DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Zulassungsprüfung Stochastik

der Deutschen Aktuarvereinigung e. V.

am 14. Oktober 2023

Hinweise:

- Als Hilfsmittel ist ein Taschenrechner zugelassen.
- Die Gesamtpunktzahl beträgt 120 Punkte. Die Klausur ist bestanden, wenn mindestens 60 Punkte erreicht werden.
- Bitte prüfen Sie die Ihnen vorliegende Prüfungsklausur auf Vollständigkeit. Die Klausur besteht aus 4 Seiten.
- **Alle Antworten sind zu begründen und bei Rechenaufgaben muss der Lösungsweg ersichtlich sein.**

Wir gehen stets von einem Maßraum $(\Omega, \mathcal{A}, \mu)$ bzw. einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ aus. Die Borel σ-Algebra auf $\mathbb{R}^n$ wird mit $\mathcal{B}^n$ bezeichnet, das Lebesgue Maß auf $\mathbb{R}^n$ wird mit λ^n bezeichnet.

Sollten Ihnen in Teilaufgaben Ergebnisse fehlen, dann treffen Sie eine plausible Annahme dafür.

Aufgabe 1. *[24 Punkte]*

Sei $f : [1, \infty) \longrightarrow \mathbb{R}$, $f(x) = \sum_{k=1}^{\infty} ke^{-kx}$. Beweisen Sie:

(a) f ist messbar bezüglich der durch $\mathcal{B}^1$ induzierten σ-Algebra auf $[1, \infty)$.

(b) Es gilt $\int_{[1,\infty)} f(x) d\lambda(x) = \sum_{k=1}^{\infty} \int_{[1,\infty)} ke^{-kx}\, d\lambda(x)$.

(c) Es gilt $\int_{[1,\infty)} f(x) d\lambda(x) = \frac{1}{e-1}$.

Aufgabe 2. *[24 Punkte]*

Die Kaffee AG füllt Kaffeebohnen in Packungen ab. Das Abfüllgewicht X einer Packung sei normalverteilt mit Erwartungswert 1 kg und Standardabweichung 0,05 kg. Eine Packung ist unverkäuflich, wenn sie weniger als 0,935 kg wiegt. Die Wochenproduktion beträgt $n = 10.000$ Packungen.

(a) Bestimmen Sie die Wahrscheinlichkeit, dass eine Packung unverkäuflich ist.

(b) Sei Y die Zufallsvariable „Anteil der unverkäuflichen Packungen der Wochenproduktion“. Bestimmen Sie den Erwartungswert und die Varianz von Y.

(c) Mit welcher Wahrscheinlichkeit sind mehr als 10 % der Wochenproduktion unverkäuflich? Verwenden Sie dazu die Approximation durch eine geeignete Normalverteilung.

(d) Welcher Anteil der unverkäuflichen Packungen an der Wochenproduktion wird mit einer Wahrscheinlichkeit von 95 % nicht überschritten?

(e) Eine neue Geschäftsleitung beschließt die Packungsgrößen um 10 % zu verringern, für das Abfüllgewicht gelte $Z = 0,9X$. Bestimmen Sie die neue Grenze d für unverkäufliche Verpackungen so, dass die Wahrscheinlichkeit für Unverkäuflichkeit 5 % beträgt.

Aufgabe 3. *[24 Punkte]*

Gegeben seien die Zufallsvariablen N und Θ mit $\Theta \sim U(a,b)$ wobei $a < b$, $[a,b] \subset (0,1)$. Für $\theta \in [a,b]$ sei $N \sim NB(1,\theta)$, d.h. N besitzt die Zähldichte

$$f(n|\theta) = \theta(1-\theta)^n, \quad n \in \mathbb{N}_0.$$

(a) Bestimmen Sie $E\left(\frac{1}{\Theta}\right)$ und $E\left(\frac{1}{\Theta^2}\right)$ $\left[\text{Ergebnis: } \frac{1}{b-a}\ln\left(\frac{b}{a}\right), \frac{1}{ab}\right]$

(b) Bestimmen Sie $E(N|\Theta)$ und $Var(N|\Theta)$.

(c) Bestimmen Sie $E(N)$ und $Var(N)$.

(d) Sind die Zufallsvariablen N und Θ unabhängig?

(e) Sei nun $[a,b] = [0,1]$. Bestimmen Sie $P(N = n)$ für $n \in \mathbb{N}_0$.
Hinweis: Für $p, q \in \mathbb{N}_0$ gilt $\int_0^1 t^p(1-t)^q\,dt = \dfrac{p!q!}{(p+q+1)!}$.

Aufgabe 4. *[24 Punkte]*

Seien $X_1, \dots, X_n$, $Y_1, \dots, Y_n$ und $W_1, \dots, W_n$ unabhängige Zufallsvariablen mit

$$X_i \sim N(\mu_1, 1),\ Y_i \sim N(\mu_2, 1),\ W_i \sim N(\mu_1 + \mu_2, 1)$$

und $\mu_1, \mu_2 \in \mathbb{R}$. Wir bezeichnen mit $\overline{X}$, $\overline{Y}$ usw. die Mittelwerte der X_i, Y_i, usw. Auf der Basis aller $3n$ Zufallsvariablen erhält man eindeutig bestimmte ML-Schätzer für μ_1 und μ_2 (wird also für die Aufgabe vorausgesetzt).

(a) Beweisen Sie, dass die Maximum Likelihoodschätzer $\hat{\mu}_1$, $\hat{\mu}_2$ für μ_1, μ_2 notwendigerweise das lineare Gleichungssystem

$$\begin{aligned} 2\hat{\mu}_1 + \hat{\mu}_2 &= \overline{X} + \overline{W} \\ \hat{\mu}_1 + 2\hat{\mu}_2 &= \overline{Y} + \overline{W} \end{aligned}$$

erfüllen.

(b) Gegeben seien die Schätzer $\mu_1^* = \dfrac{\overline{W}}{3} + \dfrac{2\overline{X}}{3} - \dfrac{\overline{Y}}{3}$ und $\mu_2^* = \dfrac{\overline{W}}{3} + \dfrac{2\overline{Y}}{3} - \dfrac{\overline{X}}{3}$ für μ_1 und μ_2.

(i) Sind μ_1^* und μ_2^* erwartungstreu?

(ii) Es ist bekannt, dass $\overline{X}$ ein erwartungstreuer Schätzer von μ_1 ist. Würden Sie ihn dem Schätzer μ_1^* vorziehen? Begründen Sie Ihre Antwort.

Aufgabe 5. *[24 Punkte]*

In einer Studie wurde der Blutdruck von Rauchern und Nichtrauchern untersucht. Bei 11 Rauchern ergab sich ein mittlerer Blutdruck von 130. Bei 14 Nichtrauchern wurde ein Mittelwert von 123 festgestellt. Bei beiden Messreihen ergab sich eine empirische Varianz von 32,49. Es wird angenommen, dass der Blutdruck bei Rauchern und Nichtrauchern jeweils normalverteilt ist.

(a) Bestimmen Sie aus obigen Daten ein Schätzintervall für den Erwartungswert des Blutdrucks von Rauchern zum Niveau $1-\alpha = 0,95$.

(b) Wie groß müsste die Stichprobe sein, damit die Länge des Schätzintervalls in (a) kleiner als 1 ist?

Sie können ohne Beweis verwenden, dass für $\gamma > 0,5$ *die Folge* $(t_{n,\gamma})_{n\in\mathbb{N}}$ *der Quantile der t-Verteilung gegen das Quantil* u_γ *von* $N(0,1)$ *konvergiert.*

(c) Bisher ging man bei Nichtrauchern von einem Blutdruck von 120 aus. Wird diese Beobachtung bei einem Niveau von $0,05$ widerlegt?

(d) Testen Sie unter geeigneten Annahmen auf dem Niveau von $\alpha = 0,05$ die Hypothese, dass der Erwartungswert des Blutdrucks bei Rauchern und Nichtrauchern gleich ist.

Lösungsvorschläge

Aufgabe 1

Zu (a) *[8 Punkte]*

Für $n \in \mathbb{N}$ sei $f_n : [1,\infty) \longrightarrow \mathbb{R}$, $f_n(x) := \sum_{k=1}^{n} ke^{-kx}$. f_n ist als Summe stetiger Funktionen stetig, also auch messbar. Laut Definition ist $\lim_{n\to\infty} f_n = f$, also ist f als punktweiser Grenzwert messbarer Funktionen auch messbar.

Zu (b) *[8 Punkte]*

Sei f_n wie in (a). Es gilt für alle $n \in \mathbb{N}$

$$\int_{[1,\infty)} f_n(x)\,d\lambda(x) = \int_{[1,\infty)} \sum_{k=1}^{n} ke^{-kx}\,d\lambda(x) = \sum_{k=1}^{n} \int_{[1,\infty)} ke^{-kx}\,d\lambda(x)$$

und $f_n \geq 0$. Wegen $f_{n+1}(x) - f_n(x) = (n+1)e^{-(n+1)x} > 0$ ist die Funktionenfolge $(f_n)_{n\in\mathbb{N}}$ monoton wachsend. Damit erfüllt $(f_n)_{n\in\mathbb{N}}$ die Voraussetzungen des Satzes von der monotonen Konvergenz, und die Behauptung folgt nun wegen

$$\begin{aligned}\int_{[1,\infty)} f(x)\,d\lambda(x) &= \int_{[1,\infty)} \lim_{n\to\infty} f_n(x)\,d\lambda(x) = \lim_{n\to\infty} \int_{([,\infty)} f_n(x)\,d\lambda(x)\\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \int_{[1,\infty)} ke^{-kx}\,d\lambda(x) = \sum_{k=1}^{\infty} \int_{[1,\infty)} ke^{-kx}\,d\lambda(x).\end{aligned}$$

Zu (c) *[8 Punkte]*

Wir wenden (b) an. Hierzu berechnet man mittels des uneigentlichen Riemann-Integrals

$$\int_{[1,\infty)} ke^{-kx} d\lambda(x) = \int_1^\infty ke^{-kx} dx = -e^{-kx}\Big|_1^\infty = e^{-k}.$$

Mit (b) und der geometrischen Reihe folgt

$$\begin{aligned}\int_{[1,\infty)} f(x)\,d\lambda(x) &= \int_{[1,\infty)} \sum_{k=1}^{\infty} ke^{-kx}\,d\lambda(x) = \sum_{k=1}^{\infty} \int_{[1,\infty)} ke^{-kx}\,d\lambda(x) = \sum_{k=1}^{\infty} e^{-k}\\ &= \sum_{k=1}^{\infty} \left(e^{-1}\right)^k = \sum_{k=0}^{\infty} \left(e^{-1}\right)^k - 1 = \frac{1}{1-e^{-1}} - 1 = \frac{1}{e-1}.\end{aligned}$$

Aufgabe 2

Zu (a) *[2 Punkte]*

$$P(X < 0,935) = \Phi\left(\frac{0,935-1}{0,05}\right) = \Phi(-1,3) = 1 - \Phi(1,3) = 0,0968$$

Zu (b) *[7 Punkte]*

Sei N die Zufallsvariable „Anzahl der unverkäuflichen Packungen einer Wochenproduktion". Dann ist $N \sim B(n,p)$ mit $p = 0,0968$ und $n = 10.000$. Somit ist $E(N) = np$ und $Var(N) = np(1-p)$. Es gilt $Y = \frac{N}{n}$ mit

$$\begin{aligned} E(Y) &= E\left(\frac{N}{n}\right) = \frac{1}{n}E(N) = \frac{1}{n}np = p = 0,0968 \\ Var(Y) &= Var\left(\frac{N}{n}\right) = \frac{1}{n^2}Var(N) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n} = \frac{0,08743}{10000} \\ &= 8,743 \cdot 10^{-6}. \end{aligned}$$

Zu (c)*[4 Punkte]*

Näherungsweise gilt $N \sim N(np, np(1-p))$ und damit auch $Y \sim N\left(p, \dfrac{p(1-p)}{n}\right)$. Es folgt

$$\begin{aligned} P(Y > 0,1) &= 1 - P(Y \leq 0,1) \\ &= 1 - \Phi\left(\frac{0,1-0,0968}{\sqrt{p(1-p)/n}}\right) = 1 - \Phi\left(\frac{0,1-0,0968}{\sqrt{p(1-p)}}\sqrt{n}\right) \\ &\approx 1 - \Phi(1,08) = 1 - 0,8599 = 0,1401. \end{aligned}$$

Zu (d) *[5 Punkte]*

Gesucht ist y mit $P(Y \leq y) = 0,95$. Mit der Normalverteilungsannahme folgt

$$\Phi(u_{0,95}) = 0,95 = P(Y \leq y) = \Phi\left(\frac{y-p}{\sqrt{p(1-p)}}\sqrt{n}\right) \text{ also}$$

$$\frac{y-p}{\sqrt{p(1-p)}}\sqrt{n} = u_{0,95} = 1,64 \text{ und damit } y = p + u_{0,95}\frac{\sqrt{p(1-p)}}{\sqrt{n}} = 0,1016.$$

Der Anteil, der mit einer Wahrscheinlichkeit von 95 % nicht überschritten wird, beträgt 10,16 %.

Zu (e) *[6 Punkte]*

Gesucht ist d mit $P(Z \leq d) = 0,05$. Da $Z = 0,9X$ und $X \sim N(1,\ 0,05^2)$ folgt

$$0,05 = P(Z \leq d) = P\left(X \leq \frac{d}{0,9}\right) = \Phi\left(\frac{d-0,9}{0,9 \cdot 0,05}\right) \text{ also}$$

$$\frac{d-0,9}{0,9 \cdot 0,05} = u_{0,05} = -u_{0,95} = -1,64 \text{ und damit}$$

$$d = 0,9 - 0,045 \cdot 1,64 = 0,826.$$

Variante für die erste Gleichung: $Z = 0,9X \sim N(1,(0,9 \cdot 0,05)^2) = N(1,\ 0,045^2)$ und

$$0,05 = P(Z \leq d) = \Phi\left(\frac{d-0,9}{0,045}\right).$$

Aufgabe 3

Zu (a) *[4 Punkte]*

Wegen $\Theta \sim U[a,b]$ folgt

$$E\left(\frac{1}{\Theta}\right) = \frac{1}{b-a}\int_a^b \frac{1}{\theta}\,d\theta = \frac{1}{b-a}\ln\frac{b}{a}$$
$$E\left(\frac{1}{\Theta^2}\right) = \frac{1}{b-a}\int_a^b \frac{1}{\theta^2}\,d\theta = \frac{1}{b-a}\left(\frac{1}{a}-\frac{1}{b}\right) = \frac{1}{ab}.$$

Zu (b) *[5 Punkte]*

$$E(N|\Theta=\theta) = \frac{1-\theta}{\theta} = \frac{1}{\theta}-1 \Longrightarrow E(N|\Theta) = \frac{1}{\Theta}-1$$
$$Var(N|\Theta=\theta) = \frac{1-\theta}{\theta^2} \Longrightarrow Var(N|\Theta) = \frac{1-\Theta}{\Theta^2} = \frac{1}{\Theta^2}-\frac{1}{\Theta}.$$

Zu (c) *[8 Punkte]*

Mit dem Satz von der iterierten Erwartung folgt

$$E(N) = E(E(N|\Theta)) = E\left(\frac{1}{\Theta}-1\right) = E\left(\frac{1}{\Theta}\right)-1$$
$$\begin{aligned} Var(N) &= Var(E(N|\Theta)) + E(Var(N|\Theta)) = Var\left(\frac{1}{\Theta}-1\right) + E\left(\frac{1}{\Theta^2}-\frac{1}{\Theta}\right) \\ &= Var\left(\frac{1}{\Theta}\right) + E\left(\frac{1}{\Theta^2}\right) - E\left(\frac{1}{\Theta}\right) = E\left(\frac{1}{\Theta^2}\right) - E\left(\frac{1}{\Theta}\right)^2 + E\left(\frac{1}{\Theta^2}\right) - E\left(\frac{1}{\Theta}\right) \\ &= 2E\left(\frac{1}{\Theta^2}\right) - E\left(\frac{1}{\Theta}\right)^2 - E\left(\frac{1}{\Theta}\right) \end{aligned}$$

Damit folgt mit (a)

$$E(N) = \frac{1}{b-a}\ln\left(\frac{b}{a}\right)-1$$
$$Var(N) = \frac{2}{ab} - \frac{1}{(b-a)^2}\ln^2\left(\frac{b}{a}\right) - \frac{1}{b-a}\ln\left(\frac{b}{a}\right).$$

Zu (d) *[3 Punkte]*

Wegen $E(N|\Theta) \neq E(N)$ sind N und Θ nicht unabhängig.

Zu (e) *[4 Punkte]*

Mit dem Hinweis erhält man

$$P(N=n) = \int_0^1 P(N=n|\Theta=\theta)\,d\theta = \int_0^1 \theta(1-\theta)^n\,d\theta = \frac{n!}{(n+2)!} = \frac{1}{(n+1)(n+2)}.$$

Aufgabe 4

Zu (a) *[15 Punkte]*

Wegen der Unabhängigkeit der Zufallsvariablen lautet die gemeinsame Dichte

$$\begin{aligned} &f(x_1,\ldots,x_n,y_1,\ldots,y_n,w_1,\ldots,w_n) \\ &= \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}}\right)^3 \exp\left(-\frac{1}{2}\left[(x_i-\mu_1)^2+(y_i-\mu_2)^2+(w_i-(\mu_1+\mu_2))^2\right]\right). \end{aligned}$$

Daraus ergibt sich für die Parameter μ_1, μ_2 die Likelihood

$$L(\mu_1,\mu_2) = \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} \exp\left(-\frac{1}{2}\sum_{i=1}^n\left[(x_i-\mu_1)^2+(y_i-\mu_2)^2+(w_i-(\mu_1+\mu_2))^2\right]\right).$$

Durch Logarithmieren und anschließendes Differenzieren nach μ_1 folgt

$$\begin{aligned} \ell(\mu_1,\mu_2) &= \ln L(\mu_1,\mu_2) \\ &= \ln\left(\frac{1}{\sqrt{2\pi}}\right)^{3n} - \sum_{i=1}^n \frac{1}{2}\left((x_i-\mu_1)^2+(y_i-\mu_2)^2+(w_i-(\mu_1+\mu_2))^2\right), \\ \ell_{\mu_1}(\mu_1,\mu_2) &= \frac{2}{2}\sum_{i=1}^n\left((x_i-\mu_1)+(w_i-(\mu_1+\mu_2))\right) \\ &= -2n\mu_1 - n\mu_2 + \sum_{i=1}^n (x_i+w_i). \end{aligned}$$

Aus Symmetrieüberlegungen folgt analog in dem man μ_1, x_i durch μ_2, y_i ersetzt

$$\ell_{\mu_2}(\mu_1,\mu_2) = -2n\mu_2 - n\mu_1 + \sum_{i=1}^n (y_i+w_i).$$

Nullsetzen des Gradienten führt zu den behaupteten Gleichungen.

Zu (b) *[4+5=9 Punkte]*

(i) Wegen $E(\overline{X}) = \mu_1$, $E(\overline{Y}) = \mu_2$, $E(\overline{W}) = \mu_1+\mu_2$ folgt

$$\begin{aligned} E(\mu_1^*) &= -\frac{1}{3}\mu_2 + \frac{1}{3}(\mu_1+\mu_2) + \frac{2}{3}\mu_1 = \mu_1 \\ E(\mu_2^*) &= -\frac{1}{3}\mu_1 + \frac{1}{3}(\mu_1+\mu_2) + \frac{2}{3}\mu_2 = \mu_2, \end{aligned}$$

also sind die beiden Schätzer erwartungstreu.

(ii) Beide Schätzer sind erwartungstreu, also sind die mittleren quadratischen Fehler gleich den Varianzen. Wegen $Var(X_i) = Var(Y_i) = Var(W_i) = 1$ für alle $i \in \{1, ..., n\}$ gilt $Var(\overline{X}) = Var(\overline{Y}) = Var(\overline{W}) = 1/n$. Daraus folgt

$$\begin{aligned} mse(\mu_1^*) = Var(\mu_1^*) = Var\left(\frac{\overline{W} + 2\overline{X} - \overline{Y}}{3}\right) &= \frac{1}{9} Var\left(\overline{W} + 2\overline{X} - \overline{Y}\right) \\ &= \frac{1}{9}\left(Var(\overline{W}) + 4Var(\overline{X}) + Var(\overline{Y})\right) = \frac{1}{9} \cdot \frac{6}{n} = \frac{2}{3n} \end{aligned}$$

Wegen $Var(\overline{X}) = Var(\overline{Y}) = \frac{1}{n}$ folgt

$$mse(\overline{X}) = Var(\overline{X}) = \frac{1}{n} > \frac{2}{3n} = mse(\mu_1^*).$$

Der $mse(\mu_1^*)$ ist also kleiner als $mse(\overline{X})$, es ist also μ_1^* vorzuziehen. Dies überrascht nicht, da mehr Informationen (neben X_i auch die Y_j) einfließen.

Aufgabe 5

Sei $X \sim N(\mu_x, \sigma_x^2)$ bzw. $Y \sim N(\mu_y, \sigma_y^2)$ die Zufallsvariable Blutdruck bei Rauchern bzw. Nichtrauchern. Die Varianz ist unbekannt, aus diesem Grund benötigen wir die Quantile der Student-Verteilungen.

Zu (a) *[6 Punkte]*

Als Konfidenzintervall ergibt sich ein symmetrisches Intervall um $\overline{x}$. Wegen

$$\underbrace{t_{10,0.975}}_{2{,}228} \frac{5,7}{\sqrt{11}} \approx 3,83$$

erhalten wir

$$[126,17; 133,83].$$

Zu (b) *[6 Punkte]*

Zu lösen ist

$$t_{n,0.975} \frac{5,7}{\sqrt{n}} \leq 0,5.$$

Es wird $t_{n,\,0,975} \approx u_{0,975} = 1,96$ verwendet, also ist zu lösen

$$\frac{1,96 \cdot 5,7}{\sqrt{n}} \leq 0,5 \iff (2 \cdot 1,96 \cdot 5,7)^2 \leq n.$$

Es ergibt sich $n = 500$.

Zu (c) *[6 Punkte]*

Beim t-Test ergibt sich die Testgröße $t = \frac{123-120}{5,7}\sqrt{14} = 1,96$. Die Hypothese wird nicht verworfen wegen $t_{13;0,975} = 2.160$ und $|t| < t_{13;0,975}$.

Zu (d) *[6 Punkte]*

Bei unbekannter Varianz und zwei unabhängigen Stichproben wird der t-Test verwendet. Zu prüfen ist

$$H_0\colon \mu_x = \mu_y.$$

Testgröße ist

$$T(x,y) = \frac{\overline{x}-\overline{y}}{s}\sqrt{\frac{14\cdot 11}{11+14}},\ \text{mit } s^2 = \frac{10s_x^2 + 13s_y^2}{10+13} = 5,7^2 \text{ also}$$
$$T(x,y) \approx 3,05$$

Wegen $|T(x,y)| > t_{23,\,0,975} = 2,069$ wird H_0 verworfen.