DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Zulassungsprüfung Stochastik

der Deutschen Aktuarvereinigung e. V.

am 9. Oktober 2020

Hinweise:

- Alle Hilfsmittel sind zugelassen (open book).
- 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 11 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]*

Seien $f_k : \Omega \longrightarrow \mathbb{R}$, $n \in \mathbb{N}$ messbar, für alle $\omega \in \Omega$ konvergiere $\sum\limits_{k=1}^{\infty} |f_k(\omega)|$.

Beweisen Sie:

(a) *[4 Punkte]* Die Funktionen $\sum\limits_{k=1}^{\infty} f_k$ und $\sum\limits_{k=1}^{\infty} |f_k|$ sind messbar.

(b) *[10 Punkte]* Es gilt

$$\int_\Omega \sum_{k=1}^{\infty} |f_k| \, d\mu = \sum_{k=1}^{\infty} \int_\Omega |f_k| \, d\mu.$$

(c) *[10 Punkte]* Es sei zusätzlich $\sum\limits_{k=1}^{\infty} \int_\Omega |f_k| d\mu < \infty$. Dann ist die Funktion $\sum\limits_{k=1}^{\infty} f_k$ integrierbar und es gilt

$$\int_\Omega \sum_{k=1}^{\infty} f_k \, d\mu = \sum_{k=1}^{\infty} \int_\Omega f_k \, d\mu.$$

Aufgabe 2. *[20 Punkte]*

Seien X, Y Zufallsvariablen, $Y \sim \Gamma(2, \lambda)$ mit $\lambda > 0$. Die Zufallsvariable X sei bei gegebenem $Y = y$ gleichverteilt $U[0, y]$.

(a) *[8 Punkte]* Bestimmen Sie den Erwartungswert und die Varianz von X.

(b) *[8 Punkte]* Zeigen Sie, dass für die gemeinsame Dichte f von X, Y gilt:

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} 1_{[x,\infty)}(y) & \text{falls } x, y > 0 \\ 0 & \text{sonst.} \end{cases}$$

(c) *[4 Punkte]* Bestimmen Sie die Dichte von X. Welche Verteilung besitzt X?

Aufgabe 3. *[28 Punkte]*

Das Ergebnis T eines IQ-Tests (Intelligenzquotient) ist $N(100, 15^2)$ verteilt. Um Mitglied bei Mensa zu werden, muss man einen IQ von mindestens 130 nachweisen.

(a) *[2 Punkte]* Bestimmen Sie $P(T > 130)$.

(b) *[10 Punkte]* Sei Φ die Verteilungsfunktion der Standardnormalverteilung. Weisen Sie nach, dass für $t > 130$

$$P(T \leq t | T > 130) = \frac{\Phi\left(\frac{t-100}{15}\right) - \Phi(2)}{1 - \Phi(2)}$$

gilt, und bestimmen Sie daraus die Dichte und den Erwartungswert der Verteilung des IQ von potentiellen und tatsächlichen Mensa-Mitgliedern.

Hinweis: Sie können verwenden, dass für die Dichte φ von $N(0, 1)$

$$\int t\varphi(t)\,dt = -\varphi(t) + C$$

gilt.

(c) *[16 Punkte]* Seien $T_1, \ldots, T_n \overset{\text{iid}}{\sim} N(100, 15^2)$ Ergebnisse von n IQ-Tests und $\overline{T} := \frac{1}{n}\sum_{i=1}^{n} T_i$.

(i) Bestimmen Sie Erwartungswert und Varianz von $\overline{T}$.

(ii) Bestimmen Sie $\delta > 0$ so, dass $P(\overline{T} \in I) = 0,9$ für das symmetrische Intervall $I = [100 - \delta, 100 + \delta]$ gilt .

(iii) Bestimmen Sie das kleinste $n \in \mathbb{N}$ so, dass in (ii) $\delta < 1$ gilt.

Aufgabe 4. *[24 Punkte]* Seien

$$Y_i = px_i + \varepsilon_i,\ i = 1, \ldots, n$$

mit $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $p \in \mathbb{R}$. Es werden $(y_1, x_1), \ldots, (x_n, y_n)$ beobachtet, $x_i \in \mathbb{R}$ fest, und y_i eine Realisierung von Y_i.

(a) *[16 Punkte]* Bestimmen Sie einen Maximum Likelihood Schätzer $\hat{p}$ für p. Ist er erwartungstreu?

(b) *[4 Punkte]* Sei $\hat{\hat{p}} := \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{x_i}$. Ist $\hat{\hat{p}}$ ein erwartungstreuer Schätzer für p?

(c) *[4 Punkte]* Folgende Realisationen und Kennzahlen sind gegeben:

i	1	2	3	4	5	Summe
y_i	2992	3497	4534	5699	7330	24052
x_i	510551	538701	572967	581037	733990	2937246
$\frac{y_i}{x_i}$	0,59%	0,65%	0,79%	0,98%	1%	4,01%

$\sum_{i=1}^{n} x_i y_i = 1,4701 \cdot 10^{10}$, $\sum_{i=1}^{n} x_i^2 = 1,755498 \cdot 10^{12}$

Welche Schätzwerte ergeben sich mit den Schätzern aus (a) und (b)?

Aufgabe 5. *[24 Punkte]* Es wird die Anzahl der positiven Testergebnisse Y und die Anzahl der durchgeführten Tests x untersucht. Für die Zufallsvariable Y wird für festes $x \in [5 \cdot 10^5,\ 1,5 \cdot 10^6]$ das Modell

$$Y = a + bx + \varepsilon \tag{1}$$

angenommen, wobei $\varepsilon \sim N(0, \sigma^2)$-verteilt ist.

Die zur Untersuchung vorliegenden Daten bestehen aus den Anzahlen x_i und den Realisierungen y_i der Zufallsvariablen $Y_i = a + bx_i + \varepsilon_i$ mit $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $i = 1, \ldots, n$. Die Größe der Stichprobe sei $n = 11$. Es werden die folgenden Kennzahlen berechnet:

$$\hat{b} = 1,001 \cdot 10^{-2}, \quad \hat{a} = -1,123 \cdot 10^3, \quad \text{se}(\hat{b}) = 1,884 \cdot 10^{-3}, \quad \text{se}(\hat{a}) = 1,648 \cdot 10^3,$$

In Abbildung 1 sind die Daten und die Anpassungsgerade eingezeichnet.

(a) *[8 Punkte]* Führen Sie die beiden Hypothesentests zu $a = 0$ und $b = 0$ zu einem Niveau von 5 % durch. Interpretieren Sie die Ergebnisse.

(b) *[3 Punkte]* Geben Sie ein 95 % Konfidenzintervall für b an.

(c) *[7 Punkte]* Gegeben ist das symmetrische 95% Prognoseintervall $[5232,\ 12542]$ für die Anzahl der positiven Tests.

  (i) Bestimmen Sie x_0 die Anzahl der Tests, zu dem dieses Prognoseintervall bestimmt wurde.

  (ii) Erläutern Sie die Bedeutung dieses Prognoseintervalls.

(d) *[6 Punkte]* Kommentieren Sie die Behauptungen:

  (i) Die erwartete Anzahl der positiven Tests steigt mit der Anzahl der durchgeführten Tests.

  (ii) Die Anzahl der positiven Tests steigt mit der Anzahl der durchgeführten Tests.

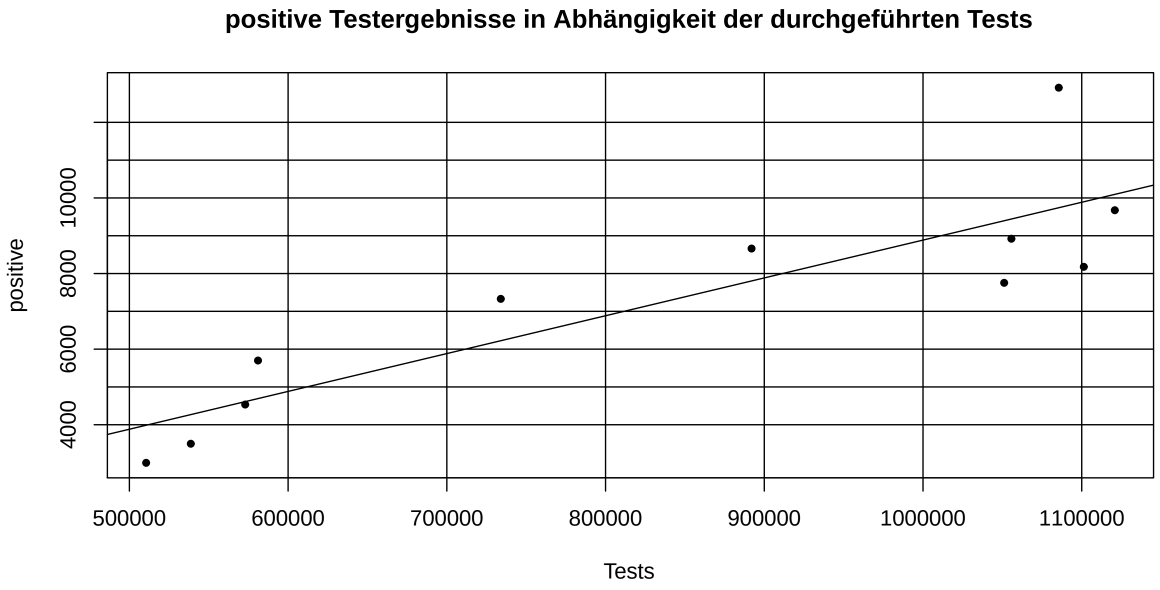

Abbildung 1: Anzahl der durchgeführten Test (auf der x-Achse) und positiven Tests (auf der y-Achse)

Lösungsvorschläge

Aufgabe 1 [24 Punkte]

Zu (a)

Für $n \in \mathbb{N}$ sei $g_n := \sum_{k=1}^{n} f_k$ und $h_n := \sum_{k=1}^{n} |f_k|$. Als Summe messbarer Funktionen sind g_n, h_n messbar. Somit sind auch g und h als punktweiser Grenzwerte messbarer Funktionen messbar.

Zu (b)

Sei h_n wie in (a). Offensichtlich gilt $0 \leq h_n \uparrow h := \sum_{k=1}^{\infty} |f_k|$ punktweise. Damit kann man den Satz von der monotonen Konvergenz anwenden, es folgt

$$\int_\Omega h\, d\mu = \lim_{n\to\infty} \int_\Omega h_n\, d\mu = \lim_{n\to\infty} \int_\Omega \sum_{k=1}^{n} |f_k|\, d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} \int_\Omega |f_k|\, d\mu = \sum_{k=1}^{\infty} \int_\Omega |f_k|\, d\mu.$$

Zu (c)

Die Funktion h aus (b) ist wegen (b) und nach Voraussetzung integrierbar. Für $n \in \mathbb{N}$ gilt mit den Bezeichnugen aus (a)

$$|g_n| \leq h \in L^1(\Omega).$$

Damit folgt mit dem Satz von der dominierten Konvergenz, dass g integrierbar ist und

$$\int_\Omega g\, d\mu = \lim_{n\to\infty} \int_\Omega g_n\, d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} \int_\Omega f_k\, d\mu = \sum_{k=1}^{\infty} \int_\Omega f_k\, d\mu.$$

Aufgabe 2 *[20 Punkte]* **Lösung**

Zu (a)

Es gilt $E(X|Y=y) = \frac{y}{2}$ und $\mathrm{Var}(X|Y=y) = \frac{y^2}{12}$ und somit

$$E(X|Y) = \frac{Y}{2}, \quad \mathrm{Var}(X|Y) = \frac{Y^2}{12}.$$

Mit dem Satz vom iterierten Erwartungswert gilt

$$\begin{aligned}
E(X) &= E(E(X|Y)) = E\left(\frac{Y}{2}\right) = \frac{1}{2}E(Y) = \frac{1}{2}\frac{2}{\lambda} = \frac{1}{\lambda} \\
\mathrm{Var}(X) &= \mathrm{Var}(E(X|Y)) + E(\mathrm{Var}(X|Y)) = \mathrm{Var}\left(\frac{Y}{2}\right) + E\left(\frac{Y^2}{12}\right) \\
&= \frac{1}{4}Var(Y) + \frac{1}{12}\left(\mathrm{Var}(Y) + E(Y)^2\right) = \frac{1}{4}\frac{2}{\lambda^2} + \frac{1}{12}\left(\frac{2}{\lambda^2} + \frac{4}{\lambda^2}\right) = \frac{1}{\lambda^2}.
\end{aligned}$$

Zu (b)

Es gilt für die bedingte Dichte von X gebeben Y

$$f(x|y) = \begin{cases} \frac{1}{y} 1_{[0,y]}(x) & \text{falls } y > 0, x \in [0,y] \\ 0 & \text{sonst.} \end{cases}$$

Für $x, y > 0$ gilt ferner

$$1_{[0,y]}(x) = 1_{[x,\infty)}(y).$$

Die Dichte von Y ist $f_Y(y) = \lambda^2 y e^{-\lambda y} 1_{[0,\infty)}(y)$. Somit ergibt sich für die gemeinsame Dichte für $x, y > 0$

$$f(x,y) = f(x|y) f_Y(y) = \frac{1}{y} 1_{[x,\infty)}(y) \lambda^2 y e^{-\lambda y} = 1_{[x,\infty)}(y) \lambda^2 e^{-\lambda y}.$$

Ist $x < 0$ oder $y < 0$ dann ist f offensichtlich 0.

Zu (c)

Die Dichte f_X von X ergibt sich durch Integration von f nach y. Für $x < 0$ ergibt sich 0. Für $x > 0$ folgt mit (b)

$$f_X(x) = \int_0^\infty 1_{[x,\infty)}(y) \lambda^2 e^{-\lambda y} \, dy = \int_x^\infty \lambda^2 e^{-\lambda y} \, dy = \lambda e^{-\lambda x}.$$

Somit ist $X \sim \text{Exp}(\lambda)$-verteilt (passt auch zum Ergebnis in (a)).

Aufgabe 3 *[28 Punkte]*

Zu (a)

Es gilt

$$\begin{aligned} P(T > 130) &= 1 - P(T \leq 130) = 1 - \Phi\left(\frac{130 - 100}{15}\right) = 1 - \Phi(2) = 1 - 0,9772 \\ &= 0,0228. \end{aligned}$$

Zu (b)

Sei $t \geq 130$. Nach der Definition der bedingten Wahrscheinlichkeit gilt

$$\begin{aligned} P(T \leq t | T > 130) &= \frac{P(130 < T \leq t)}{P(T > 130)} = \frac{P(t) - P(130)}{P(T > 130)} = \frac{\Phi\left(\frac{t-100}{15}\right) - \Phi\left(\frac{130-100}{15}\right)}{1 - \Phi(2)} \\ &= \frac{\Phi\left(\frac{t-100}{15}\right) - \Phi(2)}{1 - \Phi(2)}. \end{aligned}$$

Die Dichte erhält man, indem man nach t differenziert

$$\frac{d}{dt}P(T \leq t|T > 130) = \frac{\frac{1}{15}\varphi\left(\frac{t-100}{15}\right)}{1-\Phi(2)} = \frac{\varphi\left(\frac{t-100}{15}\right)}{15(1-\Phi(2))}$$

wobei φ die Dichte von $N(0,1)$ ist. Den Ewartungswert erhält man mittels Integration

$$\begin{aligned} E(T|T > 130) &= \frac{1}{15(1-\Phi(2))}\int_{130}^{\infty} t\varphi\left(\frac{t-100}{15}\right)dt \qquad (\text{Substitution } \tau = \tfrac{t-100}{15}) \\ &= \frac{1}{(1-\Phi(2))}\int_{2}^{\infty}(100+15\tau)\varphi(\tau)\,d\tau. \end{aligned}$$

Wegen

$$\begin{aligned} &\int_2^\infty \varphi(\tau)\,d\tau = 1-\Phi(2) \\ &\int_2^\infty \tau\varphi(\tau)\,d\tau = -\varphi(\tau)\Big|_2^\infty = \varphi(2) = \frac{1}{\sqrt{2\pi}}e^{-2} \end{aligned}$$

folgt

$$E(T|T > 130) = 100 + \frac{15}{[1-\Phi(2)]\sqrt{2\pi}}e^{-2} = 135,52.$$

Zu (c)

(i) Es gilt wegen der Linearität des Erwartungswerts und der identischen Verteilung der T_i

$$E(\overline{T}) = \frac{1}{n}\sum_{i=1}^{n}E(T_i) = \frac{1}{n}\sum_{i=1}^{n}100 = 100.$$

Ferner gilt

$$Var(\overline{T}) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n}T_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}Var(T_i) = \frac{1}{n^2}\sum_{i=1}^{n}15^2 = \frac{15^2}{n}.$$

Das zweite Gleichheitszeichen gilt wegen der Unabhängigkeit der T_i, das dritte wegen deren identischer Verteilung.

(ii) Es gilt $\overline{T} \sim N\left(100, \frac{15^2}{n}\right)$

$$\begin{aligned} P\left(\overline{T} \in [100-\delta, 100+\delta]\right) &= P\left(\overline{T} \leq 100+\delta\right) - P\left(\overline{T} \leq 100-\delta\right) \\ &= \Phi\left(\frac{\delta\sqrt{n}}{15}\right) - \Phi\left(-\frac{\delta\sqrt{n}}{15}\right) \\ &= \Phi\left(\frac{\delta\sqrt{n}}{15}\right) - \left(1-\Phi\left(\frac{\delta\sqrt{n}}{15}\right)\right) \\ &= 2\Phi\left(\frac{\delta\sqrt{n}}{15}\right) - 1. \end{aligned}$$

Löst man

$$P\left(\overline{T} \in [100-\delta, 100+\delta]\right) = 0,9 \text{ also } 2\Phi\left(\frac{\delta\sqrt{n}}{15}\right) - 1 = 0,9$$

nach δ auf dann gilt

$$u_{0,95} = \frac{\delta\sqrt{n}}{15} \iff \delta = \frac{15u_{0,95}}{\sqrt{n}} = \frac{24,6}{\sqrt{n}}$$

(iii) Mit (ii) gilt

$$\delta < 1 \iff \frac{15u_{0,95}}{\sqrt{n}} < 1 \iff n > 15^2 u_{0,95}^2 = 605,16.$$

Somit ergibt sich $n = 606$.

Aufgabe 4 *[24 Punkte]*

Zu (a)

Es gilt $Y_i \sim N(px_i, \sigma^2)$. Die Likelihood-Funktion (gemeinsame Dichte der $Y_1, \ldots, Y_n$) ist gegeben durch

$$L(p) = \prod_{i=1}^{n} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - px_i)^2\right)$$

Mit Hilfe der Loglikelihood folgt weiter

$$\ell(p) = \ln(L(p)) = -\frac{n\ln(2\pi)}{2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - px_i)^2$$

$$\frac{d}{dp}\ell(p) = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(y_i - px_i)$$

$$\frac{d^2}{dp^2}\ell(p) = -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 < 0.$$

Nullsetzen der ersten Ableitung und Auflösen ergibt

$$\sum_{i=1}^{n} x_i y_i = \hat{p}\sum_{i=1}^{n} x_i^2$$

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Somit ist

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$

der ML-Schätzer und wegen

$$E(\hat{p}) = \frac{\sum_{i=1}^{n} x_i E(Y_i)}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i p x_i}{\sum_{i=1}^{n} x_i^2} = p$$

erwartungstreu.

Zu (b)

Auch $\hat{\hat{p}}$ ist erwartungstreu, denn

$$E\left(\hat{\hat{p}}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{E(Y_i)}{x_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{px_i}{x_i} = p.$$

Zu (c)

Es ergeben sich

$$\hat{p} = \frac{4,01\%}{5} = 0,80\%,$$
$$\hat{\hat{p}} = \frac{1,4701 \cdot 10^{10}}{1,755498 \cdot 10^{12}} = 0,84\%.$$

Aufgabe 5*[24 Punkte]*

Zu (a)

Die Testgrößen sind

$$\frac{\hat{a}}{\mathrm{se}(\hat{a})} = \frac{-1123}{1648} = -0,68, \qquad \frac{\hat{b}}{\mathrm{se}(\hat{b})} = \frac{10,01}{1,84} = 5,31.$$

Die Testgrößen sind t_9 verteilt. Die Hypothese wird verworfen, wenn der Betrag der Testgröße größer als $t_{9;0.975} = 2,261$ ist. Damit wird $a = 0$ nicht verworfen, $b = 0$ wird verworfen. Damit kann angenommen werden, dass die Anzahl der positiven Tests von der Anzahl der durchgeführten Tests abhängt. Der Intercept a kann als 0 angenommen werden.

Zu (b) Es ergibt sich

$$[\hat{b} - t_{n-2;0,975}\mathrm{se}(\hat{b}), \hat{b} + t_{n-2;0,975}\mathrm{se}(\hat{b})] = [0,0058,\ 0,0143]$$

Zu (c)

(i) Das Prognoseintervall ist symmetrisch, die Intervallmitte $\hat{a} + \hat{b}x_0$ ergibt sich aus

$$\frac{1}{2}(5232 + 12542) = 8887 = \hat{a} + \hat{b}x_0 = -1123 + 0,01001x_0.$$

Löst man nach x_0 auf, ergibt sich $x_0 = 10^6$.

(ii) Für eine Anzahl von 10^6 durchgeführten Tests wird die Anzahl der posititiven Tests in 95 % der Fälle zwischen 5232 und 12542 liegen.

Zu (d)

(i) Diese Behauptung ist richtig, da die erwartete Anzahl positiver Tests gegeben ist durch

$$a + bx.$$

Mit $b > 0$ ist diese Größe monoton wachsend in x.

(ii) Diese Behauptung ist unklar fomuliert. Ist der Erwartunswert gemeint, dann kann man zustimmen (vergleiche (i)). Sind die Realisationen gemeint, so unterliegen diese einer gewissen Schwankung, die in der Behauptung nicht angeprochen wird.