UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Einführung in die Wahrscheinlichkeitstheorie und Statistik
Prof. Dr. Rainer Dahlhaus
Nathawut Phandoidaen
Wintersemester 2019/2020

12. Übungsblatt – Lösungen

Aufgabe 45	Aufgabe 46	Aufgabe 47	Aufgabe 48	Summe:

Übungsgruppe: Tutor(in):

Namen:

Aufgabe 45 (Schwache Konvergenz, 4 = 1 + 1.5 + 1.5 Punkte).
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen.

(a) Zeigen Sie, dass das schwache Gesetz der großen Zahlen aus dem zentralen Grenzwertsatz folgt.

(b) X_n besitze die Wahrscheinlichkeitsdichte $f_n(x) = \frac{n+1}{2}|x|^n \mathbb{1}_{(-1,1)}(x)$, $x \in \mathbb{R}$. Gibt es eine Zufallsvariable Z mit $X_n \xrightarrow{\mathcal{D}} Z$? Geben Sie Z bei Existenz an.

(c) Entscheiden Sie jeweils, ob eine Zufallsvariable Z existiert mit $X_n \xrightarrow{\mathcal{D}} Z$ für

(i) $X_n \sim U[0, 1 + \frac{1}{n}]$,
(ii) $X_n \sim \mathrm{Exp}(n)$,
(iii) $X_n \sim \mathrm{Exp}(1/n)$.

Lösung:

(a) Sind die $(X_n)_{n\in\mathbb{N}}$ eine iid Folge von Zufallsvariablen mit $\mathrm{Var}(X_1) < \infty$, so gilt nach dem zentralen Grenzwertsatz

$$\sqrt{n}\frac{\overline{X_n} - \mathbb{E}[X_1]}{\sqrt{\mathrm{Var}(X_1)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

bzw.

$$\sqrt{n}(\overline{X_n} - \mathbb{E}[X_1]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}(X_1)).$$

Es gilt $\frac{1}{\sqrt{n}} \xrightarrow{p} 0$. Nach Proposition 14.9(ii) folgt

$$\overline{X_n} - \mathbb{E}[X_1] = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\overline{X_n} - \mathbb{E}[X_1]) \xrightarrow{p} 0.$$

Damit folgt $\overline{X_n} \xrightarrow{p} \mathbb{E}[X_1]$, d.h. das schwache Gesetz der großen Zahlen.

(b) Um die schwache Konvergenz zu untersuchen, müssen wir zunächst die Verteilungsfunktion von X_n bestimmen. Für $x \in (-1,1)$ gilt

$$F_n(x) = \int_{-\infty}^{x} f_n(y)\mathrm{d}y = \frac{n+1}{2}\int_{-1}^{x} |y|^n \mathrm{d}y.$$

Im Falle $x \in (-1,0]$ folgt nun

$$F_n(x) = \frac{n+1}{2}(-1)^n \int_{-1}^{x} y^n \mathrm{d}y = \frac{(-1)^n}{2}\Big[y^{n+1}\Big]_{-1}^{x} = \frac{(-1)^n}{2}\Big(x^{n+1} - (-1)^{n+1}\Big) = \frac{1}{2}(1-(-x)^{n+1}).$$

Insbesondere folgt daher $F_n(0) = \frac{1}{2} = \frac{n+1}{2}\int_{-1}^{0} |y|^n \mathrm{d}y$.
Für $x \in (0,1)$ gilt

$$F_n(x) = \frac{n+1}{2}\int_{-1}^{x} |y|^n \mathrm{d}y = \frac{n+1}{2}\int_{-1}^{0} |y|^n \mathrm{d}y + \frac{n+1}{2}\int_{0}^{x} y^n \mathrm{d}y = \frac{1}{2} + \frac{1}{2}\Big[y^{n+1}\Big]_{0}^{x} = \frac{1}{2}(1+x^{n+1}).$$

Alternativ können wir für $x \in (0,1)$ folgendes Argument verwenden:
Da f_n achsensymmetrisch zum Ursprung ist, wissen wir, dass $F_n(x) = 1 - F_n(-x)$ gelten muss. Daraus kann das Ergebnis für $x \in (0,1)$ direkt aus dem Ergebnis für $x \in (-1,0)$ abgelesen werden.

Wir weisen die schwache Konvergenz nach. Um zu ermitteln, ob X_n schwach konvergiert, untersuchen wir den punktweisen Limes von F_n für $n \to \infty$. Hier ist

$$F_n(x) = \begin{cases} 0, & x \leq -1, \\ \frac{1}{2}(1-(-x)^{n+1}), & x \in (-1,0], \\ \frac{1}{2}(1+x^{n+1}), & x \in (0,1), \\ 1, & x \geq 1 \end{cases} \to \begin{cases} 0, & x \leq -1, \\ \frac{1}{2}, & x \in (-1,1), \\ 1, & x \geq 1 \end{cases} =: \quad \tilde{F}(x).$$

$\tilde{F}$ ist noch keine Verteilungsfunktion, da sie nicht rechtsseitig stetig ist. Wir können $\tilde{F}$ aber an der Unstetigkeitsstellen $x = -1$ so modifizieren, dass sie rechtsseitig stetig und damit zur Verteilungsfunktion wird. Definiere

$$F(x) := \begin{cases} 0, & x < -1, \\ \frac{1}{2}, & x \in [-1,1), \\ 1, & x \geq 1. \end{cases}$$

F ist eine Verteilungsfunktion und es gilt $F_n(x) \to F(x)$ für alle Stetigkeitspunkte x von F (d.h. für alle $x \in \mathbb{R}\backslash\{-1,1\}$). Damit ist gezeigt, dass $X_n \xrightarrow{\mathcal{D}} Z$, wobei Z eine Zufallsvariable mit Verteilungsfunktion F ist.
Genauer ist Z eine diskrete Zufallsvariable mit $\mathbb{P}(Z=-1) = \mathbb{P}(Z=1) = \frac{1}{2}$.

(c) Wir ermitteln nun wie in (b) jeweils die Verteilungsfunktion F_n von X_n und deren Limes $\tilde{F}$, und nehmen eventuell Modifikationen vor, damit $\tilde{F}$ zur Verteilungsfunktion wird.

(i) Hier ist die Dichte $f_n(x) = \frac{1}{1+\frac{1}{n}}\mathbb{1}_{[0,1+\frac{1}{n}]}(x)$ und

$$F_n(x) = \begin{cases} 0, & x < 0, \\ \frac{x}{1+\frac{1}{n}}, & 0 \leq x \leq 1+\frac{1}{n}, \\ 1, & x > 1+\frac{1}{n} \end{cases} \to \begin{cases} 0, & x < 0, \\ x, & 0 \leq x \leq 1, \\ 1, & x > 1 \end{cases} =: \tilde{F}(x).$$

$\tilde{F}$ ist an der Stelle $x = 1$ nicht rechsseitig stetig. Wir modifizieren daher zu

$$F(x) := \begin{cases} 0, & x < 0, \\ x, & 0 \leq x < 1, \\ 1, & x \geq 1 \end{cases}.$$

Dann ist F rechtsseitig stetig (es ist die Verteilungsfunktion einer $U[0,1]$-Verteilung) und es gilt $F_n(x) \to F(x)$ für alle Stetigkeitspunkte x von F. Das bedeutet, es gibt $Z \sim U[0,1]$ mit $X_n \xrightarrow{\mathcal{D}} Z$.

(ii) Hier ist $F_n(x) = \left(1 - e^{-nx}\right)\mathbb{1}_{[0,\infty)}(x)$. Es gilt

$$F_n(x) = \left(1 - e^{-nx}\right)\mathbb{1}_{[0,\infty)}(x) \to \mathbb{1}_{(0,\infty)}(x) =: \tilde{F}(x).$$

$\tilde{F}$ ist an der Stelle $x = 0$ nicht rechtsseitig stetig. Wir modifizieren zu $F(x) := \mathbb{1}_{[0,\infty)}(x)$. Dann ist F rechtsseitig stetig (es ist die Verteilungsfunktion einer konstante Zufallsvariable $Z = 0$). Das bedeutet $X_n \xrightarrow{\mathcal{D}} 0$.

(iii) Hier ist $F_n(x) = \left(1 - e^{-x/n}\right)\mathbb{1}_{[0,\infty)}(x)$. Es gilt

$$F_n(x) = \left(1 - e^{-x/n}\right)\mathbb{1}_{[0,\infty)}(x) \to 0 =: \tilde{F}(x).$$

Angenommen, es gäbe eine Zufallsvariable Z mit $X_n \xrightarrow{\mathcal{D}} Z$. Dann müsste auch $F_n(x) \to 0 = F_Z(x)$ für alle Stetigkeitspunkte x von F_Z gelten.
Da F_Z nur abzählbar viele Unstetigkeitsstellen hat, gilt sicher $F_Z(x) = 0$ für beliebig große $x \in \mathbb{R}$. Aufgrund der Monotonie von F_Z folgt dann aber schon $F_Z \equiv 0$. Das ist ein Widerspruch zu der Eigenschaft $\lim_{x\to\infty} F_Z(x) = 1$ einer Verteilungsfunktion. Daher kann X_n nicht schwach konvergieren.

Aufgabe 46 (Schwache Konvergenz, Delta-Methode, 4 = 1 + 1.5 + 1.5 Punkte). Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und X eine weitere Zufallsvariable. Weiter sei $c \in R$ eine deterministische Konstante.

(a) Sei $X \sim \text{Bin}(1, \frac{1}{2})$ Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $\frac{1}{2}$, und $X_n := 1 - X$ für alle $n \in \mathbb{N}$. Zeigen Sie, dass $X_n \xrightarrow{\mathcal{D}} X$ aber nicht $X_n \xrightarrow{p} X$.

(b) Nun gelte $X_n \xrightarrow{\mathcal{D}} c$. Zeigen Sie, dass dann sogar $X_n \xrightarrow{p} c$ folgt.
Hinweis: Drücken Sie $\mathbb{P}(|X_n - c| > \varepsilon)$ mit der Verteilungsfunktion F_{X_n} von X_n aus.

(c) *Die Delta-Methode:* Sei $(a_n)_{n\in\mathbb{N}}$ eine Folge von positiven reellen Zahlen mit $a_n \to \infty$. Sei Z eine Zufallsvariable, sodass $a_n(X_n - c) \xrightarrow{\mathcal{D}} Z$. Sei $g : \mathbb{R} \to \mathbb{R}$ eine in einer Umgebung von c stetig differenzierbare Funktion mit $g'(c) \neq 0$. Zeigen Sie, dass dann folgt:

$$a_n(g(X_n) - g(c)) \xrightarrow{\mathcal{D}} g'(c) \cdot Z.$$

Hinweis: Nutzen Sie eine Taylor-Entwicklung von g um c: Für $x \in \mathbb{R}$ und ein $\tilde{x} \in \mathbb{R}$ mit $|\tilde{x} - c| \leq |x - c|$ gilt $g(x) = g(c) + (x-c)g'(c) + (x-c)\cdot(g'(\tilde{x}) - g'(c))$. Nutzen Sie Proposition 14.9 der Vorlesung.

Lösung:

(a) Die Zufallsvariablen $X_n = 1 - X$ besitzen dieselbe Verteilung wie die Zufallsvariable X, denn

$$\mathbb{P}((1-X)=0) = \mathbb{P}(X=1) = \frac{1}{2} = \mathbb{P}(X=0), \quad \mathbb{P}((1-X)=1) = \mathbb{P}(X=0) = \frac{1}{2} = \mathbb{P}(X=1).$$

Daher gilt $F_{X_n}(x) = F_X(x)$ für alle $x \in \mathbb{R}$ und insbesondere

$$F_{X_n}(x) \quad \stackrel{n\to\infty}{\to} \quad F_X(x) \qquad \text{für alle } x \in \mathbb{R}$$

Damit gilt $X_n \stackrel{\mathcal{D}}{\to} X$. Aber für $\varepsilon = \frac{1}{2}$ gilt

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|1-2X| > \frac{1}{2}) \stackrel{\text{Bild}(X)=\{0,1\}}{=} \mathbb{P}(X=0, X=1) = 1 \not\to 0 \quad (n \to \infty).$$

(b) Seien F_{X_n} die Verteilungsfunktionen der X_n und F_c die Verteilungsfunktion der konstanten Zufallsvariable c. Dann gilt

$$F_c(x) = \mathbb{P}(c \le x) = \begin{cases} 0, & x < c \\ 1, & x \ge c \end{cases}.$$

Wegen $X_n \stackrel{\mathcal{D}}{\to} c$ folgt dann $F_{X_n}(x) \stackrel{n\to\infty}{\to} F_c(x)$ für alle Stetigkeitspunkte x von F_c, d.h. für alle $x \in \mathbb{R} \setminus \{c\}$.

Sei nun $\varepsilon > 0$. Dann gilt

$$\begin{aligned}
\mathbb{P}(|X_n - c| > \varepsilon) &= \mathbb{P}(X_n - c > \varepsilon \text{ oder } X_n - c < -\varepsilon) \\
&= \mathbb{P}(X_n > c + \varepsilon \text{ oder } X_n < c - \varepsilon) \\
&= \mathbb{P}(X_n > c + \varepsilon) + \mathbb{P}(X_n < c - \varepsilon) \\
&\le 1 - \mathbb{P}(X_n \le c + \varepsilon) + \mathbb{P}(X_n \le c - \varepsilon) \\
&= 1 - F_{X_n}(c + \varepsilon) + F_{X_n}(c - \varepsilon) \\
&\stackrel{n\to\infty}{\to} 1 - F_c(c+\varepsilon) + F_c(c-\varepsilon) = 1 - 1 + 0 = 0.
\end{aligned}$$

Alternativ haben wir auch

$$\begin{aligned}
\mathbb{P}(|X_n - c| \ge \varepsilon) &= 1 - \mathbb{P}(|X_n - c| < \varepsilon) \\
&= 1 - \mathbb{P}(c - \varepsilon < X_n < c + \varepsilon) \\
&\le 1 - \mathbb{P}(c - \frac{\varepsilon}{2} < X_n \le c + \frac{\varepsilon}{2}) \\
&= 1 - \mathbb{P}(X_n \le c + \frac{\varepsilon}{2}) + \mathbb{P}(X_n \le c - \frac{\varepsilon}{2}) \\
&= 1 - F_{X_n}(c + \frac{\varepsilon}{2}) + F_{X_n}(c - \frac{\varepsilon}{2}) \\
&\stackrel{n\to\infty}{\to} 1 - F_c(c + \frac{\varepsilon}{2}) + F_c(c - \frac{\varepsilon}{2}) = 1 - 1 + 0 = 0.
\end{aligned}$$

Also gilt $X_n \stackrel{p}{\to} c$.

(c) Mit der Taylor-Entwicklung von g aus dem Hinweis folgt durch Einsetzen von $x = X_n$ mit einer reellwertigen Zufallsvariable $\tilde{X}_n$ mit $|\tilde{X}_n - c| \le |X_n - c|$, dass

$$\begin{aligned}
g(X_n) &= g(c) + g'(c) \cdot (X_n - c) + (X_n - c) \cdot (g'(\tilde{X}_n) - g'(c)) \\
\Longrightarrow \quad a_n \cdot (g(X_n) - g(c)) &= g'(c) \cdot a_n(X_n - c) + a_n(X_n - c) \cdot (g'(\tilde{X}_n) - g'(c)).
\end{aligned}$$

Nun gilt

△ trivialerweise $g'(c) \xrightarrow{p} g'(c)$,

△ $\tilde{X}_n \xrightarrow{p} c$, da $X_n \xrightarrow{p} c$ und $|\tilde{X}_n - c| \leq |X_n - c|$,
Es gilt $\varepsilon > 0$: $\mathbb{P}(|\tilde{X}_n - c| > \varepsilon) \leq \mathbb{P}(|X_n - c| > \varepsilon) \to 0$.

△ $g'(\tilde{X}_n) - g'(c) \xrightarrow{p} 0$,
Nach dem CMT (Proposition 14.9(iv)) gilt aufgrund der Stetigkeit von g' in c, dass $g'(\tilde{X}_n) \xrightarrow{p} g'(c)$. Das Verschieben von $g'(c)$ auf die linke Seite der stochastischen Konvergenz ist trivial, da $g'(\tilde{X}_n) - g'(c) \xrightarrow{p} 0$ und $g'(\tilde{X}_n) \xrightarrow{p} g'(c)$ eingesetzt in die Definition der stochastischen Konvergenz dieselben Aussagen liefern.

△ $a_n(X_n - c) \xrightarrow{\mathcal{D}} Z$ nach Voraussetzung.

Damit folgt

$$a_n \cdot (g(X_n) - g(c)) = \underbrace{\underbrace{\underbrace{g'(c)}_{\xrightarrow{p} g'(c)} \cdot \underbrace{a_n(X_n - c)}_{\xrightarrow{\mathcal{D}} Z}}_{\xrightarrow{\mathcal{D}} g'(c)\cdot Z \text{ (14.9(iii))}} + \underbrace{\underbrace{(g'(\tilde{X}_n) - g'(c))}_{\xrightarrow{p} 0} \cdot \underbrace{a_n(X_n - c)}_{\xrightarrow{\mathcal{D}} Z}}_{\xrightarrow{p} 0 \text{ (14.9(iii))}}}_{\xrightarrow{\mathcal{D}} g'(\mu)\cdot Z \text{ (14.9(iii))}} \quad \xrightarrow{\mathcal{D}} \quad g'(\mu) \cdot Z$$

Aufgabe 47 (Asymptotische Verteilung von Schätzern, 4 = 2 + 1 + 1 Punkte).

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$ exponentialverteilt mit Parameter λ. Zur Bestimmung von λ schlagen wir die folgenden beiden Schätzer vor:

$$\hat{\lambda}_1 := -\log\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > 1\}}\right), \qquad \hat{\lambda}_2 := \frac{1}{\overline{X_n}},$$

wobei $\overline{X_n} := \frac{1}{n}\sum_{i=1}^{n} X_i$ den Mittelwert bezeichnet.

(a) Zeigen Sie, dass $\sqrt{n}(\hat{\lambda}_i - \lambda) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_i^2)$ mit $\sigma_i > 0$, $i = 1, 2$, gilt.
Hinweis: Nutzen Sie den zentralen Grenzwertsatz und Aufgabe 46(c).

(b) Argumetieren Sie, warum $\hat{\lambda}_2$ zum Schätzen von λ besser geeignet ist.
Hinweis: Es gilt $e^x - 1 > x^2$ für alle $x > 0$.

(c) Sei $\alpha \in (0, 1)$. Zeigen Sie, dass $\mathbb{P}(\lambda \in S(X_1, ..., X_n)) \to 1 - \alpha$ für $n \to \infty$ gilt, wobei

$$S(X_1, ..., X_n) := \left[\hat{\lambda}_2\left(1 - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right), \hat{\lambda}_2\left(1 + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)\right]$$

mit $u_{1-\frac{\alpha}{2}}$ als $(1 - \frac{\alpha}{2})$-Quantil der Standardnormalverteilung.

Lösung:

(a) Wir betrachten $\hat{\lambda}_1$: Die Verteilungsfunktion von X_1 ist gegeben durch $F_{X_1}(x) = 1 - e^{-\lambda x}$. Es ist $\mathbb{E}[\mathbb{1}_{\{X_i > 1\}}] = \mathbb{P}(X_i > 1) = 1 - F_{X_1}(1) = e^{-\lambda}$ und

$$\text{Var}(\mathbb{1}_{\{X_i > 1\}}) = \mathbb{E}[\mathbb{1}^2_{\{X_i > 1\}}] - \mathbb{E}[\mathbb{1}_{\{X_i > 1\}}]^2 = e^{-\lambda}(1 - e^{-\lambda}) < \infty.$$

Damit ist der zentrale Grenzwertsatz anwendbar und wir erhalten

$$\begin{aligned} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > 1\}} - e^{-\lambda}\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i > 1\}} - \mathbb{E}[\mathbb{1}_{\{X_i > 1\}}]\right) \\ &\xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \text{Var}(\mathbb{1}_{\{X_i > 1\}})\Big) = \mathcal{N}(0, e^{-\lambda}(1 - e^{-\lambda})). \end{aligned}$$

Wir wenden nun die Delta-Methode mit der Funktion $g : \mathbb{R} \to \mathbb{R}, g(x) = -\log(x)$ an und ergänzen g im Bereich $x \in (-\infty, 0]$ beliebig. Die Funktion ist stetig differenzierbar in $c := e^{-\lambda} > 0$ mit $g'(c) = -\frac{1}{c} = -e^{\lambda}$. Damit erhalten wir

$$\sqrt{n}(\hat{\lambda}_1 - \lambda) = \sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i>1\}}\right) - g(e^{-\lambda})\right) \xrightarrow{\mathcal{D}} g'(c)\cdot\mathcal{N}(0, e^{-\lambda}(1-e^{-\lambda})) = \mathcal{N}(0, e^{\lambda}-1).$$

Wir betrachten $\hat{\lambda}_2$: Da X_1 exponentialverteilt ist, gilt $\mathbb{E}[X_1] = \frac{1}{\lambda}$ und $\operatorname{Var}(X_1) = \frac{1}{\lambda^2} < \infty$. Damit ist der zentrale Grenzwertsatz anwendbar und liefert

$$\sqrt{n}\left(\overline{X_n} - \frac{1}{\lambda}\right) = \sqrt{n}(\overline{X_n} - \mathbb{E}[X_1]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \operatorname{Var}(X_1)) = \mathcal{N}\left(0, \frac{1}{\lambda^2}\right)$$

Wir wenden nun die Delta-Methode mit der Funktion $g : \mathbb{R} \to \mathbb{R}, g(x) = \frac{1}{x}$ an. Diese Funktion ist stetig differenzierbar in $c := \frac{1}{\lambda} > 0$ mit $g'(c) = -\frac{1}{c^2} = \lambda^2$. Damit erhalten wir

$$\sqrt{n}(\hat{\lambda}_2 - \lambda) = \sqrt{n}\left(g(\overline{X_n}) - g\left(\frac{1}{\lambda}\right)\right) \xrightarrow{\mathcal{D}} g'(c)\cdot\mathcal{N}\left(0, \frac{1}{\lambda^2}\right) = \mathcal{N}(0, \lambda^2).$$

(b) Laut dem Hinweis ist stets $e^{\lambda} - 1 > \lambda^2$. Das bedeutet, die Varianz der asymptotischen Verteilung von $\hat{\lambda}_2$ ist kleiner als diejenige von $\hat{\lambda}_1$. Da Schätzer mit kleinerer (asymptotischer) Varianz zu bevorzugen sind (das bedeutet weniger Streuung auch im Fall für endliches $n \in \mathbb{N}$), sollte man hier eher $\hat{\lambda}_2$ zur Schätzung nutzen.

(c) Wir wissen aus (a), dass

$$\sqrt{n}(\hat{\lambda}_2 - \lambda) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda^2).$$

Insbesondere folgt mit Proposition 14.9(ii)

$$\hat{\lambda}_2 - \lambda = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\hat{\lambda}_2 - \lambda) \xrightarrow{\mathcal{D}} 0.$$

Aus Aufgabe 46(b) schließen wir $\hat{\lambda}_2 - \lambda \xrightarrow{p} 0$ bzw. $\hat{\lambda}_2 \xrightarrow{p} \lambda$.
Damit folgt mit Proposition 14.9(iii) gerade

$$Z_n := \sqrt{n}\frac{\hat{\lambda}_2 - \lambda}{\hat{\lambda}_2} \xrightarrow{p} \frac{1}{\lambda}\mathcal{N}(0, \lambda^2) = \mathcal{N}(0, 1).$$

Wir erhalten insgesamt

$$\begin{aligned}\mathbb{P}(\lambda \in S(X_1, ..., X_n)) &= \mathbb{P}\left(\left|\sqrt{n}\frac{\hat{\lambda}_2 - \lambda}{\hat{\lambda}_2}\right| \le u_{1-\frac{\alpha}{2}}\right) \\ &= \mathbb{P}(|Z_n| \le u_{1-\frac{\alpha}{2}}) \overset{Z_n \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0,1)}{\longrightarrow} \mathbb{P}(|Z| \le u_{1-\frac{\alpha}{2}}) = 1 - \alpha.\end{aligned}$$

Aufgabe 48 (Pearsons χ^2-Test, 4 Punkte).

Wir vermuten, dass bei einem 400-Meter-Rennen die Startposition einen Einfluss auf die Gewinnchancen ausübt. Die folgende Tabelle gliedert 144 Siegerinnen und Sieger nach der Nummer ihrer Startposition auf (wobei die Startpositionen von innen nach außen nummeriert sind):

Position	1	2	3	4	5	6	7	8
Häufigkeit	29	19	18	25	17	10	15	11

Testen Sie die Hypothese „gleiche Gewinnchancen“ gegen die Alternative „ungleiche Gewinnchancen“ zum Niveau $\alpha = 0.01$ mit Pearsons χ^2-Test!
Quantile: $\chi^2_{7,0.99} = 18.48$, $\chi^2_{8,0.99} = 20.09$.

Lösung:

Wir wollen Pearson's χ^2-Test (14.12) aus der Vorlesung anwenden. Wir haben Beobachtungen von $r = 8$ Typen (Gewinner/in startet von Startposition 1, ..., Gewinner/in startet von Startposition 8). Der Typ j (d.h. Gewinner/in startet von Startposition j) trete mit Wahrscheinlichkeit p_j auf. Dann ist $\sum_{j=1}^r p_j = 1$.
Wir haben eine Stichprobe $N = (N_1, ..., N_r)$, $n = \sum_{i=1}^r N_i = 144$ gegeben mit

$$N_1 = 29, \quad N_2 = 19, \quad N_3 = 18, \quad N_4 = 25, \quad N_5 = 17, \quad N_6 = 10, \quad N_7 = 15, \quad N_8 = 11$$

und wollen auf Basis dieser Stichprobe zum Niveau $\alpha = 0.01$ testen, ob die Nullhypothese

$$H_0 : p_j = \pi_j := \frac{1}{8}, \qquad (j = 1, ..., r) \qquad \text{(Gleiche Gewinnchancen auf jeder Startposition)}$$

vorliegt. Gemäß 14.12 lautet Pearson's χ^2-Test:

$$\phi(N_1, ..., N_r) = \begin{cases} 0, & \sum_{i=1}^r \frac{(N_i - n\cdot\pi_i)^2}{n\cdot\pi_i} \le c^* \\ 1, & \sum_{i=1}^r \frac{(N_i - n\cdot\pi_i)^2}{n\cdot\pi_i} > c^* \end{cases}$$

mit $c^* = \chi^2_{r-1,1-\alpha}$.
In diesem Fall ist

$$\begin{aligned} c^* &= \chi^2_{r-1,1-\alpha} = \chi^2_{7,0.99} = 18.48, \\ \sum_{i=1}^r \frac{(N_i - n \cdot \pi_i)^2}{n \cdot \pi_i} &= \sum_{i=1}^8 \frac{\left(N_i - 144 \cdot \frac{1}{8}\right)^2}{144 \cdot \frac{1}{8}} = \frac{(29-18)^2}{18} + ... + \frac{(11-18)^2}{18} = 16.33 \end{aligned}$$

Damit ist $\sum_{i=1}^r \frac{(N_i - n\cdot\pi_i)^2}{n\cdot\pi_i} \le c^*$, also $\phi(N_1, ..., N_8) = 0$. Auf Basis der Stichprobe lässt sich mit dem Test zum Niveau $\alpha = 0.01$ kein Einfluss der Startposition auf die Gewinnchancen feststellen.

Abgabe:
In Zweiergruppen, bis spätestens Donnerstag, den **23. Januar 2020, 11:00 Uhr**.
(Die Zettelkästen für das Abgabeblatt sind im 1. OG, INF 205, vor dem Dekanat.)

Homepage der Vorlesung:
`https://ssp.math.uni-heidelberg.de/ews-WS1920/index.html`