uran 238 zerfällt in der nächsten Stunde mit Wkt $2 \cdot 10^{-14}$.
Eine Probe von $20\mathrm{mg}$ enthält etwa $5 \cdot 10^{19}$ Atome. Wir nehmen Zerfälle als stochastisch unabhängig an, insbesondere ohne Kettenreaktion. U461

Aufgabe: Sei X die Gesamtzahl der Zerfälle in der nächsten Stunde.

(1) Nennen Sie die exakte Verteilung (in diesem vereinfachten Modell).

(2) Berechnen Sie Erwartung $\mu = \mathbf{E}(X)$, Varianz $\sigma^2 = \mathbf{V}(X)$, Streuung. Runden Sie jedes Ergebnis auf die nächstgelegene ganze Zahl.

(3) Wie klein ist der totale Abstand zur Poisson–Verteilung $\mathrm{P}(\mu)$?

(4) Wie klein ist der Näherungsfehler zur Normalverteilung $\mathrm{N}(\mu, \sigma^2)$?

(5) Bestimmen Sie die Wahrscheinlichkeit $\mathbf{P}\big(|X - \mu| \leq 2500\big)$ in Prozent, gerundet auf den nächstgelegenen Prozentpunkt.

(6) Ist unser Modell für Uran 238 realistisch? Warum / warum nicht?

(7) Würden Sie diese Rechnung ebenso durchführen bei Uran 235? (Spaltbares Material wie Uran 235 *kann* zur Kettenreaktion führen. Dies wird genutzt zum Beispiel als Kernreaktor oder als Kernwaffe.)

Anwendungsbeispiel: radioaktiver Zerfall

Lösung: (1) Es handelt sich um die Binomialverteilung $\mathrm{B}(n,p)$:

$$\mathbf{P}(X{=}k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 5 \cdot 10^{19}, \quad p = 2 \cdot 10^{-14}$$

(2) Erwartung, Varianz und Streuung von $\mathrm{B}(n,p)$ sind

$$\begin{aligned} \mu = \mathbf{E}(X) &= np = 5 \cdot 10^{19} \cdot 2 \cdot 10^{-14} &&= 10^6 \\ \sigma^2 = \mathbf{V}(X) &= np(1-p) = 10^6 - 2 \cdot 10^{-8} &&\approx 10^6 \\ \sigma = \sqrt{\mathbf{V}(X)} &= \sqrt{10^6 - 2 \cdot 10^{-8}} &&\approx \sqrt{10^6} = 10^3 \end{aligned}$$

(3) Es gilt $\big\|\mathbf{P}_X - \mathrm{P}(\mu)\big\| \leq np^2 = 5 \cdot 10^{19} \cdot 4 \cdot 10^{-28} = 2 \cdot 10^{-8}$.

(4) Im LGS bzw. ZGS gilt für den Approximationsfehler δ die Schranke

$$|\delta| \;\leq\; \frac{|1-2p|}{10\sigma} + \frac{1}{3\sigma^2} \;<\; \frac{1}{6\sigma} \;<\; \frac{1}{\sigma} \;=\; 10^{-3}.$$

Mancher verwendet die Grenzwertsätze nur gefühlt „nach Erfahrung“. Fehlerschranken sind verlässlicher als ein vages subjektives Gefühl. Sie helfen insbesondere auch dort, wo noch keine Erfahrung vorliegt!

Anwendungsbeispiel: radioaktiver Zerfall

(5) Dank (4) ist die Normalverteilung ausreichend genau:

$$\mathbf{P}\big(|X - \mu| \leq 2500\big) \approx \int_{-\alpha}^{\alpha} \varphi(t)\,\mathrm{d}t = 2 \cdot \int_{t=0}^{2.5} \varphi(t)\,\mathrm{d}t$$

$$\approx 2 \cdot 0.49379 \qquad = 0.98758 \qquad \approx 99\%$$

🙂 Wir entscheiden uns hier für den lokalen Grenzwertsatz V3A und nähern die Binomialverteilung durch die Normalverteilung. Dank unserer Fehlerabschätzung (4) ist das genau genug.

Genau genommen müssten wir mit $\alpha = 2500.5/\sigma = 2.5005$ rechnen, aber die Stetigkeitskorrektur ist hier klein und fällt nicht ins Gewicht.

🙂 Die Poisson–Verteilung wäre nach (3) eine noch bessere Näherung. Für $\mathrm{P}(1\,000\,000)$ haben wir hier allerdings keine Tabelle vorliegen.

Es gibt auch kumulative Tabellen zu Poisson–Verteilungen $\mathrm{P}(\mu)$; noch bequemer ist heutzutage die Nutzung eines Computers.

⚠ Auch Computer können $\mathbf{P}_X = \mathrm{B}(n,p)$ für $n = 5 \cdot 10^{19}$ in realistischer Zeit nicht direkt berechnen, etwa die exakte Summe $\sum_{k=a}^{b} \mathrm{B}(n,p)(k)$. Hierzu nutzen wir Näherungen wie $\mathbf{P}_X \approx \mathrm{N}(\mu, \sigma^2)$ oder $\mathbf{P}_X \approx \mathrm{P}(\mu)$.

Anwendungsbeispiel: radioaktiver Zerfall

(6) Ja. Zwischen Uran 238 findet nahezu keine Wechselwirkung statt.

(7) Nein. Die stochastische Unabhängigkeit der Zerfälle gilt nicht mehr!

⚠ Allein die größere Zerfallswkt $p \in [0,1]$ ändert noch nicht das Modell. Doch unser Modell $\mathrm{B}(n,p)$ beruht auf stochastischer *Unabhängigkeit*! Andernfalls ist die Anzahl X der Zerfälle nicht mehr binomialverteilt. U461

Eine fahrlässige, falsche Modellbildung kann zu Katastrophen führen! Zwar sind die Werte innerhalb des Modells noch rechnerisch richtig, aber das gewählte Modell beschreibt keineswegs die Wirklichkeit.

Die Grundlagen der Kernphysik gehören natürlich nicht zur HM3, aber soviel Anwendungsbezug gehört zur Allgemeinbildung für technisch Gebildete, gerade angehende Ingenieur:innen. Ein qualitatives Modell kennen Sie vielleicht aus der Schule: Die Kettenreaktion entsteht, weil beim Zerfall Neutronen emittiert werden, die von anderen Atomkernen absorbiert werden und diese zum Zerfall anregen. Das ist das Gegenteil von Unabhängigkeit! Otto Hahn (1879–1968) bekam für diese wichtige Entdeckung 1945 den Nobelpreis für Chemie. Ihre Anwendungen, zum Guten wie zum Bösen, prägten das 20. Jahrhundert und wirken fort.

Anwendung des ZGS auf den fairen Würfel

Aufgabe: Sie würfeln 100 mal mit einem fairen Würfel.
Mit welcher Wkt erhalten Sie eine Augensumme $S \geq 390$?
Was sagt Chebychev? der ZGS? der LGS? die exakte Rechnung?

Lösung: Für die Augenzahl X bei einem fairen Würfel gilt:

$$\mathbf{E}(X) = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6} = 21/6 = 3.5$$
$$\mathbf{E}(X^2) = 1^2 \cdot \tfrac{1}{6} + 2^2 \cdot \tfrac{1}{6} + 3^2 \cdot \tfrac{1}{6} + 4^2 \cdot \tfrac{1}{6} + 5^2 \cdot \tfrac{1}{6} + 6^2 \cdot \tfrac{1}{6} = 91/6$$
$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = 91/6 - 49/4 = 35/12 \approx 2.9167$$

Für einmaliges Würfeln erhalten wir also folgende Kenngrößen:

$$\mu(X) = 3.5, \quad \sigma^2(X) \approx 2.9167, \quad \sigma(X) \approx 1.7078$$

Für die Augensumme S bei 100 unabhängigen Würfen gilt demnach:

$$\mu(S) = 350, \quad \sigma^2(S) = 291.67, \quad \sigma(S) \approx 17.078$$

☺ Man bewundere die Magie der großen Zahlen!

Erwartungswert und Varianz werden mit $n = 100$ multipliziert, die Streuung jedoch nur mit $\sqrt{n} = 10$. Im Verhältnis zum Erwartungswert 350 wird die Streuung 17 deutlich geringer! Aus diesem Grund liefern größere Stichproben zuverlässigere Ergebnisse. Doch alles im Leben hat seinen Preis: Für 10mal höhere Genauigkeit benötigt man eine 100mal größere Stichprobe!

Anwendung des ZGS auf den fairen Würfel

Abweichung vom Erwartungswert: Für $S \geq 390 = \mu + \alpha\sigma$ gilt

$$\alpha = (390 - \mu)/\sigma = (390 - 350)/17 \approx 2.34.$$

Chebychev gilt allgemein und liefert eine erste grobe Abschätzung:

$$\mathbf{P}(S \geq 390) \leq \frac{1}{1+\alpha^2} \lessapprox 0.16$$

Der **LGS** lässt sich nicht anwenden, da S nicht binomialverteilt ist!
Der **zentrale Grenzwertsatz** nutzt die Normalverteilung als Näherung:

$$\mathbf{P}(S \geq 390) \approx \int_\alpha^\infty \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \approx 0.01$$

Hierzu genügt es, die Tabelle abzulesen: $\mathbf{P}(S \geq 390) \approx 0.5 - 0.49036 = 0.00964$.
Der ZGS garantiert nur die pessimistische Fehlerschranke $|\delta| < 13\%$, mit $\rho^3(X) = 6.375$.
Wir nehmen optimistisch an, dass die Normalverteilung hier wesentlich besser approximiert.
In diesem konkreten Fall können wir das durch eine exakte Rechnung direkt nachprüfen:

Exakte Rechnung: Durch mühsames Aufsummieren erhalten wir:

$$\mathbf{P}(S \geq 390) = 0.0102875\ldots \approx 1\%$$

Anwendung des ZGS auf gezinkten Würfel

Aufgabe: Sie würfeln 100 mal mit unserem gezinkten Würfel (T107).
Mit welcher Wkt erhalten Sie dabei eine Augensumme $S \geq 390$?
Was sagt Chebychev? der ZGS? der LGS? die exakte Rechnung?

Lösung: Für die Augenzahl Y bei unserem gezinkten Würfel gilt:

$$\mathbf{E}(Y) = 1 \cdot 0.15 + 2 \cdot 0.16 + 3 \cdot 0.16 + 4 \cdot 0.16 + 5 \cdot 0.16 + 6 \cdot 0.21 = 3.65$$
$$\mathbf{E}(Y^2) = 1^2 \cdot 0.15 + 2^2 \cdot 0.16 + 3^2 \cdot 0.16 + 4^2 \cdot 0.16 + 5^2 \cdot 0.16 + 6^2 \cdot 0.21$$
$$\mathbf{V}(Y) = \mathbf{E}(Y^2) - \mathbf{E}(Y)^2 = 16.35 - 13.3225 = 3.0275$$

Für einmaliges Würfeln erhalten wir also folgende Kenngrößen:

$$\mu(Y) = 3.65, \quad \sigma^2(Y) \approx 3.0275, \quad \sigma(Y) \approx 1.74$$

Für die Augensumme S bei 100 unabhängigen Würfen gilt demnach:

$$\mu(S) = 365, \quad \sigma^2(S) = 302.75, \quad \sigma \approx 17.4$$

☺ Man bewundere die Magie der großen Zahlen!

Erwartungswert und Varianz werden mit $n = 100$ multipliziert, die Streuung jedoch nur mit $\sqrt{n} = 10$. Im Verhältnis zum Erwartungswert 365 wird die Streuung 17.4 deutlich geringer! Aus diesem Grund liefern größere Stichproben zuverlässigere Ergebnisse. Doch alles im Leben hat seinen Preis: Für 10mal höhere Genauigkeit benötigt man eine 100mal größere Stichprobe!

Anwendung des ZGS auf gezinkten Würfel

Abweichung vom Erwartungswert: Für $S \geq 390 = \mu + \alpha\sigma$ gilt

$$\alpha = (390 - \mu)/\sigma = (390 - 365)/17.4 \approx 1.44.$$

Chebychev gilt allgemein und liefert eine erste grobe Abschätzung:

$$\mathbf{P}(S \geq 390) \leq \frac{1}{1+\alpha^2} \lessapprox 0.33$$

Der **LGS** lässt sich nicht anwenden, da S nicht binomialverteilt ist!
Der **zentrale Grenzwertsatz** nutzt die Normalverteilung als Näherung:

$$\mathbf{P}(S \geq 390) \approx \int_\alpha^\infty \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \approx 0.075$$

Hierzu genügt es, die Tabelle abzulesen: $\mathbf{P}(S \geq 390) \approx 0.5 - 0.42507 = 0.07493 \approx 7.5\%$
Der ZGS garantiert nur die pessimistische Fehlerschranke $|\delta| < 13\%$, mit $\rho^3(Y) = 6.745$.
Wir nehmen optimistisch an, dass die Normalverteilung hier wesentlich besser approximiert.
In diesem konkreten Fall können wir das durch eine exakte Rechnung direkt nachprüfen:

Exakte Rechnung: Durch mühsames Aufsummieren erhalten wir:

$$\mathbf{P}(S \geq 390) = 0.0794448\ldots \approx 8\%$$

Anwendung: fair oder gezinkt?

Aufgabe: Sie haben drei Würfel, zwei faire und unseren gezinkten (alle äußerlich gleich). Sie wählen einen zufällig und würfeln damit 100 mal.

(1) Mit welcher Wkt erhalten Sie dabei eine Augensumme $S \geq 390$?

(2) Angenommen $S \geq 390$: Mit welcher Wkt ist der Würfel gezinkt?

Lösung: (1) Die totale Wahrscheinlichkeit (T2B) ist:

$$\mathbf{P}(S \geq 390) = \mathbf{P}(S \geq 390|\text{fair})\,\mathbf{P}(\text{fair}) + \mathbf{P}(S \geq 390|\text{zink})\,\mathbf{P}(\text{zink})$$

$$\approx 0.01 \cdot 2/3 + 0.08 \cdot 1/3 = 1/30 \approx 3.33\%$$

(2) Nach der Formel von Bayes (T2B) gilt:

$$\mathbf{P}(\text{zink}|S \geq 390) = \frac{\mathbf{P}(S \geq 390|\text{zink})\,\mathbf{P}(\text{zink})}{\mathbf{P}(S \geq 390)} \approx \frac{0.08 \cdot 1/3}{1/30} = 80\%$$

Das Auftreten einer Augensumme ≥ 390 ist demnach ein starkes Indiz, dass der von uns gewählte Würfel nicht fair, sondern der gezinkte ist.

🙂 Leistungsstarke Theorie ermöglicht effiziente Berechnung.
Zu Klarheit, Verständnis und Korrektheit hilft eine präzise Notation.

Anwendung: fair oder gezinkt?

Wir frischen hier nochmals unsere Grundlagen aus Kapitel T auf.
Dank der Formel von Bayes sind solche Rechnungen meist leicht – und doch oft überraschend. Üben Sie sich mit diesem Werkzeug!
Berechnung und Interpretation des Ergebnisses erfordern Übung.

Ohne weitere Information ist der von Ihnen gewählte Würfel mit Wkt $1/3$ gezinkt und mit Wkt $2/3$ fair. Nach Durchführung des Tests mit dem Ergebnis $S \geq 390$ ist er mit deutlich höherer Wkt von 80% gezinkt.
Das ist plausibel; die genaue Wkt erfordert präzise Rechnung.

Wir können nicht 100% sicher sein, aber der Test spricht dafür, dass der Würfel nicht fair ist. Ist eine höhere Sicherheit nötig, dann sollten Sie weitere Tests durchführen. Eine größere Datenreihe liefert größere Genauigkeit bzw. größere Sicherheit (wenn auch nie 100%).

Dieses einfache Anwendungsbeispiel ist leicht, doch repräsentativ.
Stochastik (WRechnung und Statistik) ist die Mathematik des Zufalls.
Unser Ziel dabei sind rationale Entscheidungen unter Unsicherheit.
Wir suchen nachvollziehbar begründete, quantitative Aussagen.

Anwendung: ZGS oder LGS?

Aufgabe: In einer Geburtsklinik beträgt die Wahrscheinlichkeit für eine Jungengeburt $t = 0.514$. Wie groß ist die Wahrscheinlichkeit, dass von den nächsten $n = 1000$ Neugeborenen genau $k = 514$ Jungen sind?

Was sagt Chebychev? der ZGS? der LGS? die exakte Rechnung?

Lösung: Wir haben $S \sim \mathrm{B}(n,t)$. Wir suchen $\mathbf{P}(S = k)$. Es gilt:

$$\mu = \mathbf{E}(S) = nt = 514, \quad \sigma^2 = \mathbf{V}(S) = nt(1-t) \approx 249.8, \quad \sigma \approx 15.8$$

⚠ Chebychev nützt für Intervalle um μ, aber nicht für einzelne Punkte!
Das gilt auch für den ZGS. Mangels Stetigkeitskorrektur liefert er nur

$$\mathbf{P}\big(k \leq S \leq k\big) = \int_{(k-\mu)/\sigma}^{(k-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \quad + \quad \delta$$

Pech: Das Integral verschwindet! Die Fehlerschranke liefert immerhin

$$0 \leq \mathbf{P}\big(k \leq S \leq k\big) = \delta \leq \frac{1}{2\sqrt{nt(1-t)}} \lessapprox 0.032.$$

☹ Die Ungleichung stimmt, aber damit sind wir noch nicht zufrieden!

Anwendung: ZGS oder LGS?

🙂 Der LGS ist besser: Dank seiner Stetigkeitskorrektur liefert er

$$\mathbf{P}\big(k \leq S \leq k\big) = \binom{n}{k} t^k (1-t)^{n-k} \quad \approx \int_{(k-1/2-\mu)/\sigma}^{(k+1/2-\mu)/\sigma} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi$$

$$\approx \frac{\mathrm{e}^{-(k-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \quad \overset{(k\equiv\mu)}{=} \frac{1}{\sigma\sqrt{2\pi}} \quad \approx \quad 0.0252412\ldots$$

Die Fehlerschranke ist wegen $\kappa \approx 0$ hier $|\varepsilon| < 1/(3\sigma^2) < 0.0014$.

🙂 Noch besser: Die exakte Rechnung ist hier direkt und mühelos!

$$\mathbf{P}\big(S = k\big) = \binom{1000}{514} (0.514)^{514}\,(0.486)^{486} = 0.0252349\ldots$$

Was lernen wir daraus? Augen auf bei der Methodenwahl!

Diese bemerkenswerten Grenzwertsätze bilden einen umfangreichen Werkzeugkasten: Sie liefern mächtige Werkzeuge und ganz praktische Rechenregeln. Um sie zu nutzen, muss man die Voraussetzungen kennen und die Schlussfolgerungen verstehen. Es lohnt sich!

Anwendung: Fehlerwahrscheinlichkeiten

Aufgabe: Ein Elektroniksystem habe $n = 10^6$ unabhängige Bauteile mit den Fehlerwkten p_k. Wir wissen nur $10^{-4} \le p_k \le 10^{-2}$ laut Spezifikation. Sie interessiert die Gesamtzahl $S = \sum_{k=1}^n X_k$ der auftretenden Fehler.

Wir denken an eine mehrjährige Marsmission, Störungen durch Sonnenwind, oder ähnliches. Fehler sind zahlreich, daher wollen wir ausreichend Redundanz für ein solches System. Ein explizites Beispiel für $n = 3$ haben wir auf Seite T409 ausführlich diskutiert.

(1) Berechnen Sie Erwartung $\mu = \mathbf{E}(S)$ und Varianz $\sigma^2 = \mathbf{V}(S)$.
Ist hier die Verteilung durch die Kenngrößen (μ, σ) bereits festgelegt?

$$\mu = \mathbf{E}(S) = \sum_{k=1}^n \mathbf{E}(X_k) = \sum_{k=1}^n p_k \quad \in \quad [100, 10\,000]$$

Da die $X_1, \dots, X_n$ unabhängig sind, addieren sich auch die Varianzen:

$$\sigma^2 = \mathbf{V}(S) = \sum_{k=1}^n \mathbf{V}(X_k) = \sum_{k=1}^n p_k(1-p_k) \quad \in \quad [99.99, 9900].$$

🙂 Wegen $(1 - p_k) \in [0.99, 1]$ gilt hier $\sigma^2 \approx \mu$ bis auf 1% genau.

⚠ Die exakte Verteilung ist allein durch (μ, σ) noch nicht festgelegt.

Anwendung: Fehlerwahrscheinlichkeiten

(2) Welche Modelle beschreiben die Verteilung von S exakt / approx.?
Binomial? Poisson? LGS? ZGS? mit welchen Fehlerschranken?

☹ Die Summe ist nicht binomialverteilt, da die p_k verschieden sind.

Erinnerung: Die Binomialverteilung $\mathrm{B}(n, p)$ entsteht bei n unabhängigen Experimenten mit identischer Trefferwkt p. Das ist hier nicht der Fall, also ist dieses Modell eher ungeeignet. Allgemein siehe `de.wikipedia.org/wiki/Verallgemeinerte_Binomialverteilung`.

🙂 Poissons Gesetz der kleinen Zahlen liefert $\mathbf{P}_S \approx \mathrm{P}(\mu)$ mit Fehler

$$\|\mathbf{P}_S - \mathrm{P}(\mu)\| \le p_1^2 + \ldots + p_n^2 \quad \in \quad [0.01, 100].$$

☹ Der LGS lässt sich nicht anwenden, da S nicht binomialverteilt ist.

🙂 Der ZGS ist anwendbar und liefert $\mathbf{P}_S \approx \mathrm{N}(\mu, \sigma^2)$ mit Fehler

$$|\delta| \le \frac{\rho_1^3 + \ldots + \rho_n^3}{(\sigma_1^2 + \ldots + \sigma_n^2)^{3/2}} \approx \frac{\mu}{\mu^{3/2}} = \frac{1}{\sqrt{\mu}} \quad \in \quad [0.01, 0.1].$$

Daher nutzen wir den ZGS für große p_k und Poisson für kleine p_k.

Die einzelnen Summanden von $S = X_1 + \ldots + X_n$ sind leicht zu verstehen: $X_k \sim \mathrm{B}(1, p_k)$ hat Erwartung $\mu_k = \mathbf{E}(X_k) = p_k$, Varianz $\sigma_k^2 = \mathbf{V}(X_k) = p_k(1-p_k) \approx p_k$, drittes Moment $\rho_k^3 = \mathbf{E}(|X_k - \mu_k|^3) = p_k(1-p_k)(p_k^2 + (1-p_k)^2) \approx p_k$. Trotz der großen Zahl $n = 10^6$ ist die Fehlerschranke nicht besonders gut. Für Poisson sieht's allerdings auch nicht besser aus.

Anwendung: Fehlerwahrscheinlichkeiten

(3a) Wie beurteilen Sie einen Testlauf mit mehr als $11\,000$ Fehlern?
Im Extremfall $\mu = 10\,000$ und $\sigma = 100$ liefert Chebychev ganz grob

$$\mathbf{P}(S \ge 11\,000) \le \frac{1}{1+k^2} < 0.01 \quad \text{mit} \quad k = 10.$$

Für die Normalverteilung sind 10σ–Abweichungen unwahrscheinlich:

$$\int_{10}^{\infty} \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi < 10^{-23}, \quad \text{aber leider nur} \quad |\delta| \le 1/\sqrt{\mu} \in [0.01, 0.1].$$

Wenn wir der Normalverteilung glauben, dann ist eine 10σ–Abweichung praktisch unmöglich: Sie ist unwahrscheinlicher als dreimal nacheinander sechs Richtige im Lotto zu tippen!

Exakt liegt hier allerdings nicht die Normalverteilung $\mathrm{N}(\mu, \sigma^2)$ vor, daher müssen wir zusätzlich den Approximationsfehler berücksichtigen: Unsere Schranke aus (2) ist hier leider recht grob.

Für $\mu < 10\,000$ wird alles nur noch schlimmer. Nach Beobachtung von $11\,000$ Fehlern scheint das Modell jedenfalls nicht mehr glaubwürdig! Dann ist die Annahme $p_k \le 10^{-2}$ nicht haltbar oder die Unabhängigkeit nicht gegeben: Vielleicht beeinflusst ein Fehler die anderen (z.B. durch lokale Überhitzung) oder externe Einflüsse provozieren massenhafte Fehler (z.B. Strahlung). Dem Modell liegen wichtige Annahmen zu Grunde; diese müssen kritisch überprüft werden.

Anwendung: Fehlerwahrscheinlichkeiten

(3b) Wie beurteilen Sie einen Testlauf mit etwa $10\,100$ Fehlern?

Das ist etwa mit $\mu = 10\,000$ und $\sigma = 100$ durchaus vereinbar.

(4) Durch $N = 25$ Testläufe ermitteln Sie $\hat{x} \approx 3600$ und $\sigma^2 \approx 3600$.
Welches Konfidenzintervall überdeckt μ mit 95%iger Sicherheit?

Wir haben die Streuung $\sigma \approx 60$ und $\sqrt{N} = 5$, also $2\sigma/\sqrt{N} = 24$.
Als 2σ–Intervall finden wir $[3576, 3624]$. Das ist bereits recht genau.

(5) Wie viele Testläufe brächten eine 2σ–Genauigkeit von ± 10?

Wir verlangen $2\sigma/\sqrt{N} \le 10$, im oben Beispiel $\sigma \approx 60$ also $N \ge 144$.
Das ist genauer, aber vielleicht auch zu aufwändig, das heißt zu teuer.

Vielleicht ist das Intervall aus (4) oder noch gröber bereits gut genug; das ist eine strategische Entscheidung und hängt vom Kontext ab, vor allem von der weiteren Verwendung dieser Daten und Schätzungen.

Wir suchen nachvollziehbar begründete, quantitative Aussagen.
🙂 Mit den Werkzeugen der Mathematik sind Sie dafür gewappnet.

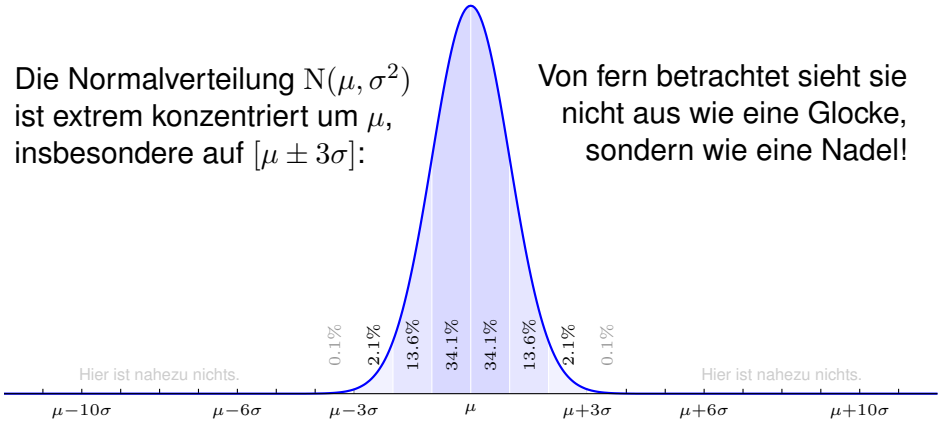
Wie selten sind seltene Ereignisse?

„We were seeing things that were 25-standard deviation moves, several days in a row.“, sagte der damalige Finanzvorstand von Goldman-Sachs, David Viniar, der Financial Times im August des Krisenjahres 2007.

Aufgabe: Ist das plausibel? Meint er ernsthaft die Normalverteilung? Sollte er sein Modell für diese Wahrscheinlichkeiten überdenken?

Die Normalverteilung $\mathrm{N}(\mu,\sigma^2)$ ist extrem konzentriert um μ, insbesondere auf $[\mu\pm 3\sigma]$:

Von fern betrachtet sieht sie nicht aus wie eine Glocke, sondern wie eine Nadel!



Wie selten sind seltene Ereignisse?

Lösung: Für die Normalverteilung ist das sicher nicht plausibel!
Die Wkt für $k\sigma$–Abweichungen nach Normalverteilung und Chebychev:

$$p_k \;=\; \int_{x=k}^{\infty} \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x \;\le\; \frac{1}{1+k^2}$$

Chebychev gilt für alle Verteilungen. Die Wkt $< 1/2500 = 0.04\%$ ist klein, und das auch noch mehrfach hintereinander! Das wäre extremes Pech.
Die genaueren Werte von p_k entnehmen wir der (erweiterten) Tabelle:

$k=$	2	3	4	5	6
$p_k \lessapprox$	0.02275	0.00135	$3.2\cdot10^{-5}$	$2.9\cdot10^{-7}$	$9.9\cdot10^{-10}$

$k=$	7	10	15	20	25
$p_k \lessapprox$	$1.3\cdot10^{-12}$	$7.7\cdot10^{-24}$	$3.7\cdot10^{-51}$	$2.8\cdot10^{-87}$	$3.1\cdot10^{-136}$

Letzteres ist unwahrscheinlicher als 20mal nacheinander im Lotto sechs Richtige zu tippen. Das genannte Modell scheint unplausibel / unseriös.

Durch mathematische Kompetenz können Sie überzeugen und glänzen, aber durch Inkompetenz machen Sie sich unglaubwürdig und lächerlich.

Anwendung: ZGS oder LGS?

Wir untersuchen die Summe $S = X_1 + \ldots + X_n$ von n unabhängigen ZVariablen $X_1,\ldots,X_n : \Omega \to \mathbb{R}$. Als Verteilungen betrachten wir dabei:

(1) $X_k \sim \mathrm{N}(\mu_k,\sigma_k^2)$ normalverteilt mit $\mu_k,\sigma_k \in \mathbb{R}$

(2) $X_k \sim \mathrm{B}(1,t_k)$ mit $t_k \in [0.1,0.9]$ abhängig von k

(3) $X_k \sim \mathrm{B}(1,t)$ mit $t \in [0.1,0.9]$ unabhängig von k

(4) $X_k \sim \mathrm{B}(N_k,t)$ mit $N_k \in \mathbb{N}$ und $t \in [0.1,0.9]$

(5) Denken Sie sich zehn weitere, ebenso konkrete Beispiele aus, etwa jede Zufallsvariable $X_k \sim \mathrm{P}(\mu_k)$ Poisson–verteilt oder $X_k \sim G(q_k)$ geometrisch verteilt oder $X_k \sim \mathrm{H}(N_k,K_k,n_k)$ hypergeometrisch, etc.

Aufgabe: (a) Auf welche Summen S lässt sich der ZGS anwenden?

(b) Auf welche lässt sich zudem der speziellere LGS anwenden?

(c) Was lässt sich jeweils über den Approximationsfehler δ sagen?

(d) In welchen Fällen kennen Sie die exakte Verteilung von S?

Anwendung: ZGS oder LGS?

Lösung: (1) Der ZGS lässt sich anwenden, aber nicht der LGS da die Summe S nicht binomialverteilt ist. Sie ist hier exakt normalverteilt:

$$S \sim \mathrm{N}(\mu,\sigma^2) \quad \text{mit} \quad \mu = \sum_{k=1}^{n} \mu_k \quad \text{und} \quad \sigma^2 = \sum_{k=1}^{n} \sigma_k^2$$

Der Approximationsfehler ist demnach $\delta = 0$, siehe folgende Aufgaben.

(2) Der ZGS lässt sich anwenden, der LGS jedoch nicht, da die Summe S nicht binomialverteilt ist. Den Approximationsfehler δ entnehmen wir Satz W1D. Ein solches Beispiel diskutiert die obige Aufgabe W333.

(3) Die exakte Verteilung ist hier die Binomialverteilung $S \sim \mathrm{B}(n,t)$. Der ZGS lässt sich anwenden, der LGS auch und liefert eine bessere Näherung. Den Fehler δ bzw. korrigiert ε entnehmen wir Satz V3A.

(4) Die exakte Verteilung ist hier $S \sim \mathrm{B}(N,t)$ mit $N = N_1 + \ldots + N_n$. Der ZGS lässt sich anwenden, der LGS auch, wenn man es richtig anstellt. Zur Summe von binomialverteilten ZVariablen siehe unten.

Summe normalverteilter Zufallsvariablen

W341
Ausführung

Aufgabe: Die Zufallsvariablen $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ und $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ seien unabhängig und normalverteilt mit den angegebenen Parametern.

Bestimmen Sie für die Summe $S = X + Y$ den Erwartungswert $\mathbf{E}(S)$ und die Varianz $\mathbf{V}(S)$ sowie die Verteilung auf $\mathbb{R}$. Ist S normalverteilt?

Lösung: Dank Linearität addieren sich die Erwartungswerte:

$$\mathbf{E}(S) = \mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y) = \mu_1 + \mu_2$$

Dank Unabhängigkeit addieren sich auch die Varianzen:

$$\mathbf{V}(S) = \mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) = \sigma_1^2 + \sigma_2^2$$

Schließlich berechnen wir die Verteilung von S dank Unabhängigkeit:

$$\mathrm{N}(\mu_1, \sigma_1^2) * \mathrm{N}(\mu_2, \sigma_2^2) = \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Das ist keineswegs offensichtlich! Sie können die Faltung geduldig direkt ausrechnen oder noch geschickter mit Fourier–Transformation. K417

Summe normalverteilter Zufallsvariablen

W342
Ausführung

Aufgabe: Seien $X_1, X_2, X_3, \ldots$ unabhängige Zufallsvariablen, und jede sei normalverteilt gemäß $X_k \sim \mathrm{N}(\mu_k, \sigma_k^2)$ mit $\mu_k, \sigma_k \in \mathbb{R}$.

Bestimmen Sie für die Summe $S = X_1 + \ldots + X_n$ die Erwartung $\mathbf{E}(S)$ und die Varianz $\mathbf{V}(S)$ sowie die Verteilung auf $\mathbb{R}$. Gilt hier der ZGS?

Es darf gelacht werden: Wie groß ist der Approximationsfehler δ?

Lösung: Dank Linearität addieren sich die Erwartungswerte:

$$\mathbf{E}(S) = \mathbf{E}(X_1 + \ldots + X_n) = \mathbf{E}(X_1) + \ldots + \mathbf{E}(X_n) = \mu_1 + \ldots + \mu_n =: \mu$$

Dank Unabhängigkeit addieren sich auch die Varianzen:

$$\mathbf{V}(S) = \mathbf{V}(X_1 + \ldots + X_n) = \mathbf{V}(X_1) + \ldots + \mathbf{V}(X_n) = \sigma_1^2 + \ldots + \sigma_n^2 =: \sigma^2$$

Aus der vorigen Aufgabe erhalten wir die Verteilung $S \sim \mathrm{N}(\mu, \sigma^2)$.
(Wir haben dies für zwei Summanden explizit nachgerechnet.
Für n Summanden folgt es dann offensichtlich durch Induktion.)

🙂 Hier gilt der ZGS exakt, d.h. mit Approximationsfehler $\delta = 0$.

Summe binomialverteiler Zufallsvariablen

W343
Ausführung

Aufgabe: Seien $X \sim \mathrm{B}(1, p)$ und $Y \sim \mathrm{B}(1, q)$ unabhängig mit $p, q > 0$. Bestimmen Sie für die Summe $S = X + Y$ die Erwartung $\mathbf{E}(S)$ und die Varianz $\mathbf{V}(S)$ sowie die Verteilung auf $\{0, 1, 2\}$. Ist S binomialverteilt?

Lösung: Dank Linearität addieren sich die Erwartungswerte:

$$\mathbf{E}(S) = \mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y) = p + q$$

Dank Unabhängigkeit addieren sich auch die Varianzen:

$$\mathbf{V}(S) = \mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) = p(1-p) + q(1-q).$$

Schließlich berechnen wir die Verteilung von S dank Unabhängigkeit:

$$\mathbf{P}(S{=}0) = (1-p)(1-q),\ \mathbf{P}(S{=}1) = p(1-q) + (1-p)q,\ \mathbf{P}(S{=}2) = pq$$

Wäre dies binomialverteilt, so hätten wir $S \sim \mathrm{B}(2, t)$ für ein $t \in [0, 1]$.
Der Vergleich von $\mathbf{E}(S) = 2t = p + q$ und $\mathbf{P}(S{=}2) = t^2 = pq$ zeigt $p = q$.
Eine Binomialverteilung erhalten wir demnach nur für den Fall $p = q$!
`de.wikipedia.org/wiki/Verallgemeinerte_Binomialverteilung`
Im Sonderfall $p = 0$ gilt $S \sim \mathrm{B}(1, q)$; für $q = 0$ gilt $S \sim \mathrm{B}(1, p)$.

Summe binomialverteiler Zufallsvariablen

W344
Ausführung

Aufgabe: Seien $X \sim \mathrm{B}(m, t)$ und $Y \sim \mathrm{B}(n, t)$ unabhängig, $m, n \in \mathbb{N}$.
Bestimmen Sie für $S = X + Y$ die Erwartung $\mathbf{E}(S)$ und Varianz $\mathbf{V}(S)$
Ist die Summe $S = X + Y$ binomialverteilt? Mit welchen Parametern?

Lösung: Wir wissen $\mathbf{E}(X) = mt$ und $\mathbf{V}(X) = mt(1-t)$, entsprechend $\mathbf{E}(Y) = nt$ und $\mathbf{V}(Y) = nt(1-t)$. Dank Linearität der Erwartung folgt

$$\mathbf{E}(S) = \mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y) = (m+n)t.$$

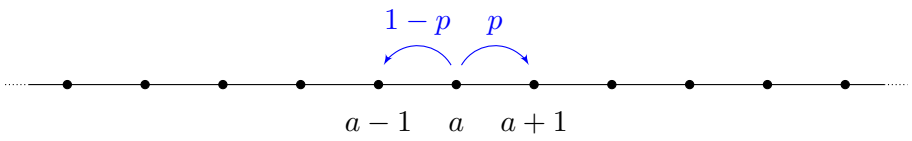
Dank Unabhängigkeit addieren sich auch die Varianzen:

$$\mathbf{V}(S) = \mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) = (m+n)t(1-t).$$

Wie finden wir geschickt die Verteilung von S? durch direkte Rechnung?
Wir nutzen die Darstellung $X = X_1 + \ldots + X_m$ und $Y = Y_1 + \ldots + Y_n$
als Summe unabhängiger Zufallsvariablen $X_i, Y_j \sim \mathrm{B}(1, t)$.
Hieraus folgt $S \sim \mathrm{B}(m+n, t)$ ohne weitere Rechnung.

🙂 Dieser Rechentrick vereinfacht die Aufgabe ungemein! Zur expliziten Faltung nutzen wir dankend gerne erzeugende Funktionen (Satz U4E).

Zufällige Irrfahrt / Random Walk



Zufällige Irrfahrt: Zur Zeit $t = 0$ starten Sie im Punkt $S_0 = a$.

Im Schritt von S_t nach S_{t+1} gehen Sie mit Wahrscheinlichkeit $p \in [0,1]$ nach rechts und entsprechend mit Wahrscheinlichkeit $(1-p)$ nach links.

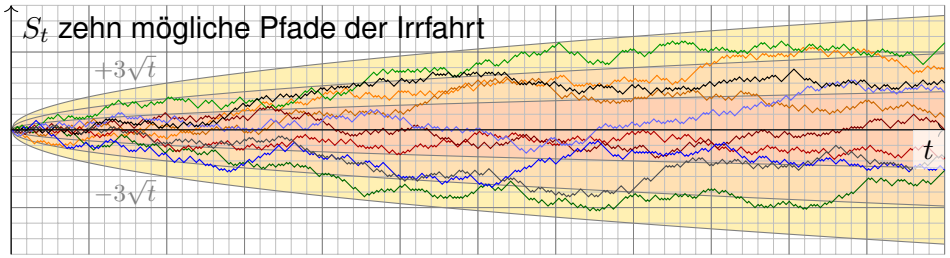
Das heißt, $S_t : \Omega \to \mathbb{Z}$ ist gegeben durch $S_t = a + X_1 + \ldots + X_t$ mit unabhängigen Zuwächsen, $\mathbf{P}_a(X_t{=}{+1}) = p$ und $\mathbf{P}_a(X_t{=}{-1}) = 1-p$.

Das ist ein einfaches aber wichtiges Modell. Mögliche Anwendung: Kontostand bei zufälligen Gewinnen und Verlusten. Daher werden solche Irrfahrten als Modelle für **Aktienkurse** genutzt.

Ähnlich entsteht die **Brownsche Bewegung** durch Wärmebewegung. Der schottische Botaniker Robert Brown (1773–1858) entdeckte 1827 unter dem Mikroskop das unregelmäßige Zittern von Pollen in Wasser. Anfangs hielt er Pollen für belebt, doch er fand dasselbe bei Staubteilchen.

Albert Einstein erklärte die Zitterbewegung 1905 durch die ungeordnete Wärmebewegung der Wassermoleküle, die aus allen Richtungen in großer Zahl gegen die Pollen stoßen. Quantitativ konnte er so die Größe von Atomen bestimmen und die Anzahl pro Mol, die **Avogadro–Zahl**. Die präzisen quantitativen Vorhersagen wurden in den Folgejahren experimentell bestätigt.

Zufällige Irrfahrt / Random Walk



Aufgabe: (1) Bestimmen Sie zu S_t die Verteilung, Erwartung, Streuung. Wir untersuchen speziell den symmetrischen Fall $p = 1/2$, ohne Drift.

(2) Sie beginnen im Startpunkt $a \in \mathbb{Z}$ und fixieren einen Zielpunkt $z \in \mathbb{Z}$. Wie groß ist die Wkt $u(a) \in [0,1]$, das Ziel z irgendwann zu erreichen?

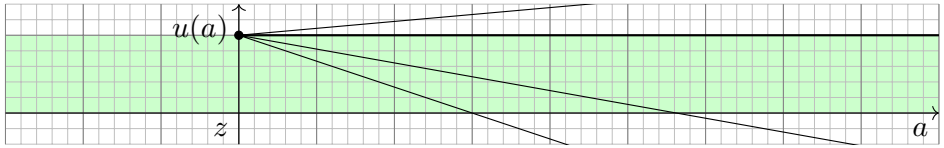
(3) Wie groß ist hierbei die erwartete Reisezeit von a nach z?

Lösung: (1) Wir erhalten eine Binomialverteilung, affin transformiert: $\mathbf{P}_a\big(S_t = a - t + 2k\big) = \binom{t}{k} p^k (1-p)^{t-k} = \binom{t}{k} \frac{1}{2^t}$ für $t, k \in \mathbb{N}$ und $p = \frac{1}{2}$.

Somit gilt $\mathbf{E}(S_t) = a$ und $\mathbf{V}(S_t) = t$, also $\sigma(S_t) = \sqrt{t}$ und $S_t \approx \mathrm{N}(a,t)$.

☺ Wir erkennen die Lösung D5D der Wärmeleitungsgleichung!

Zufällige Irrfahrt / Random Walk



(2) Offensichtlich gilt $u(z) = 1$, denn hier ist der Start auch das Ziel.
Für $a > z$ gilt die Mittelwerteigenschaft $u(a) = \frac{1}{2}u(a+1) + \frac{1}{2}u(a-1)$.
Somit ist $u : \mathbb{Z}_{\geq z} \to [0,1]$ eine Gerade, $u(a) = 1 + m(a-z)$. (Warum?)
Zudem ist u beschränkt, $0 \leq u \leq 1$, daher folgt $m = 0$. Ebenso auf $\mathbb{Z}_{\leq z}$.

Die Gerade finden wir durch vollständige Induktion: Aus $u(z) = 1$ und $u(z+1) = 1 + m$ folgt $u(a+1) = 2u(a) - u(a-1) = [2 + 2m(a-z)] - [1 + m(a-1-z)] = 1 + m(a+1-z)$.

☺ Bei einer symmetrischen Irrfahrt ($p = 1/2$, ohne Drift) erreichen wir jeden Punkt mit Wkt 1! George Pólya (1887–1985) zeigte 1921: Jeden Punkt in $\mathbb{Z}$ besuchen wir mit Wkt 1 unendlich oft. Dies gilt ebenso in Dimension 2 bei Irrfahrt auf dem ebenen Gitter $\mathbb{Z}^2$. Erstaunlicherweise gilt es nicht mehr in Dimension $n \geq 3$ bei Irrfahrt auf dem Gitter $\mathbb{Z}^n$. Anschaulich bedeutet das: Ein betrunkener Mensch findet sicher irgendwann nach Hause, ein betrunkener Fisch hingegen nicht!

📖 Ausführung bei Feller, *Introduction to Probability*, vol. 1 (1968), §XIV.7: Das Sprichwort „Alle Wege führen nach Rom." stimmt zumindest zweidimensional. Dreidimensional ist die Rückkehrwahrscheinlichkeit nur etwa 34%, siehe `en.wikipedia.org/wiki/Random_walk`.

Zufällige Irrfahrt / Random Walk

(3) Sei $T_z \in \mathbb{N}$ die Zeit des ersten Besuchs im Zielpunkt z und $\mathbf{E}_a(T_z)$ die erwartete Reisezeit vom Startpunkt a zum Zielpunkt z.
Dies ist invariant unter Verschiebungen, also $\mathbf{E}_{a+k}(T_{z+k}) = \mathbf{E}_a(T_z)$.

Zunächst gilt $\mathbf{E}_a(T_a) = 0$. Für $a \neq z$ zeigen wir nun $\mathbf{E}_a(T_z) = \infty$:

$$x := \mathbf{E}_0(T_1) = \tfrac{1}{2}\big[\underbrace{\mathbf{E}_1(T_1)}_{=0} + \mathbf{E}_{-1}(T_1)\big] + 1 = \tfrac{1}{2}\mathbf{E}_{-1}(T_1) + 1$$

$$\mathbf{E}_{-1}(T_1) = \underbrace{\mathbf{E}_{-1}(T_0)}_{=x} + \underbrace{\mathbf{E}_0(T_1)}_{=x} = 2x$$

Hieraus folgt $x = \frac{1}{2}(2x) + 1 = x + 1$. Das ist für $x \in \mathbb{R}$ unmöglich.
Es bleibt nur $\mathbf{E}_0(T_1) = x = \infty$, erst recht $\mathbf{E}_0(T_z) = \infty$ für alle $z \neq 0$.

☺ Bei einer symmetrischen Irrfahrt ($p = 1/2$, ohne Drift) erreichen wir jeden Punkt mit Wkt 1, aber die erwartete Reisezeit ist unendlich!

Die Rechnung ist einfach, dank unserer geschickten Formalisierung.
Die Interpretation hingegen muss man erst einmal verarbeiten.
Die naive Anschauung kann einen hier leicht narren.

Laplace–Transformation von WVerteilungen

☺ Die Laplace–Transformation ist in vielen Bereichen ein nützliches Werkzeug, insbesondere bei Differentialgleichungen, siehe Kapitel L. Mathematische Abstraktion hilft: *Learn once, apply everywhere!*

☺ Erfreulicherweise lassen sich solche Integraltransformationen auch in der WTheorie ganz wunderbar nutzen. Diese effiziente und elegante Rechentechnik will ich im Folgenden kurz skizzieren.

Motivation: Zu $X : \Omega \to \mathbb{N}$ haben wir die erzeugende Funktion (U4E)

$$G(z) = \mathbf{E}\big[z^X\big] \quad \text{zumindest für } |z| \le 1.$$

In der Wahrscheinlichkeitsrechnung treten erzeugende Funktionen $G(z) = \mathbf{E}\big[z^X\big]$ natürlich auf, wenn man Zufallsvariablen $X : \Omega \to \mathbb{N}$ effizient behandeln will. Wir haben dies in Beispielen ausgeführt.

Entsprechendes wünschen wir uns für Zufallsvariablen $X : \Omega \to \mathbb{R}_{\ge 0}$. Der Erwartungswert $\mathbf{E}(z^X)$ existiert für $z = 1$, denn $\mathbf{E}(1^X) = \mathbf{E}(1) = 1$. Für $0 < z \le 1$ gilt die Majorante $0 < z^X \le 1$, also existiert auch $\mathbf{E}\big[z^X\big]$. Wir parametrisieren $z = \mathrm{e}^{-s} \in \,]0,1]$ durch $s \in \mathbb{R}_{\ge 0}$ und vereinbaren:

Laplace–Transformation von WVerteilungen

Definition W4A: Laplace–Transformierte eines WMaßes auf $\mathbb{R}_{\ge 0}$

Für jedes WMaß $\mathbf{P}_X$ auf $\mathbb{R}_{\ge 0}$ ist seine **Laplace–Transformierte**:

$$F : \mathbb{R}_{\ge 0} \to \mathbb{R} : s \mapsto F(s) := \mathbf{E}\big[\mathrm{e}^{-sX}\big] = \int_{x=0}^{\infty} \mathrm{e}^{-sx} \,\mathrm{d}\mathbf{P}_X$$

Für eine kontinuierliche Verteilung $\mathbf{P}_X$ mit Dichte f erhalten wir also

$$F(s) = \int_{x=0}^{\infty} \mathrm{e}^{-sx} f(x) \,\mathrm{d}x.$$

Die Majorante $|\mathrm{e}^{-sx}| \le 1$ sichert absolute Integrierbarkeit und $|F(s)| \le 1$.
Allgemeiner erhalten wir so die stetige Funktion $F : \mathbb{C}_{\mathrm{Re}\ge 0} \to \mathbb{C}$.
Sie ist holomorph auf der offenen Halbebene $\mathbb{C}_{\mathrm{Re}>0}$.

☺ Hier tritt die Laplace–Transformierte $F(s) = \mathbf{E}\big[\mathrm{e}^{-sX}\big]$ natürlich auf, um kontinuierliche Zufallsvariablen $X : \Omega \to \mathbb{R}_{\ge 0}$ effizient zu behandeln. Dies möchte ich mit einem kurzen Rechenbeispiel illustrieren.

Laplace–Transformation von WVerteilungen

Aufgabe: (1) Für die Laplace–Transformierte $F(s) = \mathbf{E}\big[\mathrm{e}^{-sX}\big]$ gilt

$$F(0) = \mathbf{E}(1) = 1, \quad F'(0) = -\mathbf{E}(X), \quad F''(0) = \mathbf{E}(X^2).$$

(2) Berechnen Sie $F(s)$ sowie $\mathbf{E}(X)$ und $\mathbf{V}(X)$ für $X \sim E(\lambda)$, also die Exponentialverteilung mit Dichte $f(x) = \lambda\,\mathrm{e}^{-\lambda x}$ für $x \ge 0$.

Lösung: (1) Wir nehmen an, dass diese Erwartungswerte existieren:

$$\mathbf{E}(X) = \int_{x=0}^{\infty} x f(x) \,\mathrm{d}x, \quad \mathbf{E}(X^2) = \int_{x=0}^{\infty} x^2 f(x) \,\mathrm{d}x.$$

Wir dürfen dann die Ableitung unter das Integral ziehen und erhalten

$$F'(s) = \int_{x=0}^{\infty} -x\,\mathrm{e}^{-sx} f(x) \,\mathrm{d}x, \quad F''(s) = \int_{x=0}^{\infty} x^2\,\mathrm{e}^{-sx} f(x) \,\mathrm{d}x.$$

Für $s = 0$ ergeben sich hieraus obige Formeln. Sehr praktisch!

☺ Die Laplace–Transformierte ist sogar holomorph, siehe Satz L1B. Erneut lohnt sich Ihre Investition in solide mathematische Grundlagen.

Laplace–Transformation von WVerteilungen

(2) Sei X exponentialverteilt gemäß $\mathrm{E}(\lambda)$ mit $\lambda > 0$, mit Dichte

$$f : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0} : x \mapsto f(x) = \lambda\,\mathrm{e}^{-\lambda x}.$$

Ihre Laplace–Transformierte rechnen wir direkt aus:

$$\begin{aligned} F(s) &= \int_{x=0}^{\infty} \mathrm{e}^{-sx} f(x) \,\mathrm{d}x = \int_{x=0}^{\infty} \mathrm{e}^{-sx} \cdot \lambda\,\mathrm{e}^{-\lambda x} \,\mathrm{d}x \\ &= \lambda \int_{x=0}^{\infty} \mathrm{e}^{-(s+\lambda)x} \,\mathrm{d}x = \frac{\lambda}{s+\lambda}\Big[-\mathrm{e}^{-(s+\lambda)x}\Big]_{x=0}^{\infty} = \frac{\lambda}{s+\lambda} \end{aligned}$$

Plausibilitätscheck: Es gilt tatsächlich $F(0) = 1$, so wie es sein muss.
Aus der Funktion F berechnen wir leicht Erwartungswert und Varianz:

$$F'(s) = \frac{-\lambda}{(s+\lambda)^2} \quad \Longrightarrow \quad \mathbf{E}(X) = -F'(0) = \frac{1}{\lambda},$$

$$F''(s) = \frac{2\lambda}{(s+\lambda)^2} \quad \Longrightarrow \quad \mathbf{E}(X^2) = F''(0) = \frac{2}{\lambda^2}.$$

Wir erhalten $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{1}{\lambda^2}$, wie zuvor berechnet. V243

Fourier–Transformation von WVerteilungen

Motivation: Die vorigen Rechnungen sind sehr effizient und elegant.
Zur Zufallsvariablen $X : \Omega \to \mathbb{N}$ nutzen wir die erzeugende Funktion

$$G(z) = \mathbf{E}\big[z^X\big] = \sum_{k=0}^{\infty} \mathbf{P}(X{=}k) \cdot z^k \quad \text{zumindest für } |z| \leq 1.$$

Allgemein für $X : \Omega \to \mathbb{Z}$, eventuell mit negativen Werten, haben wir

$$G(z) = \mathbf{E}\big[z^X\big] = \sum_{k=-\infty}^{\infty} \mathbf{P}(X{=}k) \cdot z^k \quad \text{zumindest für } |z| = 1.$$

Mit diesen Potenzreihen können wir wunderbar arbeiten, siehe U4D.
Die erzeugende Funktion $G(z)$ ist eine konzise Schreibweise für die (beliebig komplizierte) Verteilung der Zufallsvariablen $X : \Omega \to \mathbb{Z}$.

Entsprechendes wünschen wir uns auch für Zufallsvariablen $X : \Omega \to \mathbb{R}$.
Der Erwartungswert $\mathbf{E}(z^X)$ existiert für $|z| = 1$, denn $\mathbf{E}(1^X) = \mathbf{E}(1) = 1$.
Wir parametrisieren daher $z = \mathrm{e}^{-\mathrm{i}\xi} \in \mathbb{S}^1$ durch $\xi \in \mathbb{R}$ und vereinbaren:

Fourier–Transformation von WVerteilungen

Definition W4B: Fourier–Transformierte eines WMaßes auf $\mathbb{R}$

Für jedes WMaß $\mathbf{P}_X$ auf $\mathbb{R}_{\geq 0}$ ist seine **Fourier–Transformierte**

$$F : \mathbb{R} \to \mathbb{C} : \xi \mapsto F(\xi) := \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X}\big] = \int_{x \in \mathbb{R}} \mathrm{e}^{-\mathrm{i}\xi x} \, \mathrm{d}\mathbf{P}_X$$

Für eine kontinuierliche Verteilung $\mathbf{P}_X$ mit Dichte f erhalten wir also

$$F(\xi) = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\xi x} f(x) \, \mathrm{d}x.$$

Die Majorante $|\mathrm{e}^{-\mathrm{i}\xi x}| \leq 1$ sichert absolute Integrierbarkeit und $|F(s)| \leq 1$.

⚠ Man beachte die hier gewählte Konvention der Normierung:
Zur Bequemlichkeit lasse ich den Normierungsfaktor $1/\sqrt{2\pi}$ weg.
In der Rücktransformation benötigen wir dann den Faktor $1/2\pi$.

🙂 Fourier $\mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X}\big]$ ist ein Spezialfall von Laplace $\mathbf{E}\big[\mathrm{e}^{-sX}\big]$ eingeschränkt auf die imaginäre Achse $s = \mathrm{i}\xi$ mit $\xi \in \mathbb{R}$.

Fourier–Transformation von WVerteilungen

Wiederholung: Nennen Sie die Wahrscheinlichkeitsdichte $f(x)$ einer normalverteilten Zufallsvariablen $X \sim \mathrm{N}(\mu, \sigma^2)$ sowie die Fourier–Transformierte $F(\xi) = \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X}\big]$ und ihre Rücktransformation.

Lösung: Für die Normalverteilung $\mathrm{N}(\mu, \sigma^2)$ wissen wir K128

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \Longrightarrow \quad F(\xi) = \mathrm{e}^{-\mathrm{i}\mu\xi - \sigma^2\xi^2/2}.$$

Durch Rücktransformation von F lässt sich f rekonstruieren:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\xi) \, \mathrm{e}^{\mathrm{i}\xi x} \, \mathrm{d}\xi$$

⚠ Man beachte die hier gewählte Konvention der Normierung.

🙂 Die Fourier–Transformation $f \mapsto F$ verliert keine Information:
Die Rücktransformation rekonstruiert die ursprüngliche Verteilung.
Erneut lohnt sich Ihre Investition in solide mathematische Grundlagen.

Fourier–Transformation von WVerteilungen

🙂 Die Fourier–Transformation ist ein mächtiges Universalwerkzeug.
Wir haben dieser wichtigen Technik daher ein eigenes Kapitel gewidmet.
Die dort erlernten Rechenmethoden zahlen sich bei der Behandlung vieler Probleme aus, so auch hier für die Wahrscheinlichkeitsrechnung.
Zudem haben wir wichtige Beispiele und Erfahrungen gesammelt, die in vielen Problemstellungen erneut und abgewandelt auftreten.

🙂 Erzeugende Funktionen $\mathbf{E}[z^X]$, Laplace–Transformierte $\mathbf{E}[\mathrm{e}^{-sX}]$ und Fourier–Transformierte $\mathbf{E}[\mathrm{e}^{-\mathrm{i}\xi X}]$ sind nützliche Rechenmethoden.
Sie entspringen alle derselben natürlichen Idee, wie hier skizziert:
Transformation reorganisiert alle Informationen auf praktische Weise.
Die Momente der Verteilung entsprechen nun Ableitungen, so werden insbesondere Erwartung und Varianz rechnerisch leicht zugänglich.

Durch Rücktransformation lässt sich jeweils die ursprüngliche Verteilung rekonstruieren, es geht also tatsächlich keinerlei Information verloren.
Zur Illustration will ich als Hauptergebnis den zentralen Grenzwertsatz (ZGS W1D) mit der Fourier–Transformation erklären und nachrechnen.

Erwartungswert und Varianz

Aufgabe: Berechnen Sie $F(0)$, $F'(0)$, $F''(0)$ und $F'''(0)$ sowie hiermit für $\mu = 0$ die Taylor–Entwicklung zweiter Ordnung von $F(\xi)$ um $\xi = 0$.

Lösung: Existieren die Momente $\int_{\mathbb{R}} |x|^k f(x)\,\mathrm{d}x < \infty$ für $k = 1, 2, 3, \ldots$, so dürfen wir für $F^{(k)}$ die Ableitung unters Integral ziehen K205:

$$\begin{aligned} F'(\xi) &= \frac{\mathrm{d}}{\mathrm{d}\xi} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}\xi x} f(x)\,\mathrm{d}x &&= \int_{\mathbb{R}} \frac{\partial}{\partial \xi} \mathrm{e}^{-\mathrm{i}\xi x} f(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}} (-\mathrm{i}x)\,\mathrm{e}^{-\mathrm{i}\xi x} f(x)\,\mathrm{d}x \quad &&\Longrightarrow \quad F'(0) = -\mathrm{i}\mathbf{E}(X) \\ F''(\xi) &= \int_{\mathbb{R}} (-x^2)\,\mathrm{e}^{-\mathrm{i}\xi x} f(x)\,\mathrm{d}x \quad &&\Longrightarrow \quad F''(0) = -\mathbf{E}(X^2) \\ F'''(\xi) &= \int_{\mathbb{R}} (\mathrm{i}x^3)\,\mathrm{e}^{-\mathrm{i}\xi x} f(x)\,\mathrm{d}x \quad &&\Longrightarrow \quad F'''(0) = \mathrm{i}\mathbf{E}(X^3) \end{aligned}$$

☺ Für $\mu = 0$ und $\sigma^2 > 0$ hat F demnach die Taylor–Entwicklung

$$F(\xi) = 1 - \sigma^2\xi^2/2 + \ldots$$

Das dritte Moment liefert eine Fehlerschranke $O(\xi^3)$ für das Restglied.
📖 Ausführung bei Feller, *Introduction to Probability*, vol. 2 (1971), §XVI

Summe unabhängiger Zufallsvariablen

Lemma W4C: Summe unabhängiger Zufallsvariablen
Seien $X, Y : \Omega \to \mathbb{R}$ unabhängige ZV mit Transformierten F_X, F_Y.
Die Summe $X + Y$ hat dann die Transformierte $F_{X+Y} = F_X \cdot F_Y$.

Nachrechnen: Auch $\mathrm{e}^{-\mathrm{i}\xi X}$ und $\mathrm{e}^{-\mathrm{i}\xi Y}$ sind unabhängig, also gilt

$$\begin{aligned} F_{X+Y}(\xi) &= \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi(X+Y)}\big] \\ &= \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X} \cdot \mathrm{e}^{-\mathrm{i}\xi Y}\big] \\ &= \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X}\big] \cdot \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi Y}\big] \;=\; F_X(\xi) \cdot F_Y(\xi) \end{aligned}$$

☺ Das ist der entscheidende Trick! Er vereinfacht unsere Rechnung.

Damit haben wir alle Werkzeuge, um den zentralen Grenzwertsatz zu beweisen! Ich formuliere ihn in vereinfachter Form für identisch verteilte, unabhängige Zufallsvariablen. Vor allem aber zeige ich die Konvergenz ohne eine explizite Fehlerschranke, um alle Argumente so einfach wie möglich zu halten. Für eine detaillierte Ausarbeitung müsste man die Fehlerterme kontrollieren, die hier nur mit „. . . “ abgekürzt werden.

Beweis des zentralen Grenzwertsatzes

Seien $X_1, X_2, X_3, \ldots : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen.
Es gelte $\mathbf{E}(X_k) = \mu$ und $\mathbf{E}(X_k^2) = \sigma^2 > 0$ sowie $\mathbf{E}(|X_k|^3) \le \rho^3 < \infty$.
Die Summe $S_n = X_1 + \ldots + X_n$ hat Erwartung $n\mu$ und Varianz $n\sigma^2$.
Wir normieren S_n zu Erwartung 0 und Varianz 1 durch die Skalierung

$$S_n^* := \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Wir wollen nun $\mathbf{P}_{S_n} \approx \mathrm{N}(n\mu, n\sigma^2)$ zeigen, normiert also $\mathbf{P}_{S_n^*} \approx \mathrm{N}(0,1)$.
Fehlerschranken sind für die Praxis wichtig, aber schwerer zu zeigen.
In dieser Skizze begnügen wir uns mit dem Grenzwert für $n \to \infty$.

Satz W4D: zentraler Grenzwertsatz, ohne Fehlerschranken
Die Verteilung von S_n^* konvergiert gegen die Normalverteilung $\mathrm{N}(0,1)$.

☺ Das ist die qualitative Aussage $\mathbf{P}_S \approx \mathrm{N}(n\mu, n\sigma^2)$ aus Satz W1D.
Die quantitative Fehlerabschätzung lassen wir hier außer Acht.

Beweis des zentralen Grenzwertsatzes

Aufgabe: (1) Berechnen Sie die Fourier–Transformierte F_n von S_n^*.
(2) Wie folgt hieraus der zentrale Grenzwertsatz für $n \to \infty$?
(3) Warum geht das plötzlich so leicht?

Beweis: (1) Die Transformierte von $\frac{X_k - \mu}{\sigma\sqrt{n}}$ ist $F_k(\xi) = 1 - \xi^2/2n + \ldots$.
Die Transformierte der Summe S_n^* ist das Produkt der Transformierten:

$$F_n(\xi) = \big(1 - \xi^2/2n + \ldots\big)^n \quad \to \quad \exp(-\xi^2/2) = F(\xi) \quad \text{für} \quad n \to \infty.$$

(2) Der Grenzwert F ist die Transformierte der Normalverteilung $\mathrm{N}(0,1)$!
Rücktransformation: Die Verteilung von S_n^* konvergiert gegen $\mathrm{N}(0,1)$.

(3) Das geht leicht dank starker Werkzeuge wie Fourier–Transformation!
Zwecks Vereinfachung verzichten wir zudem auf Fehlerabschätzungen.

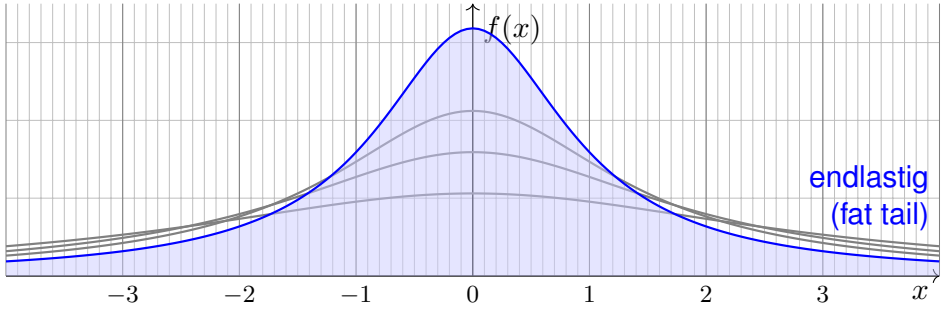
☺ Damit können wir den ZGS zumindest qualitativ nachrechnen:
Für großes n ist $X_1 + \ldots + X_n$ annähernd $\mathrm{N}(n\mu, n\sigma^2)$–verteilt.

Fourier–Transformation der Cauchy–Verteilung

Aufgabe: Ist für $a > 0$ die folgende Funktion eine WDichte?

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto f(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}$$

Bestimmen Sie Erwartung und Varianz und Fourier–Transformierte.



⚠ Der Graph erinnert grob an die Glockenkurve $\varphi(x) = \mathrm{e}^{-x^2/2}/\sqrt{2\pi}$.
Doch Vorsicht: Die Cauchy–Verteilung hat völlig andere Eigenschaften!
Vor allem klingt $f(x)$ nicht exponentiell ab, sondern nur wie $x \mapsto 1/x^2$.
Das hat dramatische Konsequenzen: Die Erwartung existiert nicht!

Fourier–Transformation der Cauchy–Verteilung

Lösung: Wir haben $f \geq 0$ und die Gesamtmasse ist

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x \;=\; \left[\frac{1}{\pi}\arctan\left(\frac{x}{a}\right)\right]_{x=-\infty}^{\infty} \;=\; 1$$

Für Erwartung und Varianz benötigen wir folgende Integrale:

$$\mathbf{E} = \frac{1}{\pi}\int_{\mathbb{R}} \frac{ax}{a^2+x^2}\,\mathrm{d}x = ? \qquad \text{und} \qquad \mathbf{V} = \frac{1}{\pi}\int_{\mathbb{R}} \frac{ax^2}{a^2+x^2}\,\mathrm{d}x = \infty.$$

Das erste ist nicht absolut integrierbar, nur als Cauchy–Hauptwert.
Der Median ist offensichtlich 0, aber die Erwartung existiert nicht!
Hier haben Erwartungswert und Varianz demnach keinen Sinn.

Die Fourier–Transformierte haben wir bereits berechnet K113:

$$F(\xi) \;=\; \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X}\big] \;=\; \int_{-\infty}^{\infty} \frac{1}{\pi}\,\mathrm{e}^{-\mathrm{i}\xi x} \cdot \frac{a}{a^2+x^2}\,\mathrm{d}x \;=\; \mathrm{e}^{-a|\xi|}$$

🙂 Die ursprüngliche WDichte f rekonstruieren wir hieraus durch Rücktransformation. Man beachte, dass F in 0 nicht differenzierbar ist.
Dies entspricht der Tatsache, dass der Erwartungswert $\mathbf{E}$ nicht existiert.

Fourier–Transformation der Cauchy–Verteilung

Die **Cauchy–Verteilung** $\mathrm{C}(a)$ mit $a > 0$ ist gegeben durch die Dichte

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto f(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}.$$

Aufgabe: Seien $X \sim \mathrm{C}(a)$ und $Y \sim \mathrm{C}(b)$ unabhängig Cauchy–verteilt.
Ist ihre Summe $S = X + Y$ Cauchy–verteilt? Mit welchem Parameter?

Lösung: Wir wissen $\mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X}\big] = \mathrm{e}^{-a|\xi|}$ und ebenso $\mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi Y}\big] = \mathrm{e}^{-b|\xi|}$.
Dank Unabhängigkeit von X und Y gilt für die Summe $S = X + Y$ somit

$$\begin{aligned}
\mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi S}\big] &= \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi (X+Y)}\big] \\
&= \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X} \cdot \mathrm{e}^{-\mathrm{i}\xi Y}\big] \\
&= \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi X}\big] \cdot \mathbf{E}\big[\mathrm{e}^{-\mathrm{i}\xi Y}\big] \;=\; \mathrm{e}^{-a|\xi|} \cdot \mathrm{e}^{-b|\xi|} \;=\; \mathrm{e}^{-(a+b)|\xi|}
\end{aligned}$$

Durch Rücktransformation gewinnen wir hieraus die Verteilung von S:
Demnach ist die Summe ebenfalls Cauchy–verteilt gemäß $\mathrm{C}(a+b)$.
Sie können dies auch direkt aus dem Faltungsintegral nachrechnen.

🙂 Die Cauchy–Verteilungen $(\mathrm{C}(a))_{a \in \mathbb{R}_{>0}}$ sind stabil unter Faltung!

Fourier–Transformation der Cauchy–Verteilung

Seien $X_1, \ldots, X_n \sim \mathrm{C}(a)$ unabhängig und identisch Cauchy–verteilt.
Wie im ZGS betrachten wir die (wie üblich normierte) Summe

$$S_n^* := \frac{1}{\sqrt{n}}\big(X_1 + \ldots + X_n\big).$$

Aufgabe: Welche Verteilung $\mathbf{P}_{S_n^*}$ hat die Zufallsvariable S_n^*?
Konvergiert $\mathbf{P}_{S_n^*}$ für $n \to \infty$ gegen eine Normalverteilung?

Lösung: Die Summe $X_1 + \ldots + X_n$ ist $\mathrm{C}(na)$–verteilt.
Durch den Normierungsfaktor $1/\sqrt{n}$ ist S_n^* dann $\mathrm{C}(a)$–verteilt.
Für $n \to \infty$ ist die Verteilung fix und konvergiert somit gegen $\mathrm{C}(a)$.
Insbesondere konvergiert sie keineswegs gegen eine Normalverteilung!

⚠ Die Cauchy–Verteilungen $\mathrm{C}(a)$ erfüllen nicht die Voraussetzungen des ZGS (W1D): Die geforderten Momente μ, σ, ρ existieren nicht!
Daher lässt sich der zentrale Grenzwertsatz hier nicht anwenden.

🙂 Unsere explizite Rechnung zeigt tatsächlich genauer noch mehr:
Die Verteilung der Summe konvergiert nicht gegen die Normalverteilung!

Was ist Entropie?

W417

The most misunderstood concept in physics

Veritasium, youtu.be/DxL2HoqLbyA

Hauptsätze der Thermodynamik

W418

1) *Die Energie der Welt ist constant.*
2) *Die Entropie der Welt strebt einem Maximum zu.*

Rudolf Clausius (1822–1888), *Mechanische Wärmetheorie* (1865)

You may see a cup of tea fall off a table and break into pieces on the floor... But you will never see the cup gather itself back together and jump back on the table. The increase of disorder, or entropy, is what distinguishes the past from the future, giving a direction to time.

Stephen Hawking (1942–2018), *A Brief History of Time* (1988)

The fact that you can remember yesterday but not tomorrow is because of entropy. [...] So I think that entropy is underappreciated as something that has a crucial role in how we go through life.

Sean M. Carroll (1966–), *The Passage of Time and the Meaning of Life*

youtu.be/KFXYCrqUTmI, ausführlich /EntaPoSEXog

Energie und Entropie in der Thermodynamik

W419
Erläuterung

Die **Wärmelehre** ist eine wissenschaftlich-technische Errungenschaft des 19. Jahrhunderts und prägt unsere hochtechnisierte Zivilisation.

(1) Die **technische Thermodynamik** beschreibt **makroskopische Phänomene** durch Temperatur, Wärme, Druck, Volumen, Arbeit, etc. und ihre Änderung in Prozessen (Hauptsätze, Zustandsgleichungen). Sie ist grundlegend zur Konstruktion und Analyse von Maschinen.

Ihre grobe Sichtweise (aka Abstraktion) ist didaktische Herausforderung und methodische Stärke; dadurch wird sie extrem vielseitig anwendbar auf große Klassen von Systemen und Prozessen.

(2) Die **statistische Physik** untersucht die **mikroskopischen Zustände** der Teilchen (Moleküle). Sie erklärt die makroskopischen Phänomene durch klassische Mechanik (Energie, Impuls, etc.) und dient so als theoretisches Fundament der Thermodynamik (Mikrofundierung).

Analogie in der Wirtschaftslehre: Die **Makroökonomik** untersucht übergeordnete Größen und Kennzahlen: Investition & Konsum, Export & Import, Staatsausgaben & Steuern, Geldpolitik, etc. Die **Mikroökonomik**, insbesondere die Spieltheorie, untersucht das Verhalten einzelner Akteure (Menschen, Haushalte, Unternehmen). Auch dies dient der Mikrofundierung, also dem detaillierten mikroökonomischen Verständnis der makroökonmisch beobachtbaren Phänomene.

Energie und Entropie in der Thermodynamik

W420
Erläuterung

0. Hauptsatz: Die Temperatur $T > 0$ eines Systems (im thermischen Gleichgewicht) beschreibt die mittlere Energie ungeordneter Bewegung.

Die Thermodynamik handelt von Temperatur, Wärme und Entropie. Allgemein beschreibt sie Vielteilchensysteme und hat daher eine fundamentale Stellung in Naturwissenschaft und Technik. Gegenüber Systemen mit wenigen Teilchen zeigen sich hier völlig neue Erscheinungen wie die Irreversibilität fast aller makroskopischen Prozesse. Die Entropie kennt nur eine Richtung!

1. Hauptsatz: Die Energie U bleibt erhalten gemäß $\mathrm{d}U = \mathrm{d}Q + \mathrm{d}W$.

Es gibt kein Perpetuum Mobile erster Art: Keine Maschine kann aus nichts Arbeit verrichten. Die Thermodynamik wirkt als Systemtheorie: Jedes Teilsystem wird als Black Box durch nur wenige Parameter beschrieben (Temperatur T, Druck p, Volumen V, etc.) und wechselwirkt durch Ströme (Energie $\mathrm{d}U$, Arbeit $\mathrm{d}W = p\,\mathrm{d}V +$ weitere, Wärme $\mathrm{d}Q$, Entropie $\mathrm{d}S$, etc.)

2. Hauptsatz: Die Entropie S wächst gemäß $\mathrm{d}S = \mathrm{d}Q/T \geq 0$.

In jedem abgeschlossenen System kann die Entropie nur anwachsen: $S_1 - S_0 = \int_{t_0}^{t_1} \mathrm{d}Q/T \geq 0$. Bestenfalls bleibt sie konstant, bei reversiblen Prozessen. Das hat weitreichende Konsequenzen: **Carnot:** Wärmeenergie kann niemals vollständig in mechanische Arbeit umgewandelt werden. **Kelvin, Planck:** Es gibt kein Perpetuum Mobile zweiter Art. Keine Maschine kann mechanische Arbeit verrichten allein durch Abkühlung eines Wärmespeichers.

3. Hauptsatz: Im idealen Grundzustand gilt $U = 0$ und $T = 0$ und $S = 0$.

Was ist Entropie? . . . in mikroskopischer Sicht?

Wir betrachten einen endlichen WRaum (Ω, P) mit $\Omega = \{\omega_1, \ldots, \omega_n\}$ als M „Lostopf“, P „Zustandsraum“ oder I „Quelle zufälliger Nachrichten“.
Das Ergebnis ω_i hat Wkt $p_i \in [0,1]$ und Überraschung $-\log_2 p_i \in [0,\infty]$.

p	1	$1/2$	$1/4$	$1/8$	. . .	0
$-\log_2 p$	0	1	2	3	. . .	∞

Das misst den Informationsgehalt der Nachricht ω_i in Bit, zur Basis 2.
Jede andere Basis b skaliert die Information gemäß $\log_b p = \ln(p)/\ln(b)$.
Unabhängigkeit heißt Wkten multiplizieren, Überraschungen addieren.

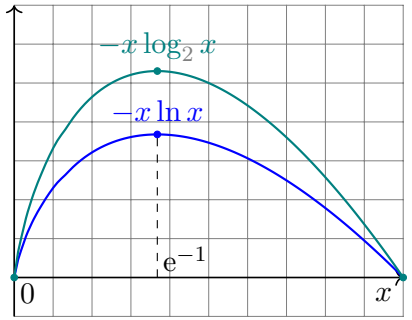
Die **Entropie** ist die erwartete Überraschung oder mittlere Information:

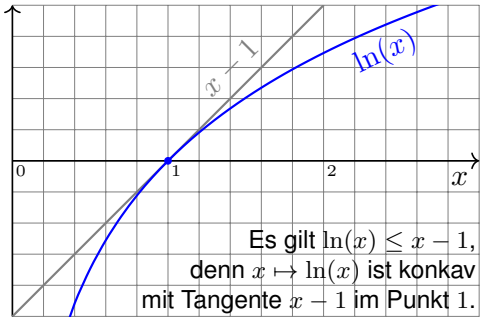
$$\mathrm{H}_2(P) := \mathbf{E}_P\big[\omega_i \mapsto -\log_2 p_i\big] = -\sum_{i=1}^{n} p_i \log_2 p_i \quad \in \mathbb{R}_{\geq 0}$$

Beispiel: Münzwurf $\Omega = \{\,0 = \text{Kopf},\ 1 = \text{Zahl},\ 2 = \text{Rand}\,\}$
mit den Wkten $q = (\,0.50,\ 0.50,\ 0.00\,)$ und $p = (\,0.49,\ 0.49,\ 0.02\,)$.

$$\begin{aligned} \mathrm{H}(Q) &\overset{\text{Def}}{=} 0.50 \cdot 1.00 + 0.50 \cdot 1.00 + 0.00 \cdot \infty &&= 1.00 \\ \mathrm{H}(P) &\approx 0.49 \cdot 1.03 + 0.49 \cdot 1.03 + 0.02 \cdot 5.64 &&\approx 1.12. \end{aligned}$$

Eintropie nach Clausius, Boltzmann, Gibbs, Shannon, . . .



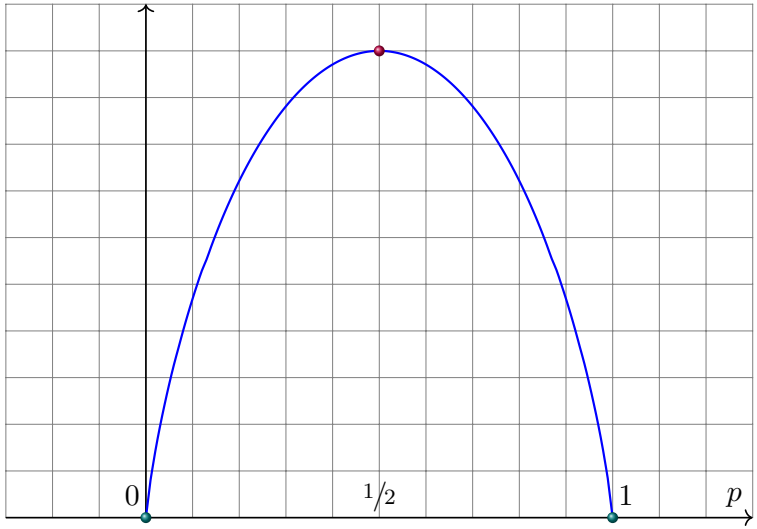


Eingeführt wurde die Entropie 1865 von Rudolf Clausius im 2. Hauptsatz der Thermodynamik. Ludwig Boltzmann und Willard Gibbs erklärten sie 1875 in der statistischen Mechanik durch $S = -\mathrm{k}\sum_i p_i \ln(p_i)$ mit den Wkten p_i aller Mikrozustände i, skaliert durch die Boltzmann–Konstante $\mathrm{k} = 1.38 \times 10^{-23}$ J/K. Claude Shannon klärte 1948 optimale Codierung von Information durch die Entropie und gründete die Informationstheorie.
📖 Thomas M. Cover, Joy A. Thomas: *Information Theory.* Wiley 2006
Entropie spielt überall eine zentrale Rolle, von Quanten über Strömung zu schwarzen Löchern, von Kommunikation zu maschinellem Lernen.

Welche WVerteilung hat maximale Entropie?

Auf $\Omega = \{1,2\}$ hat die WVerteilung $P := (p, 1-p)$ die Entropie

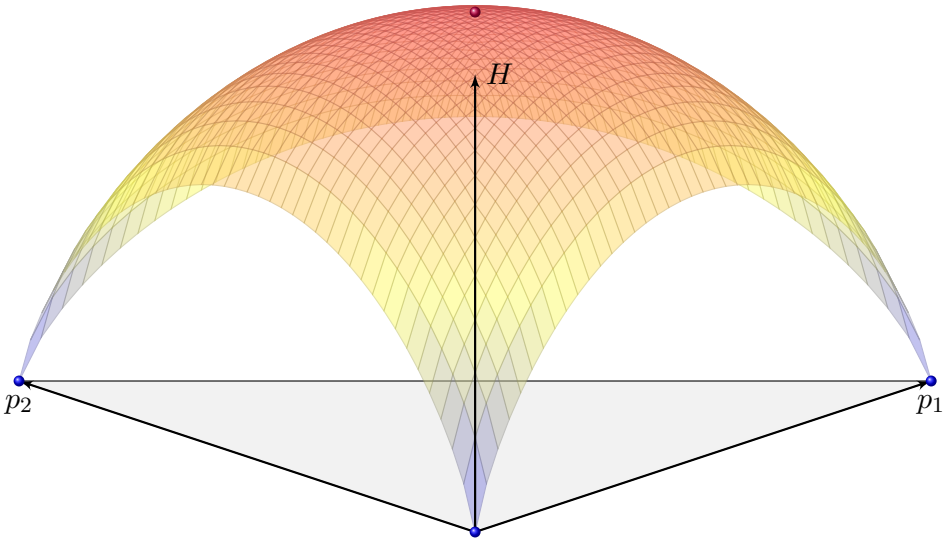
$$\mathrm{H}(P) = -p\log(p) - (1-p)\log(1-p).$$



Welche WVerteilung hat maximale Entropie?

Auf $\Omega = \{1,2,3\}$ hat die WVerteilung $P := (p_1, p_2, p_3)$ die Entropie

$$\mathrm{H}(P) = -p_1\log p_1 - p_2 \log p_2 - p_3 \log p_3.$$



Die Kreuzentropie $\mathrm{H}(P,Q)$ misst die Überraschung.

Oft schätzen wir die wahre Verteilung P (engl. *ground truth*) durch ein Modell Q. Das Ergebnis ω_i hat Wkt p_i, aber die Überraschung $-\log q_i$ als unser Maß der Information. Erstere ist objektiv, also physikalisch real, zweitere subjektiv, eine Eigenschaft unseres Modells der Wirklichkeit.

$$\mathrm{H}(P,Q) := \underbrace{\mathbf{E}_P\big[\omega_i \mapsto -\log q_i\big]}_{\text{experimentell messbar}} = -\sum_{i=1}^{n} p_i \log q_i \quad \in [0,\infty]$$

Die **Kreuzentropie** von P und Q ist auch hier die „mittlere Information“ oder die „erwartete Überraschung“ des Modells Q in der Wirklichkeit P. Diese kann durch Stichprobe geschätzt werden (Monte Carlo Methode).

▶ *The Key Equation Behind Probability.* youtu.be/KHVR587oW8I

Beispiel: Weiter sei $q = (0.50,\ 0.50,\ 0.00)$ und $p = (0.49,\ 0.49,\ 0.02)$.

$$\mathrm{H}(P,Q) \;=\; 0.49\cdot 1.00 + 0.49\cdot 1.00 + 0.02\cdot\infty \;=\; \infty$$

Ist das Modell $q' = (0.46,\ 0.46,\ 0.08)$ näher an der Realität?

$$\mathrm{H}(P,Q') \;\approx\; 0.49\cdot 1.12 + 0.49\cdot 1.12 + 0.02\cdot 3.64 \;\approx\; 1.17$$
$$\mathrm{H}(P,P) \;\approx\; 0.49\cdot 1.03 + 0.49\cdot 1.03 + 0.02\cdot 5.64 \;\approx\; 1.12$$

Die relative Entropie $\mathrm{D}(P,Q)$ misst die Abweichung.

Die **relative Entropie** des Modells Q bezüglich der Wirklichkeit P, oder **Kullback–Leibler–Divergenz**, ist die „zusätzliche Überraschung“:

$$\mathrm{D}(P,Q) \;:=\; \underbrace{\mathrm{H}(P,Q)}_{\text{exp. messbar}} - \underbrace{\mathrm{H}(P)}_{\text{konstant}} \;=\; \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \quad \in [0,\infty]$$

Divergenz: Gilt $\mathrm{D}(P,Q') < \mathrm{D}(P,Q)$, so weicht Q' weniger ab als Q.

Satz W4E: Gibbs' Informationsungleichung

Es gilt $\mathrm{D}(P,Q) \geq 0$, also $\mathrm{H}(P,Q) \geq \mathrm{H}(P)$, mit Gleichheit nur für $P = Q$.

Beweis: Für $x \in \mathbb{R}_{\geq 0}$ gilt $\ln(x) \leq x - 1$, mit Gleichheit gdw $x = 1$. Also:

$$-\textstyle\sum_i p_i \ln(q_i/p_i) \overset{(1)}{\geq} -\sum_i p_i (q_i/p_i - 1) \overset{(2)}{\geq} \sum_i p_i - q_i \overset{\text{Wkt}}{=} 0$$

In den ersten beiden Summen setzen wir $0 \cdot (\pm\infty) := 0$. Hier zählen also nur Indizes i mit $p_i > 0$. In der dritten Summe zählen alle Indizes i, auch mit $p_i = 0$, daher die Ungleichung (2). Gleichheit gilt genau dann, wenn $p_i = q_i$ für alle i gilt: Für $p_i > 0$ folgt dies aus (1), für $p_i = 0$ aus (2). □

Maximale Entropie: die diskrete Gleichverteilung

☺ Auf $\Omega = \{1,\dots,n\}$ maximiert $Q := (1/n,\dots,1/n)$ die Entropie:

Satz W4F: Die diskrete Gleichverteilung maximiert die Entropie.

Jedes WMaß P auf $\Omega = \{1,\dots,n\}$ erfüllt $0 \leq \mathrm{H}(P) \leq \log n$.
(1) Maximum: Genau dann gilt $\mathrm{H}(P) = \log n$, wenn P gleichverteilt ist.
(2) Minimum: Genau dann gilt $\mathrm{H}(P) = 0$, wenn $p_i = 1$ für ein $i \in \Omega$ gilt.

Beweis: (1) Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \leq \mathrm{H}(P,Q)$ gilt:

$$\mathrm{H}(P) \overset{\text{W4E}}{\leq} \mathrm{H}(P,Q) \overset{\text{Def}}{=} -\textstyle\sum_{i=1}^n p_i \log q_i \overset{Q}{=} \sum_{i=1}^n p_i \log n \overset{P}{=} \log n$$

Gleichheit gilt genau dann, wenn $p_i = q_i$ für alle $i \in \Omega$, also $P = Q$.

(2) Für alle $x \in [0,1]$ gilt $-x \log x \geq 0$, also $\mathrm{H}(P) = -\sum_i p_i \log p_i \geq 0$. Gleichheit $-x \log x = 0$ gilt nur für $x \in \{0,1\}$. Aus $\mathrm{H}(P) = 0$ folgt also $p_1,\dots,p_n \in \{0,1\}$. Mit $\sum_i p_i = 1$ folgt $p_i = 1$ für genau ein $i \in \Omega$. □

Prinzip der Indifferenz (Laplace 1812): Wenn keine Gründe vorliegen, irgendein Ergebnis zu begünstigen, so sind alle gleich wahrscheinlich. Dieses Modell maximiert die Symmetrie… und zugleich die Entropie!

Maximale Entropie: die diskrete Gleichverteilung

☺ Für $n = 2$ und $n = 3$ sehen wir das sehr schön in obigen Graphiken. Allgemein hilft uns hier die Informationsungleichung! Oder Lagrange:

Aufgabe: Maximieren Sie die Entropie $H(p_1,\dots,p_n) = \sum_{i=1}^n -p_i \ln p_i$ unter der Nebenbedingung $g(p) := p_1 + \ldots + p_n = 1$ und $p_1,\dots,p_n \geq 0$.

Lösung: Wir erinnern uns an den Satz über Lagrange–Multiplikatoren und lösen $\nabla H = -\lambda \nabla g$. Aus $-\ln p_i - 1 = -\lambda$ folgt erneut $p_i = q_i := 1/n$.

Ist dies das globale Maximum? Der Definitionsbereich Δ ist kompakt, hierauf ist $H : \Delta \to \mathbb{R}$ stetig, nimmt also Minimum und Maximum an. Im Inneren Δ° ist q die einzige kritische Stelle, mit Wert $H(q) = \ln(n)$. Auf dem Rand $\partial\Delta$ gilt $H(\partial\Delta) = [0, \ln(n-1)]$ per Induktion über n. Also gilt $H(\Delta) = [0, \ln(n)]$ mit q als einziger Maximalstelle. Raffiniert!

☺ Gibbs' Ungleichung / Konkavität erledigt alles auf einen Schlag!

? Ist dieses Ergebnis nicht offensichtlich? Nein, erst im Rückblick!

☺ Das Maximum zu bestimmen, übersichtlich und nachvollziehbar, ist nicht trivial, insbesondere in höherer Dimension. Es ist eine erste, effiziente und elegante Anwendung von Gibbs' Informationsungleichung.

Maximale Entropie: die Boltzmann–Verteilung

Ein Teilchen gerät in Zustand $\omega_i \in \Omega$ mit Energie $u : \Omega \to \mathbb{R} : \omega_i \mapsto u_i$. Zustände hoher Energie sind unwahrscheinlicher, aber wie genau?

Zur Wkt macht Boltzmann den Exponentialansatz $q_i = \exp(-\alpha - \beta u_i)$. Hier ist $\beta := 1/(\mathrm{k}T)$ mit Temperatur $T > 0$ und Boltzmann–Konstante $\mathrm{k} = 1.38 \times 10^{-23}$ J/K. Normierung: Wir bestimmen die Zustandssumme $Z := \sum_i \exp(-\beta u_i) > 0$ und erhalten $\alpha := \ln Z \in \mathbb{R}$; damit gilt $\sum_i q_i = 1$.

Sie hat Energie $\mathrm{E}(Q) = \sum_i q_i u_i$ und Entropie $\mathrm{S}(Q) = \mathrm{k} \sum_i q_i(\alpha + \beta u_i)$. Die Shannon–Entropie H skalieren wir zur Gibbs–Entropie $\mathrm{S} = \mathrm{k} \ln(2)\, \mathrm{H}$.

Satz W4G: Boltzmann maximiert die Entropie.

Jedes WMaß P auf Ω mit der Energieschranke $\mathrm{E}(P) \leq \mathrm{E}(Q)$ erfüllt auch die Entropieschranke $\mathrm{S}(P) \leq \mathrm{S}(Q)$, und Gleichheit gilt nur für $P = Q$.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{S}(P) \leq \mathrm{S}(P, Q)$ gilt:

$$\mathrm{S}(P) \overset{\text{W4E}}{\leq} \mathrm{k} \sum_i p_i(\alpha + \beta u_i) \overset{\text{Lin}}{=} \mathrm{k}[\alpha + \beta \mathrm{E}(P)] \overset{\text{Vor}}{\leq} \mathrm{k}[\alpha + \beta \mathrm{E}(Q)]$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „<“. □

Maximale Entropie: geometrische Verteilung

Beispiele: (1) Bei n Zuständen $\omega_1, \ldots, \omega_n$ gleicher Energie $u_i = \text{const}$ erhalten wir erneut die Gleichverteilung $q_i = 1/n$ wie oben in Satz W4F.

(2) Auf $\Omega = \mathbb{N}$ ist die geometrische Verteilung $q_i = (1-q)q^i$ mit $q \in [0, 1[$ die Boltzmann–Verteilung zu $u_i = i$ und $\beta = -\ln q$ und $\alpha = -\ln(1-q)$. Sie hat Erwartung / Energie $\mathrm{E}(Q) = \sum_{i=0}^{\infty} q_i \cdot i = q/(1-q)$ und Entropie

$$\begin{aligned} \mathrm{H}(Q) &= -\sum_{i=0}^{\infty} q_i \big[\ln(1-q) + i \ln(q)\big] \\ &= -\ln(1-q) - \ln(q) \cdot q/(1-q). \end{aligned}$$

Übung: Jedes WMaß P auf $\mathbb{N}$ mit Erwartung $\mathrm{E}(P) \leq \mathrm{E}(Q) = q/(1-q)$ erfüllt die Entropieschranke $\mathrm{H}(P) \leq \mathrm{H}(Q)$, mit Gleichheit nur für $P = Q$.

🙂 Unter allen WMaßen P auf $\Omega = \mathbb{N}$ mit Erwartung $\leq \mu < \infty$ maximiert die obige geometrische Verteilung Q mit $\mu = q/(1-q)$ die Entropie; dabei gilt $q = \mu/(1+\mu)$ und $\mathrm{H}(Q) = (1+\mu)\ln(1+\mu) - \mu \ln \mu$.

Methode der maximalen Entropie (Jaynes 1957): Erlauben die Daten mehrere Verteilungen, so wählen wir diejenige mit maximaler Entropie.

Jaynes' Methode der maximalen Entropie

Wir suchen WVerteilungen mit vorgegebenen Eigenschaften, wie z.B. Erwartung und Varianz, und wählen diejenige mit maximaler Entropie. Jede mit kleinerer Entropie enthält mehr Information / Struktur als unsere Messdaten stützen. Laplace' Prinzip der Indifferenz ist der Spezialfall.

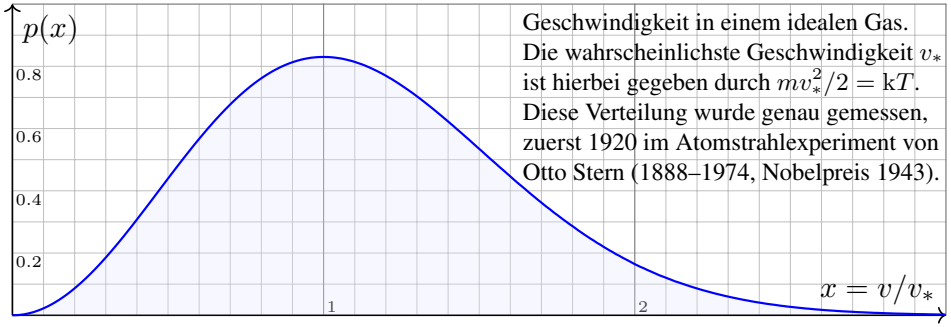
Aufgabe: Ein fairer Würfel $\Omega = \{1, \ldots, 6\}$ mit $P = (1/6, \ldots, 1/6)$ hat die erwartete Augenzahl $\mu = 3.5$. Sie messen $\mu_1 = 4.7$, andermal $\mu_2 = 2.8$. Welches Modellmaß Q auf Ω wählen Sie? Die Frage ist unterbestimmt!

Lösung: Gemäß der Methode der maximalen Entropie macht Jaynes den Ansatz $q_i = \exp(-\alpha - \beta i)$ mit $\alpha, \beta \in \mathbb{R}$. Numerisch bestimmen wir den Parameter $\beta \in \mathbb{R}$ mit $\mu(\beta) = (\sum_{i=1}^{6} \mathrm{e}^{-\beta i} \cdot i)/(\sum_{i=1}^{6} \mathrm{e}^{-\beta i})$ und finden $\beta_1 \approx -0.463$ und $Q_1 \approx (0.039, 0.062, 0.098, 0.157, 0.249, 0.395)$ ebenso $\beta_2 \approx +0.249$ und $Q_2 \approx (0.284, 0.221, 0.172, 0.134, 0.104, 0.081)$. *Whaaa?*

Für Jaynes war dies zuerst eine Frage der Logik, dann erst der Physik. Die statistische Mechanik liefert überzeugende weitere Argumente und begründet so den 2. Hauptsatz: Jedes physikalische System strebt zum thermischen Gleichgewicht, das heißt zum Zustand maximaler Entropie.

Maxwell–Verteilung der Geschwindigkeit im idealen Gas

Zustände hoher Energie werden seltener angenommen. Mit dem obigen Exponentialansatz konnten Boltzmann und Gibbs dies quantifizieren!



In einem idealen Gas ist die Teilchenenergie rein kinetisch, $E = mv^2/2$. Im thermischen Gleichgewicht ist die Wkt proportional zu $\exp(-E/\mathrm{k}T)$.

In 3 Dimensionen gilt $v^2 = v_x^2 + v_y^2 + v_y^2$. Die Zustände gleicher Energie füllen also die (Energie-)Kugelschale $[v, v + \mathrm{d}v]$ vom Volumen $4\pi v^2\, \mathrm{d}v$.

Für $x = v/v_*$ mit $mv_*^2/2 = \mathrm{k}T$ ist die WDichte $p(x) = \mathrm{e}^{-x^2} \cdot x^2 \cdot 4/\sqrt{\pi}$.

Kontinuierliche WVerteilungen auf $\Omega \subseteq \mathbb{R}^n$ und ihre Entropie

Auf $\Omega \subseteq \mathbb{R}^n$ sei P ein WMaß mit Dichte $p : \Omega \to \mathbb{R}_{\geq 0}$ und $\int_\Omega p(x)\,\mathrm{d}x = 1$. Ihre (natürliche) **Entropie** definieren wir (nun zur Basis e) durch

$$\mathrm{H}(P) := -\int_{x\in\Omega} p(x)\ln p(x)\,\mathrm{d}x \quad \in \bar{\mathbb{R}}$$

gemessen in Nat (*natural unit of information*), alternativ Bit zur Basis 2. Die **Kreuzentropie** kontinuierlicher WVerteilungen P und Q ist demnach

$$\mathrm{H}(P,Q) := -\int_{x\in\Omega} p(x)\ln q(x)\,\mathrm{d}x \quad \in \bar{\mathbb{R}}.$$

Die **relative Entropie** des Modells Q bezüglich der Wirklichkeit P ist

$$\mathrm{D}(P,Q) \;:=\; \mathrm{H}(P,Q) - \mathrm{H}(P) \;=\; \int_{x\in\Omega} p(x)\ln\frac{p(x)}{q(x)}\,\mathrm{d}x \quad \in \bar{\mathbb{R}}.$$

Aufgabe: Weiterhin gilt Gibbs' Informationsungleichung (Satz W4E), also $\mathrm{D}(P,Q) \geq 0$, somit $\mathrm{H}(P,Q) \geq \mathrm{H}(P)$. Gleichheit gilt genau dann, wenn $p(x) = q(x)$ für fast alle $x \in \Omega$ gilt, somit $P = Q$. Sind die Dichten $p, q : \Omega \to \mathbb{R}_{\geq 0}$ sogar stetig, so gilt $p(x) = q(x)$ in jedem Punkt $x \in \Omega$.

Maximale Entropie: kontinuierliche Gleichverteilung

Auf $\Omega \subseteq \mathbb{R}^n$ mit $0 < \mathrm{vol}_n(\Omega) < \infty$ hat die Gleichverteilung Q die Dichte

$$q \;:\; \Omega \to \mathbb{R}_{\geq 0} \;:\; q(x) = 1/\mathrm{vol}_n(\Omega).$$

Satz W4H: Die Gleichverteilung maximiert die Entropie.

(0) Für jedes kontinuierliche WMaß P auf Ω gilt die Schranke

$$\mathrm{H}(P) \leq \mathrm{H}(Q) = \ln \mathrm{vol}_n(\Omega).$$

Gleichheit gilt genau für $P = Q$, die Gleichverteilung auf Ω.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \leq \mathrm{H}(P,Q)$ gilt:

$$\mathrm{H}(P) \;\leq\; \int_{x\in\Omega} p(x)\ln \mathrm{vol}_n(\Omega)\,\mathrm{d}x \;\overset{P}{=}\; \ln \mathrm{vol}_n(\Omega)$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „$<$“. □

Beispiel: Die Gleichverteilung auf $[0,1]^n$ hat die Entropie $\ln 1 = 0$.
⚠ Die kontinuierliche Entropie kann negativ sein, die diskrete niemals.

Maximale Entropie: Exponentialverteilung $\mathrm{E}(\mu^{-1})$

Auf $\Omega = \mathbb{R}_{\geq 0}$ hat die Exponentialverteilung $Q := \mathrm{E}(\mu^{-1})$ die Dichte

$$q \;:\; \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \;:\; q(x) = \mathrm{e}^{-x/\mu}\cdot\mu^{-1}$$

mit Erwartung $\mu \in \mathbb{R}_{>0}$.Diese können wir wie folgt charakterisieren:

Satz W4H: Die Exponentialverteilung maximiert die Entropie.

(1) Für jedes kontinuierliche WMaß P auf $\mathbb{R}_{\geq 0}$ mit Erwartung $\leq \mu$ gilt

$$\mathrm{H}(P) \leq \mathrm{H}(Q) = 1 + \ln\mu.$$

Gleichheit gilt genau für $P = Q$, die Exponentialverteilung $\mathrm{E}(\mu^{-1})$.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \leq \mathrm{H}(P,Q)$ gilt:

$$\mathrm{H}(P) \;\leq\; \int_{x=0}^{\infty} p(x)\left[\frac{x}{\mu} + \ln\mu\right]\mathrm{d}x \;\overset{P}{\leq}\; 1 + \ln\mu$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „$<$“. □

Beispiel: Die Verteilung $\mathrm{E}(1)$ hat die Dichte $x \mapsto \mathrm{e}^{-x}$ und die Entropie 1.

Maximale Entropie: Normalverteilung $\mathrm{N}(\mu,\sigma^2)$

Auf $\Omega = \mathbb{R}$ hat die Normalverteilung $Q := \mathrm{N}(\mu,\sigma^2)$ die Dichte

$$q \;:\; \mathbb{R} \to \mathbb{R}_{\geq 0} \;:\; q(x) = \mathrm{e}^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\cdot\frac{1}{\sigma\sqrt{2\pi}}$$

mit Erwartung $\mu \in \mathbb{R}$ und Varianz $\sigma^2 \in \mathbb{R}_{>0}$. Entropie-Charakterisierung:

Satz W4H: Die Normalverteilung maximiert die Entropie.

(2) Für jedes kont. WMaß P auf $\mathbb{R}$ mit Erwartung μ und Varianz $\leq \sigma^2$ gilt

$$\mathrm{H}(P) \leq \mathrm{H}(Q) = \ln\sqrt{2\pi\,\mathrm{e}} + \ln\sigma$$

Gleichheit gilt genau für $P = Q$, die Normalverteilung $\mathrm{N}(\mu,\sigma^2)$.

Beweis: Dank Gibbs' Informationsungleichung $\mathrm{H}(P) \leq \mathrm{H}(P,Q)$ gilt:

$$\mathrm{H}(P) \;\leq\; \int_{\mathbb{R}} p(x)\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + \ln(\sigma\sqrt{2\pi})\right]\mathrm{d}x \;\overset{P}{\leq}\; \frac{1}{2} + \ln(\sigma\sqrt{2\pi})$$

Für $P = Q$ gilt Gleichheit, sonst die strikte Ungleichung „$<$“. □

Beispiel: Die Normalverteilung $\mathrm{N}(0,1)$ hat Entropie $\ln\sqrt{2\pi\,\mathrm{e}} \approx 1.419$.

Entropie im zentralen Grenzwertsatz (ZGS)

Die obigen Rechnungen zeigen eindrucksvolle Beispiele und lassen die große Bedeutung der Entropie erahnen. Als Ausblick skizziere ich drei Anwendungen: Grenzwertsatz, maschinelles Lernen, Thermodynamik.

Gegeben seien unabhängige ZVariablen $X_1, X_2, \ldots : \Omega \to \mathbb{R}$.
Die Summe $S_n = X_1 + \ldots + X_n$ ist annähernd normalverteilt,
also $S_n \approx \mathrm{N}(\mu_n, \sigma_n^2)$ mit $\mu_n = \sum_{i=1}^n \mathbf{E}(X_i)$ und $\sigma_n^2 = \sum_{i=1}^n \mathbf{V}(X_i)$.
Zentriert und normiert erhalten wir $S_n^* := (S_n - \mu_n)/\sigma_n \approx \mathrm{N}(0,1)$.
Die Entropie strebt für $n \to \infty$ zum Maximum, $\mathrm{H}(S_n^*) \nearrow \mathrm{H}(\mathrm{N}(0,1))$.
Die Entropie treibt so die Konvergenz der Verteilungen, $S_n^* \to \mathrm{N}(0,1)$.

Physikalische Anschauung: Die vielen unabhängigen Summanden in S_n^* sorgen für hohe Entropie. Mit wachsendem n nähert sich die Entropie dem Maximum… der Normalverteilung, wie in Satz W4H gesehen.

Oft wird der ZGS durch Fourier–Transformation bewiesen. Das ist genial, aber auch etwas unintuitiv. Das Entropie-Argument hingegen entspricht direkt unserer physikalischen Anschauung aus dem 2. Hauptsatz der Thermodynamik, ist in diesem Sinne also intuitiv näherliegend.

Entropie im zentralen Grenzwertsatz (ZGS)

🙂 Mit diesem letzten Höhepunkt endet unsere kurze Einführung in die wunderbare Welt der Wahrscheinlichkeit und insgesamt diese Vorlesung. Unser Weg war lang, auch anstrengend, doch er hat sich wieder gelohnt!

Unsere Beispiele illustrieren eindrücklich: Wahrscheinlichkeit begegnet Ihnen überall, ob Sie wollen oder nicht, in alltäglichen Entscheidungen ebenso wie in Mathematik, Physik und technischen Anwendungen bis hin zu gesellschaftlichen und ökonomischen Belangen.

Unser Ziel sind rationale Entscheidungen unter Unsicherheit.
Die Erfahrung zeigt: Auch mit dem Zufall können wir rechnen!
Wir wollen nachvollziehbar begründete, quantitative Aussagen.
Die Mathematik stellt Ihnen dafür alle Werkzeuge bereit.

Das gilt allgemein für jede physikalisch-technische Betrachtung und insbesondere für raffinierte, fortgeschrittene Konzepte wie die Entropie. Damit können Sie nun der vagen Idee vom „Maß für die Unordnung“ einen präzisen mathematischen Sinn geben.

Divergenz und Entropie im maschinellen Lernen (ML)

Shannon verband 1948 die fundamentalen Begriffe „Information“ und „Entropie“. Jüngst feiert dies im maschinellen Lernen große Erfolge:

Wie kann ein Computer aus Daten lernen? Die Grundidee ist einfach: Die Wirklichkeit entspricht einem WMaß P (*ground truth*). Das Training anhand von Beispieldaten erzeugt ein approximatives WMaß Q (*model*). Wie gut ist es? Die Divergenz $\mathrm{D}(P,Q) = \mathrm{H}(P,Q) - \mathrm{H}(P) \geq 0$ misst die Abweichung des Modells Q von der Wirklichkeit P. Diese Zielgröße wollen wir minimieren: Die Entropie misst den Lernfortschritt!

Problem: Leider kennen wir die Entropie $\mathrm{H}(P)$ der Wirklichkeit P nicht. Geniale Lösung: Es genügt, die Kreuzentropie $\mathrm{H}(P,Q)$ zu minimieren. Sie kann durch Stichprobe geschätzt werden (Monte Carlo Methode). So kann die Güte verbessert und durch Training optimiert werden.

Aus diesem Grund ist die Kreuzentropie $\mathrm{H}(P,Q)$ die zentrale Zielgröße im maschinellen Lernen (ML). Die Erfolge der generativen künstlichen Intelligenz (genAI) beruhen maßgeblich auf dieser fundamentalen Idee und ihrer effizienten Implementierung für gigantisch große Modelle.

Zum guten Schluss

So jung und mit so geringer Entropie
kommen wir nie wieder zusammen.
Tocotronic, youtu.be/CpqeCZB3sMQ

In Ihrem Leben, Studium und Karriere, werden diese zentralen Themen eine Rolle spielen. Bewährte Techniken des Ingenieurwesens werden erweitert durch Künstliche Intelligenz, demnächst Quantum Computing, zukünftig noch weitere Ideen, die derzeit noch niemand ahnt.

Ihre mathematische Ausbildung wird Ihnen überall gute Dienste leisten.

Ich wünsche Ihnen, die hier dargelegten Werkzeuge bieten Ihnen eine gute Grundlage im Studium und einen Einstieg zur weiteren Vertiefung. Ich hoffe, damit bleiben die vielen wichtigen Anwendungen nicht „most misunderstood“, sondern führen Sie zu einem besseren Verständnis.

Habe Mut, dich deines eigenen Verstandes zu bedienen!

Much to learn, you still have.
This is just the beginning.