Sommersemester 2017

Wahrscheinlichkeitstheorie

Übungsblatt 11: Lösungen

Prof. K. Panagiotou/K. Matzke

Zentralübungsaufgaben

Aufgabe Z11.1

Gegeben sei ein unabhängiges $\triangle$-Schema $\left((X_{n,l})_{l=1}^{k_n}\right)_{n\in\mathbb{N}}$ mit $\mathbb{E}[X_{n,l}] =: \mu_{n,l}$ und $\mathbb{V}\mathrm{ar}[X_{n,l}] =: \sigma_{n,l}^2$ sowie $s_n^2 := \sum_{l=1}^{k_n} \sigma_{n,l}^2$. Die Lyapunov-Bedingung für dieses $\triangle$-Schema besagt

$$\exists \delta > 0 : \quad \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{l=1}^{k_n} \mathbb{E}\left[|X_{n,l} - \mu_{n,l}|^{2+\delta}\right] = 0.$$

Zeigen Sie, dass die Lyapunov-Bedingung die Lindeberg-Bedingung impliziert.

Lösung. Wir nehmen o.B.d.A. an, das $\triangle$-Schema sei zentriert, also $\mu_{n,l} = 0$ für alle n, l. Sei $\delta > 0$ aus der Lyapunov-Bedingung. Für beliebiges $\varepsilon > 0$ gilt dann

$$\begin{aligned}
&\frac{1}{s_n^2} \sum_{l=1}^{k_n} \int_{\{X_{n,l}^2 \geq \varepsilon s_n^2\}} X_{n,l}^2 \,\mathrm{d}P \\
&\quad \leq \frac{1}{s_n^2} \sum_{l=1}^{k_n} \int_{\{|X_{n,l}| \geq \sqrt{\varepsilon} s_n\}} |X_{n,l}|^{2+\delta} \cdot \frac{1}{(\sqrt{\varepsilon} s_n)^\delta} \,\mathrm{d}P \\
&\quad \leq \frac{\varepsilon^{-\delta/2}}{s_n^{2+\delta}} \sum_{l=1}^{k_n} \mathbb{E}\left[|X_{n,l}|^{2+\delta}\right] \xrightarrow{n\to\infty} 0.
\end{aligned}$$

Aufgabe Z11.2

Seien $(X_n)_{n\in\mathbb{N}}$ unabhängig identisch verteilt mit $\mathbb{E}[X_1] = 0, \mathbb{E}[X_1^2] = 1$. Beweisen oder widerlegen Sie, dass

$$\frac{S_n}{\sqrt{n}} \xrightarrow{P} \mathcal{N}(0,1).$$

Lösung. Nehme an, es existiere eine solche Folge X_n mit $S_n/\sqrt{n} \xrightarrow{P} X \sim \mathcal{N}(0,1)$. Wir definieren $T_n = \frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}}$ und zeigen, dass nun einerseits $T_n \xrightarrow{P} 0$, aber andererseits $T_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, 2-\sqrt{2})$. Dies gibt uns einen direkten Widerspruch, denn $\mathcal{N}_{0,2-\sqrt{2}}([\varepsilon,\infty)) > 0$ für beliebiges $\varepsilon > 0$.
Die erste Behauptung folgt unmittelbar via

$$P(|T_n \geq \varepsilon|) \leq P\left(\left|\frac{S_{2n}}{\sqrt{2n}} - X\right| \geq \varepsilon\right) + P\left(\left|\frac{S_n}{\sqrt{n}} - X\right| \geq \varepsilon\right) \xrightarrow{n\to\infty} 0.$$

Für die zweite Behauptung bemerken wir, dass

$$T_n = \frac{S_{2n} - \sqrt{2}S_n}{\sqrt{2n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \underbrace{\frac{X_{n+k} + (1-\sqrt{2})X_k}{\sqrt{2}}}_{=:Y_{n,k}}.$$

Das Schema $(Y_{n,k})_{k=1}^n$ ist offenbar unabhängig und zentriert. Für beliebiges $k \le n$ gilt zudem

$$\begin{aligned}\mathbb{E}[Y_{n,k}^2] &= \tfrac{1}{2}\left(\mathbb{E}\left[X_{n+k}^2\right] + 2(1-\sqrt{2})\mathbb{E}\left[X_{n+k}X_k\right] + (1-\sqrt{2})^2\mathbb{E}\left[X_k^2\right]\right) \\ &= \tfrac{1}{2}(1 + 0 + (1 - 2\sqrt{2} + 2)) = 2 - \sqrt{2}.\end{aligned}$$

Folglich ist $\left(\frac{Y_{n,k}}{\sqrt{n}\sqrt{2-\sqrt{2}}}\right)_{k=1}^n$ ein unabhängiges, zentriertes, normiertes $\triangle$-Schema und erfüllt wegen

$$\sum_{k=1}^n \mathbb{E}\left[\frac{Y_{n,k}^2}{n(2-\sqrt{2})} \cdot \mathbb{1}_{\left\{Y_{n,k}^2 > \varepsilon n(2-\sqrt{2})\right\}}\right] = (2-\sqrt{2})^{-1}\mathbb{E}\left[Y_{1,1}^2 \cdot \mathbb{1}_{\left\{Y_{1,1}^2 > \varepsilon n(2-\sqrt{2})\right\}}\right] \xrightarrow{n\to\infty} 0$$

die Lindeberg-Bedingung. Also gilt $\frac{T_n}{\sqrt{2-\sqrt{2}}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ bzw. $T_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, 2-\sqrt{2})$, was die Behauptung war.

Aufgabe Z11.3

Seien $(X_n)_{n\in\mathbb{N}}$ unabhängig identisch verteilt mit Dichte $f_{X_1}(x) = |x|^{-3} \cdot \mathbb{1}_{\mathbb{R}\setminus[-1,1]}(x)$. Bestimmen Sie eine Folge $(a_n)_{n\in\mathbb{N}}$, sodass

$$\frac{S_n}{a_n} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Hinweis: X_1 hat keine Varianz. Gehen Sie vor wie im Beweis des starken Gesetzes der großen Zahlen: Schneiden Sie geeignet ab!

Lösung. Wir machen den Ansatz, $Y_k = X_k \mathbb{1}_{\{|X_k| \le c_k\}}$ zu definieren. Das Borel-Cantelli-Lemma liefert eine Möglichkeit zu kontrollieren, dass Y_k sich nur für endlich viele k von X_k unterscheidet. Dafür muss

$$\sum_{k\ge 1} P(|X_k| > c_k) = 2\sum_{k\ge 1}\int_{c_k}^{\infty} x^{-3}\,\mathrm{d}x = \sum_{k\ge 1} c_k^{-2}$$

endlich sein. Wir wählen daher $Y_k = X_k \cdot \mathbb{1}_{\{|X_k| \le \sqrt{k}\log k\}}$. Damit ist $Y_1 = Y_2 = 0$ und für $k \ge 3$ gilt

$$\mathbb{V}\mathrm{ar}[Y_k] = \int x^2 f_{X_1}(x)\,\mathrm{d}x = 2\int_1^{\sqrt{k}\log k} x^{-1}\,\mathrm{d}x = \log k + 2\log\log k.$$

Wir definieren $a_n^2 := \sum_{k=1}^n \mathbb{V}\mathrm{ar}[Y_k]$ und $Y_{n,k} = Y_k/a_n$. Damit ist $(Y_{n,k})$ nach Definition ein unabhängiges, zentriertes und normiertes $\triangle$-Schema. Mit der Beobachtung $a_n^2 \sim n\log n$

(Vergleich mit Integral oder Stirling-Formel) folgt zudem

$$\begin{aligned}
\sum_{k=1}^{n} \mathbb{E}\left[Y_{n,k}^2 \cdot \mathbb{1}_{\{Y_{n,k}^2 > \varepsilon\}}\right] &= \frac{1}{a_n^2} \sum_{k=1}^{n} \mathbb{E}\left[X_k^2 \cdot \mathbb{1}_{\{\sqrt{\varepsilon} a_n < |X_k| \leq \sqrt{k} \log k\}}\right] \\
&\leq \frac{n}{a_n^2} \mathbb{E}\left[X_1^2 \cdot \mathbb{1}_{\{\sqrt{\varepsilon} a_n < |X_1| \leq \sqrt{n} \log n\}}\right] \\
&= \frac{2n}{a_n^2} \int_{\sqrt{\varepsilon} a_n}^{\sqrt{n} \log n} x^2 f_{X_1}(x) \, \mathrm{d}x = \frac{2n}{a_n^2} \log\left(\frac{\sqrt{n} \log n}{\sqrt{\varepsilon} a_n}\right) \\
&\sim \frac{2n}{n \log n} \log\left(\frac{\sqrt{n} \log n}{\sqrt{\varepsilon} \sqrt{n \log n}}\right) \\
&\leq \frac{2 \log \log n - 2 \log \sqrt{\varepsilon}}{\log n} \xrightarrow{n \to \infty} 0,
\end{aligned}$$

die Lindeberg-Bedingung ist erfüllt und $\frac{1}{a_n} \sum_{k=1}^{n} Y_k \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$.
Wir fassen zusammen: Nach Borel-Cantelli ist $Y_k \neq X_k$ nur endlich oft, also $\frac{1}{a_n} \sum_{k=1}^{n} X_k - Y_k \xrightarrow{n \to \infty} 0$ fast sicher und es gilt

$$\frac{S_n}{a_n} = \frac{1}{a_n} \sum_{k=1}^{n} Y_k + \frac{1}{a_n} \left(\sum_{k=1}^{n} X_k - Y_k\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

nach Satz 5.6. Als Notiz sei vermerkt, dass wegen der bereits genutzten Asymptotik auch die Wahl $a_n = \sqrt{n \log n}$ funktioniert hätte.

Aufgabe Z11.4

Sei $X_n \sim \mathrm{Ber}(1/n)$ unabhängig und für $n \in \mathbb{N}$ und sei $S_n = \sum_{k=1}^{n} X_k$. Zeigen Sie, dass

$$\frac{S_n - \log n}{\sqrt{\log n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Lösung. Wir erinnern an die Notation der harmonischen Zahlen erster un zweiter Ordnung, $H_n = \sum_{k=1}^{n} k^{-1} \sim \log n$ und $H_n^{(2)} = \sum_{k=1}^{n} k^{-2} \xrightarrow{n \to \infty} \pi^2/6$. Wegen $\mathbb{E}[X_n] = \frac{1}{n}$ und $\mathbb{V}\mathrm{ar}[X_n] = \frac{1}{n}(1 - \frac{1}{n})$ gilt also $\mathbb{E}[S_n] = H_n$ und $\mathbb{V}\mathrm{ar}[S_n] = H_n - H_n^{(2)}$. Definieren wir

$$Y_{n,k} = \frac{1}{\sqrt{H_n - H_n^{(2)}}} \left(X_k - \frac{1}{k}\right),$$

so erhalten wir mit obigen Beobachtungen ein unabhängiges, zentriertes, normiertes $\triangle$-Schema. Um die Lindeberg-Bedingung zu verifizieren, reicht die Beobachtung, dass $(X_k - \frac{1}{k})^2 < 1$ und damit $Y_{n,k}^2 \mathbb{1}_{\{Y_{n,k}^2 > \varepsilon\}} = Y_{n,k}^2 \mathbb{1}_{\{(X_k - \frac{1}{k})^2 > \varepsilon(H_n - H_n^{(2)})\}} = 0$ für n hinreichend groß. Nach dem Satz von Lindeberg-Feller gilt also

$$\frac{S_n - H_n}{\sqrt{H_n - H_n^{(2)}}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Um zur Behauptung der Aufgabe zu gelangen, definieren wir

$$c_n := \sqrt{\frac{H_n - H_n^{(2)}}{\log n}} \xrightarrow{n \to \infty} 1, \qquad d_n := \frac{H_n - \log n}{\sqrt{\log n}} \xrightarrow{n \to \infty} 0.$$

(Die zweite Konvergenz ist klar, da $H_n - \log n \to \gamma$, die Euler-Mascheroni-Konstante.) Damit gilt

$$\frac{S_n - \log n}{\sqrt{n}} = c_n \frac{S_n - H_n}{\sqrt{H_n - H_n^{(2)}}} + d_n \xrightarrow{\text{d}} \mathcal{N}(0,1)$$

nach Satz 5.6.3.

Übungsaufgaben

Aufgabe Ü11.5

Angenommen, $\sqrt{n}(X_n - \vartheta) \xrightarrow{\text{d}} \mathcal{N}(0,\sigma^2)$ und g ist stetig differenzierbar. Was ist der schwache Grenzwert von $\sqrt{n}(g(X_n) - g(\vartheta))$?

Lösung. Definiere $S_n = \sqrt{n}(X_n - \vartheta)$. Damit gilt also $X_n = \frac{S_n}{\sqrt{n}} + \vartheta$ und $X_n \xrightarrow{\text{d}} \vartheta$ (Satz 5.6). Es existiert ein Wahrscheinlichkeitsraum und darauf Zufallsvariablen T_n, sodass $P_{T_n} = P_{S_n}$ und $T_n \xrightarrow{\text{f.s.}} \mathcal{N}(0,\sigma^2)$. Dann ist $Y_n := \frac{T_n}{\sqrt{n}} + \vartheta$ verteilt wie X_n.
Wir sind interessiert an $\lim_{n\to\infty} \sqrt{n}(g(Y_n) - g(\vartheta))$. Da nach Voraussetzung g differenzierbar ist, können wir den Mittelwertsatz anwenden, der für jedes n eine ZV $Z_n \in [Y_n \wedge \vartheta, Y_n \vee \vartheta]$ garantiert, sodass $g(Y_n) - g(\vartheta) = g'(Z_n)(Y_n - \vartheta)$. Da $Y_n \xrightarrow{\text{f.s.}} \vartheta$ konvergiert auch Z_n fast sicher gegen ϑ. Insgesamt erhalten wir

$$\sqrt{n}(g(Y_n) - g(\vartheta)) \xrightarrow{\text{f.s.}} g'(\vartheta)\mathcal{N}(0,\sigma^2) = \mathcal{N}(0, \sigma^2 g'(\vartheta)^2),$$

was auch der schwache Grenzwert von $\sqrt{n}(g(X_n) - g(\vartheta))$ ist.

Aufgabe Ü11.6

Seien $(X_n)_{n\in\mathbb{N}}$ unabhängige Zufallsvariablen mit $P(X_n = 2^n) = \frac{1}{2} = P(X_n = -2^n)$ und setze $S_n = \sum_{k=1}^n X_k$. Zeigen oder widerlegen Sie, dass

$$\frac{S_n}{\sqrt{\mathbb{V}\mathrm{ar}[S_n]}} \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

Lösung. Die Aussage ist falsch. Um dies zu zeigen, bemerken wir zuerst, dass $\mathbb{V}\mathrm{ar}[X_n] = 4^n$ und damit $\mathbb{V}\mathrm{ar}[S_n] = \sum_{k=1}^n 4^k = \frac{4^{n+1}-4}{3}$. Insbesondere ist $2^{-n}\sqrt{\mathbb{V}\mathrm{ar}[S_n]} \xrightarrow{n\to\infty} \frac{2}{\sqrt{3}} < \frac{6}{5}$.
Die nächste Beobachtung ist, dass wegen $\sum_{k=1}^n 2^k < 2^{n+1}$ der Wert von S_n maßgeblich von X_n und X_{n-1} beeinflusst wird:

$$P(S_n > 2^n) \geq P(X_{n-1} = 2^{n-1}, X_n = 2^n) = \frac{1}{4}.$$

Wir leiten hieraus einen Widerspruch ab, indem wir zeigen, dass $S_n/\sqrt{\mathbb{V}\mathrm{ar}[S_n]}$ eine höhere Wahrscheinlichkeit hat, große Werte anzunehmen als die Normalverteilung: Für große n gilt

$$P\left(\frac{S_n}{\sqrt{\mathbb{V}\mathrm{ar}[S_n]}} \geq \frac{5}{6}\right) = P\left(S_n \geq 2^n \frac{5}{6}\frac{\sqrt{\mathbb{V}\mathrm{ar}[S_n]}}{2^n}\right) \geq P(S_n \geq 2^n) \geq \frac{1}{4};$$

doch lässt sich einfach nachprüfen, dass $\mathcal{N}_{0,1}([\frac{5}{6},\infty)) < \frac{1}{4}$.

Aufgabe Ü11.7

Sei $S_n \sim \mathrm{Bin}(n, p_n)$ mit $p_n = n^{-a}$ und $a > 0$. Wann ist die Normalapproximation besser und wann die Poissonapproximation?

Lösung. Die Berry-Esseen-Ungleichung liefert die Fehlerabschätzung

$$C \cdot \frac{p_n^2 + (1-p_n)^2}{\sqrt{n}\sqrt{p_n(1-p_n)}},$$

die Poisson-Approximation liefert

$$2np_n^2.$$

Ein Vergleich für exakte Wahrscheinlichkeiten hängt von C ab und ist nur numerisch möglich. Wir können jedoch eine asymptotische Aussage treffen: Die Normalapproximation ist asymptotisch besser, falls der Ausdruck

$$C \cdot \frac{p_n^2 + (1-p_n)^2}{\sqrt{n}\sqrt{p_n(1-p_n)}} \cdot \frac{1}{2np_n^2} \sim \frac{C/2}{\sqrt{np_n}np_n^2} = \frac{C/2}{n^{\frac{3}{2}-\frac{5}{2}a}}$$

für $n \to \infty$ gegen Null strebt. Dies ist für $a < \frac{3}{5}$, der Fall, während für $a > \frac{3}{5}$ die Poisson-Approximation asymptotisch besser ist.

Aufgabe Ü11.8

Seien $(X_n)_{n\in\mathbb{N}}$ unabhängige Zufallsvariablen mit

$$P(X_n = \pm 1) = \frac{1}{2}\left(1 - \frac{1}{n^2}\right), \qquad P(X_n = \pm n) = \frac{1}{2n^2}.$$

Zeigen Sie, dass $(X_k/\sqrt{a_n})_{n\in\mathbb{N}, 1\le k\le n}$ mit $a_n := \sum_{k=1}^n (2 - k^{-2})$ ein zentriertes, normiertes, unabhängiges Dreiecksschema ist, aber

$$\frac{\sum_{k=1}^n X_k}{\sqrt{a_n}} \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \tfrac{1}{2}\right) \quad (n \to \infty).$$

Lösung.

- Unabhängig: Die (X_k) sind unabhängig, also auch die $(X_k/\sqrt{a_n})$.
- Zentriert: $\mathbb{E}[a_n^{-1/2} X_k] = a_n^{-1/2}\mathbb{E}[X_k] = 0$.
- Normiert: $\sum_{k=1}^n \mathbb{V}\mathrm{ar}[a_n^{-1/2} X_k] = a_n^{-1} \sum_{k=1}^n \mathbb{E}[X_k^2] = a_n^{-1} \sum_{k=1}^n (2 - \frac{1}{k^2}) = 1$.

Es verbleibt, die Konvergenz zu zeigen. Definiere hierfür $Y_n = \mathrm{sgn} X_n$ (also ist Y_n die einfache symmetrische Irrfahrt), setze weiterhin $S_n = \sum_{k=1}^n X_k, T_n = \sum_{k=1}^n Y_k$. Dann ist $\mathbb{E}[Y_n] = 0, \mathbb{V}\mathrm{ar}[Y_n] = 1$ und nach dem ZGS gilt $T_n/\sqrt{n} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$.
Wir zeigen nun, dass $a_n^{-1/2} T_n \xrightarrow{\mathrm{d}} \mathcal{N}(0, \frac{1}{2})$ und $|a_n^{-1/2} T_n - a_n^{-1/2} S_n| \xrightarrow{P} 0$. Denn damit folgt mit den Rechenregeln zur schwachen Konvergenz, dass

$$\underbrace{a_n^{-1/2} S_n - a_n^{-1/2} T_n}_{\xrightarrow{\mathrm{d}} 0} + \underbrace{a_n^{-1/2} T_n}_{\xrightarrow{\mathrm{d}} \mathcal{N}(0,\frac{1}{2})} \xrightarrow{\mathrm{d}} 0 + \mathcal{N}(0, \tfrac{1}{2}).$$

Für die erste Behauptung beweisen wir folgendes

Lemma 1. *Es gelte $Z_n \xrightarrow{\mathrm{d}} Z$ und $c_n \xrightarrow{n\to\infty} c$. Dann gilt $c_n X_n \xrightarrow{\mathrm{d}} cX$.*

Die Aussage ist Teil des Satzes 5.6.

Beweis. Wir nehmen $c > 0$ an ($c \leq 0$ lässt sich analog behandeln). Sei $\varepsilon > 0$, dann ist $c_n \leq c + \varepsilon$ für hinreichend große n. Damit gilt

$$P(c_n X_n \leq x) \geq P(X_n \leq \tfrac{x}{c+\varepsilon}) \xrightarrow{n\to\infty} P(X \leq \tfrac{x}{c+\varepsilon})$$

falls $x/(c+\varepsilon)$ ein Stetigkeitspunkt der Verteilung von X ist. Senden wir $\varepsilon \to 0$ und ist x ein Stetigkeitspunkt der Verteilung von cX, so folgt

$$\liminf_{n\to\infty} P(c_n X_n \leq x) \geq P(cX \leq x).$$

Sei nun $0 < \varepsilon < c$ und n groß genug, dass $c_n \geq c - \varepsilon$. Ähnlich wie zuvor gilt $P(c_n X_n \leq x) \leq P(X_n \leq \frac{x}{c-\varepsilon})$. Mit selber Argumentation ergibt sich

$$\limsup_{n\to\infty} P(c_n X_n \leq x) \leq P(cX \leq x),$$

was das Lemma beweist. □

Nun gilt $T_n/\sqrt{n} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ und $\sqrt{\frac{n}{a_n}} \xrightarrow{n\to\infty} \sqrt{\frac{1}{2}}$. Mit Anwendung des Lemmas folgt also $T_n/\sqrt{a_n} \xrightarrow{\mathrm{d}} \sqrt{\frac{1}{2}}\,\mathcal{N}(0,1) = \mathcal{N}(0,\frac{1}{2})$.
Um die zweite Behauptung zu zeigen, bemerken wir, dass $\sum_{k\geq 1} P(X_k \neq Y_k) = \sum_{k\geq 1} k^{-2} < \infty$. Nach Borel-Cantelli ist also fast sicher $X_k \neq Y_k$ für nur endlich viele k. Doch damit gilt

$$|a_n^{-1/2} T_n - a_n^{-1/2} S_n| \leq a_n^{-1/2} \sum_{k=1}^{n} |Y_k - X_k| \xrightarrow{n\to\infty} 0$$

fast sicher, da $a_n \xrightarrow{n\to\infty} \infty$. Insbesondere folgt die behauptete stochastische Konvergenz und damit, was zu zeigen war.