

Arten von stochastischer Konvergenz

In diesen Notizen erklären wir verschiedenen Arten von Konvergenz von Folgen von Zufallsvariablen, und deren Beziehungen.

Definition 1 Seien $(Y_i)_{i \in \mathbb{N}}$ und Y Zufallsvariablen (oder Vektoren) aus dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) $Y_n \rightarrow Y$ fast sicher, falls

$$Y_n(\omega) \xrightarrow{n \rightarrow \infty} Y(\omega) \quad \text{bis auf eine Menge von } \mathbb{P}\text{-Maß null.}$$

(b) $Y_n \rightarrow Y$ im p -ten Mittel (oder in L^p) für $p \geq 1$, falls

$$\mathbb{E}(|Y_n - Y|^p) \xrightarrow{n \rightarrow \infty} 0.$$

(c) $Y_n \rightarrow Y$ in Wahrscheinlichkeit, falls für jedes $\varepsilon > 0$

$$\mathbb{P}(\omega \in \Omega : |Y_n(\omega) - Y(\omega)| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

(d) $Y_n \rightarrow Y$ in Verteilung, falls für jedes $t \in C$

$$|F_{Y_n}(t) - F_Y(t)| \xrightarrow{n \rightarrow \infty} 0,$$

wobei F_{Y_n} und F_Y die Verteilungsfunktionen von Y_n und Y sind und $C = \{t \in \mathbb{R} : F_Y \text{ ist stetig in } t\}$.

Bemerkung 2 Verteilungsfunktion sind immer rechts-stetig, aber in der Menge C in Teil (d) wird zwei-seitige Stetigkeit gefordert.

Häufige Notationen sind: $Y_n \rightarrow Y$ f.s. für fast sichere Konvergenz, $Y_n \rightarrow_W Y$ für Konvergenz in Wahrscheinlichkeit und $Y_n \rightarrow_d Y$ oder $Y_n \Rightarrow Y$ für Konvergenz in Verteilung.

Theorem 3 Die Beziehungen zwischen diesen Arten von Konvergenz sind die folgende:

$$(a) \Rightarrow (c) \Rightarrow (d) \quad \text{und} \quad (b) \Rightarrow (c) \Rightarrow (d),$$

aber zwischen (a) und (b) gibt es keine allgemeine Implikation.

Beispiel 4 Seien (X_i) identisch verteilte, unabhängige Zufallsvariablen mit gemeinsamer Erwartung $\mu \in \mathbb{R}$ und Varianz σ^2 . Seien $S_n = X_1 + \dots + X_n$ und $Y_n = \frac{1}{n}S_n$. Dann ist es klar, dass $\mathbb{E}(Y_n) = \mu$ und jetzt fassen wir μ auch auf als diskrete Zufallsvariable Y mit $\mathbb{P}(Y = \mu) = 1$.

(a) $Y_n \rightarrow Y$ fast sicher. Dies ist das starke Gesetz der großen Zahlen.

(b) $Y_n \rightarrow Y$ im p -ten Mittel mit $p \in [1, 2]$. Um das zu zeigen, fangen wir mit $p = 2$ an. Wir definieren $\tilde{Y}_n = \frac{1}{n} \sum_{i=1}^n \underbrace{(X_i - \mu)}_{\tilde{X}_i}$. Dann $|\tilde{Y}_n| = |Y_n - \mu|$ und es reicht zu zeigen, dass

$\mathbb{E}(|\tilde{Y}_n|^2) = \text{Var}(\tilde{Y}_n) \rightarrow 0$. Aber weil die $\tilde{X}_i := X_i - \mu$ unabhängig sind, gilt

$$\text{Var}(\tilde{Y}_n) = \frac{1}{n^2} \text{Var}(\tilde{X}_1 + \dots + \tilde{X}_n) = \frac{1}{n^2} n \text{Var}(\tilde{X}_1) = \frac{1}{n} \sigma^2 \rightarrow 0.$$

Jetzt für allgemeines $p \in [1, 2)$ verwenden wir die Hölder Ungleichung mit $a = 2/p$ und $b = 2/(2-p)$, also $1/a + 1/b = 1$:

$$\mathbb{E}(|\tilde{Y}_n|^p) = \mathbb{E}(|\tilde{Y}_n|^p \cdot 1) \leq \mathbb{E}(|\tilde{Y}_n|^2)^{p/2} \mathbb{E}(1^{2/(2-p)})^{(2-p)/2} = \mathbb{E}(|\tilde{Y}_n|^2)^{p/2} \rightarrow 0.$$

Für die Konvergenz $Y_n \rightarrow Y$ im p -ten Mittel mit $p > 2$ bräuchten wir eine stärkere Annahme, nämlich dass $\mathbb{E}(|X_i|^p)$ beschränkt ist in i .

(c) $Y_n \rightarrow Y$ in Wahrscheinlichkeit. Dies ist das schwache Gesetz der großen Zahlen.

(d) $Y_n \rightarrow Y$ in Verteilung. Dass dies gilt, folgt aus (a) oder (b) eingesetzt in Theorem 3, aber um die Rolle der Menge C näher zu erklären, nehmen wir an, dass jedes X_i eine symmetrische Dichte rund um μ hat. Dann hat auch Y_n eine symmetrische Dichte um μ und

$$F_{Y_n}(\mu) = \mathbb{P}(Y_n = \mu) + \frac{1}{2}(1 - \mathbb{P}(Y_n = \mu)),$$

oder noch einfacher, wenn Y_n kontinuierlich ist, also $\mathbb{P}(Y_n = \mu) = 0$, dann $F_{Y_n}(\mu) = \frac{1}{2}$. Aber das konvergiert **nicht** gegen $F_Y(\mu) = 1$! Aber die Verteilungsfunktion $F_Y = 1_{[\mu, \infty)}$ macht genau einen Sprung bei μ , also $\mu \notin C = \mathbb{R} \setminus \{\mu\}$.

Für $t < \mu$ folgt schon aus dem schwachen Gesetz der großen Zahlen, dass $F_{Y_n}(t) \rightarrow 0$. Für $t > \mu$ gilt analog, dass $F_{Y_n}(t) \rightarrow 1$. Also für $t \in C$ haben wir tatsächlich die Konvergenz in Verteilung.

Beweis von Theorem 3. (a) \Rightarrow (c): Sei $\varepsilon > 0$ gegeben, und $\Omega_n = \{\omega \in \Omega : \exists m \geq n \text{ so dass } |Y_m(\omega) - Y(\omega)| > \varepsilon\}$. Dann gilt $\Omega_n \supset \Omega_{n+1}$ und $\lim_{n \rightarrow \infty} \mathbb{P}(\Omega_n) = \mathbb{P}(\bigcap_n \Omega_n) = 0$ wegen der fast sicheren Konvergenz. Aber $\{\omega \in \Omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\} \subset \Omega_n$, also

$$\mathbb{P}(\{\omega \in \Omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) \leq \mathbb{P}(\Omega_n) \rightarrow 0,$$

und das ist Konvergenz in Wahrscheinlichkeit.

(b) \Rightarrow (c): Sei $\varepsilon > 0$. Wegen der Markov-Ungleichung,

$$\begin{aligned} \mathbb{P}(\{\omega \in \Omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) &= \mathbb{P}(\{\omega \in \Omega : |Y_n(\omega) - Y(\omega)|^p > \varepsilon^p\}) \\ &\leq \frac{1}{\varepsilon^p} \mathbb{E}(|Y_n(\omega) - Y(\omega)|^p) \rightarrow 0. \end{aligned}$$

(c) \Rightarrow (d): Sei $\varepsilon > 0$ gegeben. Die Verteilungsfunktionen erfüllen für jedes $t \in \mathbb{R}$:

$$\begin{aligned} F_{Y_n}(t) &= \mathbb{P}(Y_n \leq t) \\ &= \mathbb{P}(Y_n \leq t \wedge Y \leq t + \varepsilon) + \mathbb{P}(Y_n \leq t \wedge Y > t + \varepsilon) \\ &\leq \mathbb{P}(Y \leq t + \varepsilon) + \mathbb{P}(|Y_n - Y| > \varepsilon) \\ &= F_Y(t + \varepsilon) + \mathbb{P}(|Y_n - Y| > \varepsilon), \end{aligned}$$

sowie

$$\begin{aligned}
F_Y(t - \varepsilon) &= \mathbb{P}(Y \leq t - \varepsilon) \\
&= \mathbb{P}(Y \leq t - \varepsilon \wedge Y_n \leq t) + \mathbb{P}(Y \leq t - \varepsilon \wedge Y_n > t) \\
&\leq \mathbb{P}(Y_n \leq t) + \mathbb{P}(|Y_n - Y| > \varepsilon) \\
&= F_{Y_n}(t) + \mathbb{P}(|Y_n - Y| > \varepsilon).
\end{aligned}$$

Diese Formeln kombinieren sich zu

$$F_Y(t - \varepsilon) - \mathbb{P}(|Y_n - Y| > \varepsilon) \leq F_{Y_n}(t) \leq F_Y(t + \varepsilon) + \mathbb{P}(|Y_n - Y| > \varepsilon).$$

Wegen der Konvergenz in Wahrscheinlichkeit haben wir $\mathbb{P}(|Y_n - Y| > \varepsilon) \rightarrow 0$ für $n \rightarrow \infty$. Also

$$F_Y(t - \varepsilon) \leq \lim_{n \rightarrow \infty} F_{Y_n}(t) \leq F_Y(t + \varepsilon).$$

Jetzt $\varepsilon \rightarrow 0$ und weil F_Y **zwei-seitig** ist in t , bekommen wir $\lim_{n \rightarrow \infty} F_{Y_n}(t) = F_Y(t)$.

(a) $\not\Rightarrow$ (b): Sei $\Omega = [0, 1]$ mit Lebesgue-Maß \mathbb{P} . Sei $Y_n = 1_{[0, 1/n]}$ und $Y \equiv 0$. Dann $Y_n(\omega) \rightarrow Y(\omega)$ für jedes $\omega \in (0, 1]$, also fast sicher. Andererseits

$$\mathbb{E}(|Y_n - Y|^p) = \int_0^{1/n} |n^{1/p} - 0|^p dx = \int_0^{1/n} n dx = 1 \not\rightarrow 0.$$

(b) $\not\Rightarrow$ (a): Sei $\Omega = \mathbb{S}^1 = [0, 1]/_{0 \sim 1}$ der Einheitskreis mit Lebesgue-Maß \mathbb{P} und $a_n = \sum_{k=1}^n \frac{1}{k}$. Dann ist $A_n := [a_n \bmod 1, a_{n+1} \bmod 1]$ ein Teilintervall von Ω von Länge $\frac{1}{n+1}$, und A_n liegt immer neben A_{n+1} .

Jetzt nehmen wir $Y_n = 1_{A_n}$ und $Y \equiv 0$. Dann

$$\mathbb{E}(|Y_n - Y|^p) = \int_{A_n} 1 dx = |A_n| = \frac{1}{n+1} \rightarrow 0.$$

Andererseits $a_n \rightarrow \infty$ und deswegen überdecken die Intervalle zusammen jeden Punkt $\omega \in \Omega$ unendlich oft. Für jedes $\omega \in \Omega$ gibt unendlich viele n , so dass $Y_n(\omega) = 1$ und auch unendlich viele n , so dass $Y_n(\omega) = 0$. In jedem Fall $Y_n(\omega) \not\rightarrow Y(\omega)$. \square

Das wichtigste Theorem über Konvergenz in Verteilung ist der Zentrale Grenzwertsatz.

Theorem 5 Sei $(X_i)_{i=1}^{\infty}$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit Varianz $\text{Var}(X_i) = \sigma^2 \in (0, \infty)$ und Erwartung $\mathbb{E}(X_i) = \mu$. Dann gibt es ein $Y \simeq \mathcal{N}(0, 1)$, so dass

$$Y_n := \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} \rightarrow Y \quad \text{in Verteilung.}$$

Konvergenz in Verteilung ausgeschrieben bedeutet hier, dass für alle $t \in \mathbb{R}$ gilt

$$\mathbb{P}\left(\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} \leq t\right) \xrightarrow{n \rightarrow \infty} \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Bemerkung 6 • *Identisch verteilt ist nicht essentiell, solange $\sup_i \text{Var}(X_i) < \infty$ (und einige technische Sachen gelten). Das heißt $\mathbb{E}(X_i) = \mu_i$ und dann existiert ein $\sigma \in (0, \infty)$ und $Y \simeq \mathcal{N}(0, 1)$, so dass*

$$Y_n := \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sigma \sqrt{n}} \rightarrow Y \quad \text{in Verteilung.}$$

- *Die Annahme von Unabhängigkeit kann auch etwas abgeschwächt werden, aber das behandeln wir nicht in diesem Kurs.*
- *Auch wenn $\text{Var}(X_i) = \infty$ (aber $\mathbb{E}(|X_i|^p) < \infty$ für irgendein $p \in [1, 2)$), kann noch eine Art von Konvergenz in Verteilung gelten, mit \sqrt{n} ersetzt durch $n^{1/p}$ oder $(n \log n)^{1/p}$. Das sind die sogenannten stabilen Gesetze, die wir aber nicht in diesem Kurs behandeln.*
- *Stärkere Konvergenz-Arten für $Y_n \rightarrow Y \simeq \mathcal{N}(0, 1)$ gibt es auch (zB das fast sichere Invarianzprinzip), aber auch das behandeln wir nicht in diesem Kurs.*