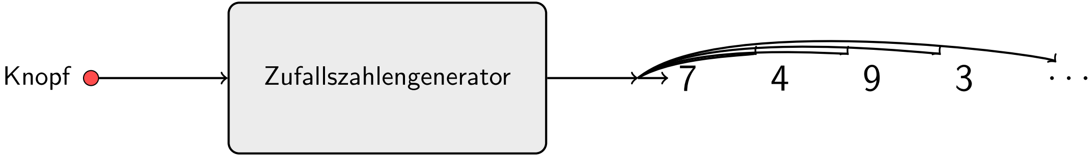
Konvergenzbegriffe in der Wahrscheinlichkeitstheorie

Warum schwache Konvergenz nicht genügt – pfadweises Zusammenwachsen und praktische Beispiele

Oliver Dürr

May 25, 2025

Der naive Zufallsbegriff – "Black Box"



- Jeder Knopfdruck liefert **eine Zahl**.
- Viele Knopfdrücke $\Longrightarrow$ eine **empirische Verteilung**.
- Die *Zufallsquelle* (das ω aus $(\Omega, \mathcal{F}, P)$) bleibt unsichtbar.

Konvergenz in Wahrscheinlichkeit

Definition

$$X_n \xrightarrow{\text{P}} X \quad \Longleftrightarrow \quad \forall \varepsilon > 0: \; P\big(|X_n - X| > \varepsilon\big) \longrightarrow 0.$$

Beispiel: empirische Verteilungsfunktion Für i.i.d. $X_1, \ldots, X_n$ mit Verteilungsfunktion F:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(X_i \leq x)} \quad \Longrightarrow \quad F_n(x) \xrightarrow{\text{P}} F(x) \quad \text{(f. j. fixes } x\text{)}.$$

Intuition

Der Schätzer $F_n(x)$ wird mit wachsendem n *wahrscheinlich* beliebig nah an die wahre Verteilungsfunktion $F(x)$ heranrücken.

Zentraler Grenzwertsatz (i.i.d.-Version)

Seien $X_1, X_2, \ldots$ i.i.d. mit $E[X_i] = \mu$ und $\mathrm{Var}(X_i) = \sigma^2 < \infty$.

$$\frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- *Konvergenz in Verteilung* – Zufallsstreuung bleibt erhalten.

Gesetz der großen Zahlen – einfache Version

Für dieselben i.i.d. X_i:

$$\bar{X}_n \xrightarrow{\text{P}} \mu.$$

Lesart

Mit wachsendem n liegt der Stichprobenmittelwert mit hoher Wahrscheinlichkeit beliebig nah am Erwartungswert μ.

Gesetz der großen Zahlen – formale Version

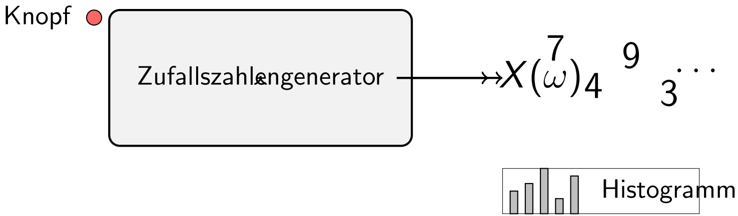
$$P\big(|\bar{X}_n - \mu| > \varepsilon\big) \ \longrightarrow_{n\to\infty}\ 0, \quad \forall \varepsilon > 0.$$

Kurzform vs. Ereignis-Sprache

$$P\underbrace{\big(\{\omega : |\bar{X}_n(\omega) - \mu| > \varepsilon\}\big)}_{\text{"zu weit weg"}} \ \to 0.$$

- Die Differenz zu einer **Zufallsvariable** $Z \equiv \mu$ wird klein.
- Schreibweise ohne ω ist nur Abkürzung – das Ereignis lebt in Ω.

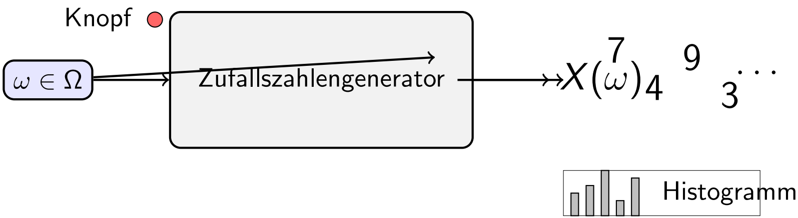
Intuition: Zufall $\rightarrow$ Verteilung (Black-Box-Bild)



- Naive Sicht: *Ein Knopfdruck* $\rightarrow$ *eine Zahl*. Viele Knopfdrücke $\Rightarrow$ Häufigkeiten $\approx$ Verteilung.
- Unsichtbar bleibt der **Master-Zufall** $\omega \in \Omega$. Eine Zufallsvariable ist nur die deterministische Abbildung $X(\omega)$.

Zentraler Punkt: Ob zwei Variablen *gemeinsam* konvergieren, hängt davon ab, ob sie denselben ω teilen.

Intuition: Zufall $\to$ Verteilung (Black-Box-Bild)



- Naive Sicht: *Ein Knopfdruck* $\to$ *eine Zahl*. Viele Knopfdrücke $\Rightarrow$ Häufigkeiten $\approx$ Verteilung.
- Unsichtbar bleibt der **Master-Zufall** $\omega \in \Omega$. Eine Zufallsvariable ist nur die deterministische Abbildung $X(\omega)$.

Zentraler Punkt: Ob zwei Variablen *gemeinsam* konvergieren, hängt davon ab, ob sie denselben ω teilen.

Zwei Konvergenzbegriffe – Definitionen

Konvergenz in Verteilung (schwach)

$X_n \xrightarrow{d} X \iff F_{X_n}(x) \to F_X(x)$ für alle Stetigkeitspunkte x.

Konvergenz in Wahrscheinlichkeit

$X_n \xrightarrow{\mathrm{P}} X \iff \forall\varepsilon > 0: \; P(|X_n - X| > \varepsilon) \to 0.$

Hierarchie a.s. $\Rightarrow$ P $\Rightarrow$ d

Merke

P vergleicht Pfade auf demselben ω; d vergleicht bloß Randverteilungen.

Beispiel 1 – Empirische Verteilungsfunktion

$X_1, \dots, X_n$ i.i.d., Verteilungsfunktion F. $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(X_i \leq x)}$.

$F_n(x) \xrightarrow{\mathrm{P}} F(x)$ für jedes feste x.

Stärkere Glivenko–Cantelli-Aussage: $\sup_x |F_n(x) - F(x)| \xrightarrow{a.s.} 0$.

Beispiel 2 – Zentraler Grenzwertsatz

(X_i) i.i.d., $E[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$.

$$\frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

Schwach $\Rightarrow$ Zufallsstruktur (Varianz 1) bleibt.

LLN – einfache Version

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathrm{P}} \mu.$$

Pfaddeutung: Für fast jedes ω liegt $\bar{X}_n$ irgendwann ε-nah an μ.

LLN – formale Version

$$P(|\bar{X}_n - \mu| > \varepsilon) \to 0 \quad (\forall \varepsilon > 0).$$

Kurzform für $P\{\omega : |\bar{X}_n(\omega) - \mu| > \varepsilon\} \to 0$.

Kopplung vieler MCMC-Ketten

Quadratisches Potential $V(k) = k^2$, gemeinsamer RNG-Strom. Startwerte verschieden $\rightarrow$ Meeting-Zeit τ_{max}.

$$\forall i,j: \; |X_t^{(i)} - X_t^{(j)}| \xrightarrow{\mathrm{P}} 0 \quad \Longrightarrow \quad X_t^{(i)} \xrightarrow{\mathrm{P}} Z, \; Z \sim \pi(k) \propto e^{-k^2}.$$

Zwei RNG-Streams – Seed-Effekt

Unterschiedliche Seeds
X_n, Y_n Uniform(0,1).

$$P(|X_n - Y_n| > \varepsilon) \not\to 0 \quad \Rightarrow \quad X_n \stackrel{P}{\not\to} Y_n.$$

Gemeinsamer Seed
$X_n = Y_n \; \forall n \;\Rightarrow\;$ sofort $X_n \xrightarrow{P} Y_n$.
Gleiche Randverteilungen $\neq$ Pfadnähe.

Take-aways

- Schwache Konvergenz: nur Randverteilungen.
- Konvergenz in Wahrscheinlichkeit: Pfade auf gleichem ω.
- Kopplungstricks zeigen den Unterschied praktisch (MCMC, Seeds).
- Burn-in: notwendig, aber unabhängig von Pfadverschmelzung.