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

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

## 11. Übungsblatt – Lösungen

| Aufgabe 41 | Aufgabe 42 | Aufgabe 43 | Aufgabe 44 | Summe: |
|---|---|---|---|---|
| | | | | |

Übungsgruppe: Tutor(in):

Namen:

**Aufgabe 41 (Einstichproben $t$-Test, $4 = 2 + 1 + 1$ Punkte).**

In einem Einkaufsladen sehen wir Packungen, die $\mu_0 = 500$ Gramm Mehl beinhalten sollen. Wir sind skeptisch und wollen dem Hersteller nachweisen, dass er weniger als 500 Gramm Mehl in die Tüten abfüllt.

Dazu nehmen wir an, dass das Mehltütengewicht normalverteilt ist mit Erwartungswert $\mu \in \mathbb{R}$ (entspricht durchschnittlichem Mehltütengewicht) und unbekannter Standardabweichung $\sigma > 0$. Wir kaufen $n = 12$ Mehltüten und messen folgende Gewichte (in Gramm):

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 494.0 | 495.0 | 498.0 | 493.4 | 494.1 | 501.0 | 498.9 | 506.1 | 498.4 | 491.2 | 501.3 | 484.6 |

(a) Formulieren Sie zunächst die Hypothesen für das Testproblem, so dass der Fehler 1. Art dem peinlichen Irrtum entspricht, dass wir den Hersteller zu Unrecht beschuldigen. Führen Sie dann einen Student $t$-Test (Satz 13.8) zum Niveau $\alpha = 0.05$ durch und formulieren Sie das Ergebnis der Testentscheidung.

(b) Sei $\alpha \in (0, 1)$. Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Weiter sei $\overline{X_n} := \frac{1}{n}\sum_{i=1}^n X_i$ der Mittelwert und $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$ die empirische Stichprobenvarianz. Zeigen Sie, dass $S(X_1, ..., X_n) := \Big( -\infty, \overline{X_n} + \frac{S}{\sqrt{n}} t_{n-1,1-\alpha}\Big]$ ein $(1-\alpha)$-Konfidenzintervall für $\mu$ ist.
*Hinweis: Nutzen Sie das Ergebnis aus Satz 13.4(ii).*

(c) Geben Sie ein 95%-Konfidenzintervall für das durchschnittliche Mehltütengewicht $\mu$ basierend auf unseren Beobachtungen an.

*Quantile der $t$-Verteilung: $t_{11,0.95} = 1.796$, $t_{12,0.95} = 1.782$.*

**Lösung:**

(a) Wir haben Beobachtungen $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ (die gemessenen Mehltütengewichte) mit unbekanntem $\sigma > 0$. Wir wollen einen $t$-Test zu den Hypothesen

$$H_0 : \mu \geq \mu_0 = 500 \quad \text{gegen} \quad H_1 : \mu < \mu_0$$

durchführen. Da die Hypothesen nicht den Hypothesen aus Satz 13.8 entsprechen, muss der Test zunächst umgeformt werden. Wir definieren neue Beobachtungen $Y_i := -X_i$ und neue Parameter $\theta := -\mu$ bzw. $\theta_0 := -\mu_0$. Dann gilt $Y_1, ..., Y_n \overset{\text{iid}}{\sim} \mathcal{N}(-\mu, \sigma^2) = \mathcal{N}(\theta, \sigma^2)$. Der $t$-Test für $Y_1, ..., Y_n$ zum Niveau $\alpha$ für die Hypothesen

$$H_0 : \theta \le \theta_0 \quad \text{gegen} \quad H_1 : \theta > \theta_0 \tag{1}$$

lautet gemäß Satz 13.8 gerade

$$\phi^*(Y_1, ..., Y_n) = \begin{cases} 1, & T_n^Y > c^*, \\ 0, & T_n^Y \le c^*, \end{cases} \tag{2}$$

wobei $c^* = t_{n-1,1-\alpha}$ das $(1-\alpha)$-Quantil der $t_n$-Verteilung ist und

$$T_n^Y = \frac{\overline{Y_n} - \theta_0}{S_Y/\sqrt{n}} = \frac{\mu_0 - \overline{X_n}}{S/\sqrt{n}} =: -T_n,$$

wobei $S_Y^2 := \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y_n})^2$, $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$.
Nun können wir die Ergebnisse von dem Test mit den $Y_i$ auf den Test mit den $X_i$ übertragen. Die Hypothesen (1) werden zu

$$H : \mu \ge \mu_0 \quad \text{gegen} \quad H_1 : \mu < \mu_0$$

und der Test (2) wird zu

$$\phi^*(X_1, ..., X_n) = \begin{cases} 1, & -T_n > c^*, \\ 0, & -T_n \le c^*, \end{cases} = \begin{cases} 1, & T_n < -c^*, \\ 0, & T_n \ge -c^*, \end{cases}.$$

Hier gilt nun für $n = 12$ und $\alpha = 0.05$,

$$\begin{aligned} \overline{X_n} &:= \frac{1}{n}\sum_{i=1}^n X_i = 496.41, \\ S^2 &:= \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2 = 30.62, \\ T_n &:= \frac{\overline{X_n} - \mu_0}{S/\sqrt{n}} = -2.25. \end{aligned}$$

Wegen $t_{n-1,1-\alpha} = t_{11,0.95} = 1.796$ folgt

$$T_n = -2.25 < -1.796 = -t_{n-1,1-\alpha} = -c^*,$$

d.h. $\phi^*(X_1, ..., X_n) = 1$. Auf Basis unserer Beobachtungen entscheiden wir uns also dafür, den Hersteller zu kontaktieren und auf seinen Fehler hinzuweisen.

(b) Eine Möglichkeit ist, mit Satz 9.3 zu argumentieren und das Konfidenzintervall aus dem Test aus (a) zu konstruieren. Wir werden hier aber den direkten Weg wählen und mit der Definition zeigen, dass $S(X_1, ..., X_n)$ ein $(1-\alpha)$-Konfidenzintervall für $\mu$ ist.
Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Aus Satz 13.4(ii) wissen wir, dass

$$T_n := \frac{\overline{X_n} - \mu}{S/\sqrt{n}} \sim t_{n-1}. \tag{3}$$

Wir beachten, dass $T_n$ dieselbe Verteilung wie $-T_n$ besitzt, da die Wahrscheinlichkeitsdichte von $T_n$ achsensymmetrisch um Null ist. Daher gilt

$$\begin{aligned}
\mathbb{P}(\mu \in S(X_1, ..., X_n)) &= \mathbb{P}\left(\mu \leq \overline{X_n} + \frac{S}{\sqrt{n}} t_{n-1,1-\alpha}\right) \\
&= \mathbb{P}\Big(\underbrace{\frac{\mu - \overline{X_n}}{S/\sqrt{n}}}_{=-T_n} \leq t_{n-1,1-\alpha}\Big) \\
&\overset{T_n \text{ hat gleiche Vtgl. wie } -T_n}{=} \mathbb{P}(T_n \leq t_{n-1,1-\alpha}) \\
&\overset{3}{=} 1 - \alpha.
\end{aligned}$$

(c) Die Werte von $\overline{X_n}$ und $S^2$ wurden schon in (a) berechnet. Wir erhalten mit (b) daher folgendes 95%-Konfidenzintervall (d.h. mit $\alpha = 0.05$ und mit $n = 12$ entsprechend $t_{n-1,1-\alpha} = t_{11,0.975} = 1.796$)

$$\begin{aligned}
S(X_1, ..., X_n) &= \Big( -\infty, \overline{X_n} + \frac{S}{\sqrt{n}} t_{n-1,1-\alpha}\Big] \\
&= (-\infty, 499.28].
\end{aligned}$$

*Anmerkung: Da das Konfidenzintervall auch aus dem Test aus (a) hätte konstruiert werden können, ist klar, dass* $S(X_1, ..., X_n)$ *den Wert* $\mu_0 = 500$ *nicht beinhaltet (sonst hätte der Test aus (a) die Nullhypothese angenommen).*

**Aufgabe 42 (Zweistichproben $t$-Test, 4 Punkte).**

Ein/e Wissenschaftler/in möchte untersuchen, ob die Kenntnisse der Vektorrechnung von Schüler/innen der 12. Klasse eines Gymnasiums besser sind als die von Mathematikstudenten im 3. Semester. Dazu hat er eine Vergleichsarbeit entwickelt, die er nun in einer 12. Klasse mit 10 Schüler/innen sowie in einem Seminar mit 6 Mathematik-Studierenden lösen lässt. Der/die Wissenschaftler/in nimmt an, dass die dabei erreichte Punktzahl (in Prozent) normalverteilt ist mit Mittelwert $\mu_S$ (Schüler) bzw. $\mu_U$ (Studenten) und gleicher, unbekannter Standardabweichung $\sigma > 0$. Er erhält folgende Ergebnisse für die Punktzahl (in Prozent):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Schüler/in | 81.4 | 67.8 | 70.0 | 66.8 | 61.9 | 83.8 | 61.9 | 70.9 | 69.1 | 78.5 |
| Studierende | 54.2 | 67.6 | 58.7 | 62.5 | 61.4 | 65.3 | | | | |

Führen Sie einen Zweistichproben $t$-Test (Satz 13.11) zum Niveau $\alpha = 0.05$ durch und entscheiden Sie, ob der/die Wissenschaftler/in mit seiner/ihrer Annahme (Schüler/innen besser als Studierende) Recht hat oder nicht. Hierbei sollte der Fehler 1. Art dem peinlichen Irrtum entsprechen, dass die Schüler/innen als besser eingeschätzt werden, obwohl es nicht stimmt.
*Quantile der $t$-Verteilung:* $t_{14,0.95} = 1.761$, $t_{15,0.95} = 1.753$, $t_{16,0.95} = 1.746$.

**Lösung:**

Laut Aufgabe beobachten wir die Ergebnisse der Schüler/innen $X_1, ..., X_m \overset{\text{iid}}{\sim} \mathcal{N}(\mu_S, \sigma^2)$ und die Ergebnisse der Studierenden $Y_1, ..., Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu_U, \sigma^2)$, wobei die $X_1, ..., X_m, Y_1, ..., Y_n$ gemeinsam unabhängig sind. Hierbei ist $n = 6$ und $m = 10$.
Wir wollen einen Zweistichproben $t$-Test für die Hypothesen

$$H_0 : \mu_S \leq \mu_U \quad \text{gegen} \quad H_1 : \mu_S > \mu_U$$

durchführen. Dann ist der Fehler 1. Art gerade die Situation, sich dafür zu entscheiden, dass Schüler/innen besser sind als Studierende, obwohl dies gar nicht stimmt.

Laut Satz 13.11 lautet der Zweistichproben $t$-Test zum Niveau $\alpha$,

$$\phi^*(X_1,...,X_m,Y_1,...,Y_n) = \begin{cases} 1, & T_{m,n} > c^* \\ 0, & T_{m,n} \le c^*, \end{cases}$$

wobei $c^* = t_{m+n-2,1-\alpha}$ das $(1-\alpha)$-Quantil der $t_{m+n-2}$-Verteilung ist und $T_{m,n}$ unten definiert wird. Es ist

$$\begin{aligned}
\overline{X_m} &:= \frac{1}{m}\sum_{i=1}^m X_i = 71.21, \\
\overline{Y_n} &:= \frac{1}{n}\sum_{i=1}^n Y_i = 61.62, \\
S_X^2 &:= \frac{1}{m-1}\sum_{i=1}^m (X_i - \overline{X_m})^2 = 58.44, \\
S_Y^2 &:= \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y_n})^2 = 22.74, \\
S^2 &:= \frac{1}{m+n-2}\Big((m-1)S_X^2 + (n-1)S_Y^2\Big) = 45.69, \\
T_{m,n} &:= \frac{\sqrt{\frac{mn}{m+n}}(\overline{X_m} - \overline{Y_n})}{S} = 2.75.
\end{aligned}$$

Nun ist $t_{m+n-2,1-\alpha} = t_{14,0.95} = 1.761$, d.h. wir haben

$$T_{m,n} = 2.75 > 1.761 = t_{m+n-2,1-\alpha} = c^*.$$

Damit ist $\phi^*(X_1,...,X_m,Y_1,...,Y_n) = 1$, d.h. auf Basis des Tests und unseren Beobachtungen können wir davon ausgehen, dass die Schüler/innen tatsächlich besser sind als die Studierenden, und die Annahme des Wissenschaftlers stimmt.

**Aufgabe 43 (Test für autoregrssive Prozesse, 4 = 2 + 1 + 1 Punkte).**

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\theta \in (-1,1)$. Seien $X_0, \varepsilon_1, ..., \varepsilon_n : \Omega \to \mathbb{R}$ unabhängige und normalverteilte Zufallsvariablen mit $X_0 \sim \mathcal{N}\left(0, \frac{1}{1-\theta^2}\right)$ und $\varepsilon_i \sim \mathcal{N}(0,1)$, $i = 1,...,n$. $X_1,...,X_n$ seien rekursiv definiert durch $X_t := \theta \cdot X_{t-1} + \varepsilon_t$, $t = 1,...,n$. Wir gehen davon aus, dass wir nur $X_0,...,X_n$ beobachten. Sei nun $\theta_0 \in (-1,1)$. Wir wollen auf die Hypothesen $H_0 : \theta = \theta_0$ gegen $H_1 : \theta \neq \theta_0$ testen. Dafür schlagen wir folgenden Test vor:

$$\phi(X_0,...,X_n) := \begin{cases} 1, & W_n > k_n, \\ 0, & W_n \le k_n \end{cases} \qquad \text{mit} \qquad W_n := \frac{1}{n}\sum_{t=1}^n (X_t - \theta_0 X_{t-1})^2.$$

(a) Berechnen Sie $k_n > 0$, so dass $\phi$ ein Test zum Niveau $\alpha \in (0,1)$ ist.

(b) Sei $\alpha \in (0,1)$. Berechnen Sie ein $(1-\alpha)$-Konfidenzintervall $S(X_0,...,X_n)$ für den unbekannten Parameter $\theta \in (-1,1)$, zum Beispiel mit Satz 9.3.

(c) Wir beobachten nun $n = 100$ Werte $X_0,...,X_n$ und erhalten

$$\hat{c}_n(0) := \frac{1}{n}\sum_{t=1}^n X_{t-1}^2 = 2.900, \quad \tilde{c}_n(0) := \frac{1}{n}\sum_{t=1}^n X_t^2 = 2.906, \quad \hat{c}_n(1) := \frac{1}{n}\sum_{t=1}^n X_t X_{t-1} = 2.436.$$

Wir vermuten einen wahren Wert von $\theta_0 = 0.3$. Wie sieht die Testentscheidung von $\phi$ zum Niveau $\alpha = 0.05$ aus? Geben Sie außerdem ein 95%-Konfidenzintervall für $\theta$ an. *Quantile der* $\chi^2$*-Verteilung:* $\chi^2_{100,0.95} = 124.3$, $\chi^2_{100,0.05} = 77.9$.

**Lösung:**

(a) Gilt die Hypothese $\theta = \theta_0$, d.h.

$$X_t = \theta_0 \cdot X_{t-1} + \varepsilon_t, \quad t = 1, ..., n,$$

so folgt

$$X_t - \theta_0 \cdot X_{t-1} = \varepsilon_t, \quad t = 1, ..., n.$$

Daher ist

$$n \cdot W_n = \sum_{t=1}^{n} (X_t - \theta_0 X_{t-1})^2 = \sum_{t=1}^{n} \varepsilon_t^2 \sim \chi_n^2.$$

Es soll gelten

$$\begin{aligned} \alpha &\stackrel{!}{=} \mathbb{P}_{\theta_0}(\phi = 1) = \mathbb{P}(W_n > k_n) = \mathbb{P}(nW_n > nk_n) = 1 - \mathbb{P}(nW_n < nk_n) \\ \iff & \quad 1 - \alpha = \mathbb{P}(nW_n < nk_n), \end{aligned}$$

d.h. $nk_n = \chi^2_{n,1-\alpha}$ ($(1-\alpha)$-Quantil der $\chi_n^2$-Verteilung) erfüllt die Gleichung. Daher kann

$$k_n = \frac{\chi^2_{n,1-\alpha}}{n}$$

gewählt werden.

(b) Schreiben wir $\phi_{\theta_0} = \phi$ für den Test aus der Aufgabenstellung. Für den Annahmebereich des Tests gilt

$$\begin{aligned} A(\theta) &:= \{(x_0, ..., x_n) \in \mathbb{R}^{n+1} : \phi_\theta(x) = 0\} \\ &= \left\{ (x_0, ..., x_n) \in \mathbb{R}^{n+1} : \frac{1}{n} \sum_{t=1}^{n} (x_t - \theta x_{t-1})^2 \leq k_n \right\}. \end{aligned}$$

Laut Satz 9.3 ist dann

$$\begin{aligned} S(X_0, ..., X_n) &= \{\theta \in (-1,1) : (X_0, ..., X_n) \in A(\theta)\} \\ &= \left\{ \theta \in (-1,1) : \frac{1}{n} \sum_{t=1}^{n} (X_t - \theta X_{t-1})^2 \leq k_n \right\}. \end{aligned}$$

Nun gilt mit $\hat{c}_n(1) := \frac{1}{n}\sum_{t=1}^{n} X_t X_{t-1}$, $\hat{c}_n(0) := \frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2$, $\hat{\theta}_n := \frac{\hat{c}_n(1)}{\hat{c}_n(0)}$ und mit $\tilde{c}_n(0) := \frac{1}{n}\sum_{t=1}^{n} X_t^2$:

$$\Delta\hat{\theta}_n := \sqrt{\hat{\theta}_n^2 - \left( \frac{\tilde{c}_n(0) - k_n}{\hat{c}_n(0)} \right)}$$

Dann folgt

$$\begin{aligned} & W_n = \frac{1}{n} \sum_{t=1}^{n} (X_t - \theta X_{t-1})^2 = \tilde{c}_n(0) - 2\theta \hat{c}_n(1) + \theta^2 \hat{c}_n(0) \stackrel{!}{=} k_n \\ \iff & \theta^2 - 2\theta \frac{\hat{c}_n(1)}{\hat{c}_n(0)} + \frac{\tilde{c}_n(0) - k_n}{\hat{c}_n(0)} = 0 \\ \iff & \theta_{1/2} = \frac{\hat{c}_n(1)}{\hat{c}_n(0)} \pm \sqrt{\left( \frac{\hat{c}_n(1)}{\hat{c}_n(0)} \right)^2 - \left( \frac{\tilde{c}_n(0) - k_n}{\hat{c}_n(0)} \right)} = \hat{\theta}_n \pm \Delta\hat{\theta}_n. \end{aligned}$$

Das bedeutet

$$S(X_0, ..., X_n) = \left[ \hat{\theta}_n - \Delta\hat{\theta}_n, \hat{\theta}_n + \Delta\hat{\theta}_n \right] \cap (-1, 1)$$

ist ein gesuchtes Konfidenzintervall.

(c) Mit den gegebenen Werten und $\theta_0 = 0.3$, $n = 100$, $\alpha = 0.05$ erhalten wir

$$W_n = \frac{1}{n}\sum_{t=1}^{n}(X_t - \theta_0 X_{t-1})^2 = \tilde{c}_n(0) - 2\theta_0 \cdot \hat{c}_n(1) + \theta_0^2 \hat{c}_n(0) = 1.706.$$

Daher gilt mit $k_n = \frac{\chi^2_{n,1-\alpha}}{n} = 1.243$ gerade $W_n > k_n$, weswegen $\phi(X_0, ..., X_n) = 1$. D.h. wir verwerfen die Vermutung $\theta_0 = 0.3$ auf Basis des Tests.
Wir berechnen

$$\hat{\theta}_n = \frac{\hat{c}_n(1)}{\hat{c}_n(0)} = 0.840, \qquad \Delta\theta_n = \sqrt{\hat{\theta}_n^2 - \left(\frac{\tilde{c}_n(0) - k_n}{\hat{c}_n(0)}\right)} = 0.363.$$

Damit ist ein gesuchtes Konfidenzintervall laut (b)

$$S(X_0, ..., X_n) = \left[\hat{\theta}_n - \Delta\hat{\theta}_n, \hat{\theta}_n + \Delta\hat{\theta}_n\right] \cap (-1,1) = [0.477, 1.203] \cap (-1,1) = [0.477, 1).$$

**Aufgabe 44 (Anwendung ZGWS, SGGZ), 4 = 1.5 + 1 + 1.5 Punkte).**

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{Laplace}(\mu, b)$ Laplace-verteilt mit Dichten $f_{\mu,b}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$, wobei $\mu \in \mathbb{R}$ und $b > 0$ Parameter sind. Seien $\overline{X_n} := \frac{1}{n}\sum_{i=1}^n X_i$ und $S_n := \frac{1}{n}\sum_{i=1}^n |X_i - \overline{X_n}|$.

(a) Zeigen Sie, dass mit geeignetem $\sigma > 0$ gilt:
$\sqrt{n}(\overline{X_n} - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$ und $\tilde{S}_n := \frac{1}{n}\sum_{i=1}^n |X_i - \mu| \xrightarrow{p} b$.

(b) Zeigen Sie, dass $\frac{\sqrt{n}}{\sqrt{2}S_n}\left(\overline{X_n} - \mu\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.
*Hinweis: Zeigen Sie zunächst, dass* $S_n - \tilde{S}_n \xrightarrow{p} 0$. *Sie dürfen Proposition 14.9 verwenden*

(c) Konstruieren Sie mittels (a) einen (möglichst sinnvollen) Test $\phi(X_1, ..., X_n)$ für die Hypothesen $H_0 : \mu = \mu_0$ gegen $H_1 : \mu > \mu_0$ bei unbekanntem $b$, der ein vorgegebenes Niveau $\alpha \in (0,1)$ zumindest asymptotisch einhält, d.h. es soll gelten $\mathbb{P}_{\mu_0}(\phi = 1) \to \alpha$ $(n \to \infty)$.

**Lösung:**

(a) Zunächst ist

$$\begin{aligned}\mathbb{E}[X_1] &= \int_{\mathbb{R}} x f_{\mu,b}(x)\mathrm{d}x \overset{y:=x-\mu}{=} \frac{1}{2b}\int_{-\infty}^{\infty}(y+\mu)\exp\left(-\frac{|y|}{b}\right)\mathrm{d}y \\ &= \frac{1}{2b}\int_{-\infty}^{\infty} y\exp\left(-\frac{|y|}{b}\right)\mathrm{d}y + \mu \cdot \int_{-\infty}^{\infty}\frac{1}{2b}\exp\left(-\frac{|y|}{b}\right)\mathrm{d}y = 0 + \mu \cdot 1 = \mu,\end{aligned}$$

da beim ersten Integral über eine punktsymmetrische Funktion integriert wird und beim zweiten über die $f_{0,b}$-Dichte.
Weiter ist

$$\begin{aligned}\mathrm{Var}(X_1) &= \mathbb{E}[(X_1-\mu)^2] \overset{y:=x-\mu}{=} \frac{1}{2b}\int_{-\infty}^{\infty} y^2\exp\left(-\frac{|y|}{b}\right)\mathrm{d}y \overset{\text{Symm.}}{=} \frac{1}{b}\int_0^{\infty} y^2\exp(-y/b)\mathrm{d}y \\ &\overset{z:=y/b}{=} b^2 \cdot \int_0^{\infty} z^2 e^{-z}\mathrm{d}z = 2b^2 < \infty,\end{aligned}$$

da $\int_0^\infty z^2 e^{-z}\mathrm{d}z = [-z^2e^{-z}]_0^\infty + \int_0^\infty 2ze^{-z}\mathrm{d}z = [-2ze^{-z}]_0^\infty + \int_0^\infty 2e^{-z}\mathrm{d}z = [-2e^{-z}]_0^\infty = 2$.
Damit ist der Zentrale Grenzwertsatz 14.4 anwendbar und liefert

$$\sqrt{n}(\overline{X_n} - \mu) = \sqrt{n}(\overline{X_n} - \mathbb{E}[X_1]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}(X_1)) = \mathcal{N}(0, 2b^2).$$

Weil die $X_i$, $i = 1, ..., n$, gemeinsam unabhängig sind, sind auch die $|X_i - \mu|$, $i = 1, ..., n$, gemeinsam unabhängig. Außerdem gilt $\mathbb{E}[|X_1 - \mu|^2] < 2\mathbb{E}[X_1^2] + 2\mu^2 < \infty$ wegen $\mathrm{Var}(X_1) < \infty$. Wir haben

$$\begin{aligned}
\mathbb{E}|X_i - \mu| \quad &= \quad \int_{-\infty}^{\infty} |x-\mu| \; f_{\mu,b}(x)\mathrm{d}x \overset{y:=x-\mu}{=} \frac{1}{2b}\int_{-\infty}^{\infty} |y| \exp\left(-\frac{|y|}{b}\right) \mathrm{d}y \\
&\overset{\text{Symm.}}{=} \frac{1}{b}\int_0^{\infty} ye^{-y/b}\mathrm{d}y \overset{z:=y/b}{=} b \cdot \int_0^{\infty} ze^{-z}\mathrm{d}z = b,
\end{aligned}$$

da $\int_0^\infty ze^{-z}\mathrm{d}z = 1$ (partielle Integration).
Es folgt mit dem schwachen Gesetz der großen Zahlen

$$\tilde{S}_n = \frac{1}{n}\sum_{i=1}^n |X_i - \mu| \xrightarrow{p} \mathbb{E}|X_i - \mu| = b.$$

(b) Es gilt mit der umgekehrten Dreiecksungleichung

$$|S_n - \tilde{S}_n| \leq \frac{1}{n}\sum_{i=1}^n \Big| |X_i - \overline{X_n}| - |X_i - \mu| \Big| \leq \frac{1}{n}\sum_{i=1}^n \Big| (X_i - \overline{X_n}) - (X_i - \mu) \Big| = |\mu - \overline{X_n}|.$$

Nach dem starken Gesetz der großen Zahlen oder aus (a) gilt

$$\overline{X_n} \xrightarrow{p} \mu.$$

Damit folgt auch $|\overline{X_n} - \mu| \xrightarrow{p} 0$ und somit $S_n - \tilde{S}_n \xrightarrow{p} 0$.
Dies zeigt also

$$S_n = S_n - \tilde{S}_n + \tilde{S}_n \xrightarrow{p} 0 + b = b.$$

Mit Proposition 14.9(iii) folgt

$$\frac{\sqrt{n}}{\sqrt{2}S_n}\left(\overline{X_n} - \mu\right) = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sqrt{2}S_n} \xrightarrow{\mathcal{D}} \frac{1}{\sqrt{2}b}\mathcal{N}(0, 2b^2) = \mathcal{N}(0,1).$$

(c) Ist $X_1, ..., X_n \sim \mathrm{Laplace}(\mu_0, b)$, so folgt nach (b) entsprechend

$$T_n := \frac{\sqrt{n}}{\sqrt{2}S_n}\left(\overline{X_n} - \mu_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Wir schlagen daher folgenden Test vor:

$$\phi(X_1, ..., X_n) = \begin{cases} 1, & T_n > u_{1-\alpha}, \\ 0, & T_n \leq u_{1-\alpha} \end{cases},$$

wobei $u_{1-\alpha}$ das $(1-\alpha)$-Quantil der Standardnormalverteilung bezeichnet.

Zur „Sinnhaftigkeit“ des Tests $\phi$: Ein Test, der (asymptotisch) das Niveau $\alpha$ einhält, muss eigentlich nur die Bedingung erfüllen, dass der Fehler 1. Art gegen $\alpha$ konvergiert für $n \to \infty$. Allerdings ist ein Test besser und „sinnvoller“, je kleiner sein Fehler 2. Art ist. Man sollte sich bewusst machen, dass z.B. auch

$$\tilde{\phi}(X_1, ..., X_n) = \begin{cases} 1, & T_n < u_\alpha, \\ 0, & T_n \geq u_\alpha \end{cases}$$

ein legitimer Test zum Niveau $\alpha$ gewesen wäre. Trotzdem ist zu erwarten, dass $\tilde{\phi}$ einen wesentlich größeren Fehler 2. Art hat als $\phi$. Dies kann wie folgt begründet werden:

Da $\mathbb{E}X_i = \mu$, erwarten wir große $X_i$, falls ein großes $\mu$ vorliegt. Die Alternative $H_1 : \mu > \mu_0$ ist gerade die Menge an $\mu$, die größer als $\mu_0$ sind. Wenn also tatsächlich die Alternative vorliegt, sollte $X_i$ durchschnittlich größer als $\mu_0$ sein. Wir erwarten dann, dass $T_n$ größer ist als das $T_n$, was im Falle der Nullhypothese $H_0 : \mu = \mu_0$ herausgekommen wäre. Mit anderen Worten scheint es sinnvoll zu sein, sich für die Alternative zu entscheiden, wenn $T_n$ groß ist. Dies wird durch den Test $\phi$ gewährleistet, durch den Test $\tilde{\phi}$ aber nicht.
Der Nachweis, dass $\phi$ asymptotisch das Niveau $\alpha$ einhält, erigbt sich aus

$$\mathbb{P}_{\mu_0}(\phi = 1) = \mathbb{P}_{\mu_0}(T_n > u_{1-\alpha}) \overset{T_n \overset{\mathcal{D}}{\to} Z \sim \mathcal{N}(0,1)}{\to} \mathbb{P}(Z > u_{1-\alpha}) = \alpha.$$

---

**Abgabe:**
In Zweiergruppen, bis spätestens Donnerstag, den **16. 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`