Zulassungsprüfung Stochastik, 07.05.16

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 (12 Punkte)
Sei $X : \Omega \longrightarrow [0, \infty)$ eine Zufallsvariable. Zeigen Sie

$$\int_\Omega X \, dP \geq 0.$$

Hinweis: Beweisen Sie diese Ungleichung zuerst für Treppenfunktionen.

Aufgabe 2 (30 Punkte)

(a) Sei $k \in \mathbb{N}$, $X \sim \Gamma(\alpha, \lambda)$ mit $\alpha > k$. Beweisen Sie, dass

$$E\left(\frac{1}{X^k}\right) = \frac{\lambda^k}{(\alpha - 1) \dots (\alpha - k)}$$

gilt.

Für die Gamma-Funktion $\Gamma : (0, \infty) \to (0, \infty)$ *mit* $\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} \, dt$ *gilt die Identität* $\Gamma(x) = (x - 1) \dots (x - k)\Gamma(x - k)$, $x > k$.

(b) Gegeben seien $X_1, \dots, X_n \sim \mathrm{Exp}(\lambda)$ unabhängig mit $n \geq 3$ und unbekanntem $\lambda > 0$. Der Mittelwert der X_i wird mit

$$\overline{X} := \frac{1}{n} \sum_{i=1}^n X_i$$

bezeichnet. Zeigen Sie, dass $\hat{\lambda} = \frac{1}{\overline{X}}$ ein Maximum Likelihood Schätzer für λ ist.

(c) Seien $X_1, \dots, X_n$ und $\overline{X}$ wie in (b).

(i) Begründen Sie $\overline{X} \sim \Gamma(n, n\lambda)$.
Hinweis: Verwenden Sie die Eigenschaften der Gammaverteilung in der Formelsammlung!

(ii) Bestimmen Sie Erwartungswert und Varianz des Schätzers $\hat{\lambda}$ aus (b).
$\left[\frac{n\lambda}{n-1} \text{ und } \left(\frac{n}{n-1}\right)^2 \cdot \frac{\lambda^2}{n-2}\right]$

(iii) Geben Sie eine erwartungstreue Modifikation λ^* des Schätzers $\hat{\lambda}$ an und bestimmen Sie $\mathrm{Var}(\lambda^*)$.

(iv) Für welchen Schätzer $\hat{\lambda}$, λ^* würden Sie sich entscheiden? Begründen Sie Ihre Antwort mit dem mittleren quadratischen Fehler.
Für einen Schätzer T *von* λ *ist der mittlere quadratische Fehler gegeben durch* $mse(T) = E((T - \lambda)^2) = (E(T) - \lambda)^2 + \mathrm{Var}(T)$.

Aufgabe 3 (12 Punkte)
Seien X, Y Zufallsvariablen mit $E(X), \mathrm{Var}(X) \in \mathbb{R}$. Die Zufallsvariablen A, B seien $\sigma(Y)$ messbar, d.h. es gibt messbare Funktionen $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ mit $A = f(Y)$ und $B = g(Y)$. Beweisen Sie

$$\begin{aligned} E(AX + B|Y) &= AE(X|Y) + B, \\ \mathrm{Var}(AX + B|Y) &= A^2 \mathrm{Var}(X|Y). \end{aligned}$$

Aufgabe 4 (36 Punkte)
Max fährt täglich um 07:30 in die Hochschule, Vorlesungsbeginn ist 08:00. Die Anfahrtzeit T in Minuten sei normalverteilt.

(a) Es gelte $T \sim N(25, 16)$.

(i) Bestimmen Sie die Wahrscheinlichkeit, dass er zwischen 07:45 und 08:05 ankommt.

(ii) Wie groß ist die Wahrscheinlichkeit des Zuspätkommens?

(iii) An wievielen Vorlesungstagen im Jahr (150 Tage pro Jahr) wird er im Mittel zu spät kommen?

(iv) Welchen Abfahrtszeitpunkt (minutengenau) schlagen Sie vor, damit Max bei unveränderter Anfahrtszeit T im Mittel an höchstens 2 Tagen im Jahr zu spät kommt?

(b) Max will weiterhin erst um 07:30 starten und probiert eine neue Route. Die Anfahrtzeit sei weiter normalverteilt. An 10 Tagen kommt er im Mittel um 07:50 an, die empirische Varianz beträgt 12,25.

(i) Max vermutet, dass sich seine Anfahrtzeit verkürzt hat. Überprüfen Sie anhand einer geeigneten Nullhypothese seine Vermutung zum Niveau von 5 %.

(ii) Geben Sie für den Erwartungswert der Anfahrtzeit ein Konfidenzintervall zum Niveau 5 % an.

(c) Kommentieren Sie Normalverteilungsannahme.

Lösungsvorschläge

Aufgabe 1 [12]
Sei $X = \sum_{i=1}^{n} \alpha_i \, 1_{A_i}$, $\alpha_i \geq 0$, $A_i \in \mathcal{A}$, $i = 1, \ldots, n$ eine Treppenfunktion. Dann gilt

$$\int_\Omega X \, dP = \sum_{i=1}^{n} \alpha_i \, P(A_i)$$

und damit $\int_\Omega X \, dP \geq 0$.
Sei nun $X \geq 0$ beliebig. Dann gibt es eine monoton wachsende Folge $(X_n)_{n \in \mathbb{N}}$ von Treppenfunktionen mit $X_n \geq 0$ und $\lim_{n\to\infty} X_n = X$. Es gilt mit dem Satz von der monotonen Konvergenz

$$\int_\Omega X \, dP = \lim_{n\to\infty} \int_\Omega X_n \, dP,$$

und aus dem vorher Gezeigten folgt nun $\int_\Omega X \, dP \geq 0$.

Aufgabe 2 [**a) 5 b) 7 c) 3+5+5+5**]
Zu (a)
Es gilt

$$\begin{aligned}
E\left(\frac{1}{X^k}\right) &= \int_0^\infty x^{-k} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-k)}{\lambda^{\alpha-k}} \underbrace{\int_0^\infty \frac{\lambda^{\alpha-k}}{\Gamma(\alpha-k)} x^{\alpha-k-1} e^{-\lambda x} \, dx}_{=1} \\
&= \frac{\lambda^k \Gamma(\alpha-k)}{\Gamma(\alpha)} = \frac{\lambda^k}{(\alpha-1)\ldots(\alpha-k)}.
\end{aligned}$$

Die Gleichung in der zweiten Zeile folgt, da der Intergrand die Dichte der $\Gamma(\alpha-k, \lambda)$-Verteilung ist. In der letzten Gleichung wurde der Hinweis verwendet.

Zu (b)
Aufgrund der Unabhängigkeit der X_i, $i = 1, \ldots, n$ gilt für die Likelihood-Funktion

$$L(\lambda; x_1, \ldots, x_n) = \lambda e^{-\lambda x_1} \cdot \ldots \cdot \lambda e^{-\lambda x_n} = \lambda^n \, \exp\left(-\lambda \sum_{i=1}^{n} x_i\right), \; x_1, \ldots, x_n > 0.$$

Für die Loglikelihood-Funktion ℓ mit $\ell(\lambda) := \ln(L(\lambda))$ ergibt sich

$$\begin{aligned}
\ell(\lambda) &:= \ln L(\lambda) = \ln(\lambda^n) - \lambda \sum_{i=1}^{n} x_i = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \\
\ell'(\lambda) &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \\
\ell''(\lambda) &= -\frac{n}{\lambda^2} < 0
\end{aligned}$$

und somit liegt in $\frac{n}{\sum\limits_{i=1}^{n} x_i}$ ein Maximum der Likelihoodfunktion vor. Damit ist die Zufallsvariable $\hat{\lambda} = \frac{n}{\sum\limits_{i=1}^{n} X_i} = \frac{1}{\overline{X}}$ ein Maximum Likelihood Schätzer.

Zu (c)
(i) Es gilt $X_i \sim \Gamma(1, \lambda) = \mathrm{Exp}(\lambda)$. Da die X_i unabhängig sind, gilt $\sum_{k=1}^{n} X_i \sim \Gamma(n, \lambda)$ und somit $\frac{1}{n} \sum_{k=1}^{n} X_i \sim \Gamma(n, n\lambda)$.
(ii) Mit (a) und (i) gilt

$$\begin{aligned}
E(\hat{\lambda}) &= E\left(\frac{1}{\overline{X}}\right) = \frac{n\lambda}{n-1} \\
E(\hat{\lambda}^2) &= E\left(\frac{1}{\overline{X}^2}\right) = \frac{n^2\lambda^2}{(n-1)(n-2)} \\
\mathrm{Var}(\hat{\lambda}) &= \mathrm{Var}\left(\frac{1}{\overline{X}}\right) = \frac{n^2\lambda^2}{(n-1)(n-2)} - \frac{n^2\lambda^2}{(n-1)^2} \\
&= n^2\lambda^2 \cdot \frac{n-1-n+2}{(n-1)^2(n-2)} \\
&= \frac{n^2\lambda^2}{(n-1)^2(n-2)} = \left(\frac{n}{n-1}\right)^2 \cdot \frac{\lambda^2}{n-2}
\end{aligned}$$

(iii) Wegen (ii) ist $\hat{\lambda}$ nicht erwartungstreu. Dagegen ist offenbar

$$\lambda^* = \frac{n-1}{n}\hat{\lambda} = \frac{n-1}{n\overline{X}}$$

ein erwartungstreuer Schätzer von λ. Es gilt

$$\mathrm{Var}(\lambda^*) = \frac{(n-1)^2}{n^2}\mathrm{Var}(\hat{\lambda}) = \frac{(n-1)^2}{n^2} \cdot \frac{n^2\lambda^2}{(n-1)^2(n-2)} = \frac{\lambda^2}{n-2}.$$

(iv) Vergleicht man (ii) und (iii) ergibt sich $\mathrm{Var}(\hat{\lambda}) > \mathrm{Var}(\lambda^*)$. Da $\hat{\lambda}$ nicht erwartungstreu ist, während λ^* erwartungstreu ist, gilt

$$\begin{aligned}
\mathrm{mse}(\hat{\lambda}) &= \underbrace{(E(\hat{\lambda}) - \lambda)^2}_{>0} + \mathrm{Var}(\hat{\lambda}) > \mathrm{Var}(\hat{\lambda}) > \mathrm{Var}(\lambda^*) = (E(\lambda^*) - \lambda)^2 + \mathrm{Var}(\lambda^*) \\
&= \mathrm{mse}(\lambda^*).
\end{aligned}$$

λ^* weist einen kleineren mse auf als $\hat{\lambda}$, somit wird man λ^* den Vorzug geben.

Aufgabe 3 [12]
Es gilt

$$\begin{aligned}
E(AX + B|Y) &= E(AX|Y) + E(B|Y) = AE(X|Y) + BE(1|Y) \\
&= AE(X|Y) + B.
\end{aligned}$$

Die erste Gleichheit gilt wegen der Linearität der bedingten Erwartung, die zweite, da A, B jeweils $\sigma(Y)$-messbar sind.

Mit der gleichen Argumentation erhalten wir

$$
\begin{aligned}
&\mathrm{Var}(AX+B|Y)\\
&= E((AX+B)^2|Y)) - (E(AX+B|Y))^2\\
&= E(A^2X^2 + 2ABX + B^2|Y) - (AE(X|Y)+B)^2\\
&= E(A^2X^2|Y) + E(2ABX|Y) + E(B^2|Y) - \left(A^2E(X|Y)^2 + 2ABE(X|Y) + B^2\right)\\
&= A^2E(X^2|Y) + 2ABE(X|Y) + B^2 - A^2E(X|Y)^2 - 2ABE(X|Y) - B^2\\
&= A^2E(X^2|Y) - A^2E(X|Y)^2\\
&= A^2(E(X^2|Y) - E(X|Y)^2)\\
&= A^2\mathrm{Var}(X|Y).
\end{aligned}
$$

Aufgabe 4 [a) 5+3+5+9=22, b) 5+5=10, c) 4]
Zu (a)
(i) Ankunftszeit liegt zwischen 07:45 und 08:00 genau dann wenn $T \in [15, 30]$ gilt.

$$
\begin{aligned}
P(15 \le T \le 30) &= P(T \le 30) - P(T \le 15) = \Phi\left(\frac{30-25}{4}\right) - \Phi\left(\frac{15-25}{4}\right)\\
&= \Phi(1,25) - \Phi(-1,52) = 0,9938 - (1 - 0,9938) = 0,9876.
\end{aligned}
$$

(ii) Max verspätet sich genau dann wenn $T > 30$ gilt.

$$P(T > 30) = 1 - P(T \le 30) = 1 - \Phi(1,25) = 0,1056.$$

(iii) Bezeichnet man mit N die Anzahl der Tage, an denen sich Max verspätet, und setzt man $p := 0,1056$, dann gilt nach (ii) $N \sim B(150, p)$ und damit

$$E(N) = 150p = 15,84.$$

Im Mittel kommt Max an 16 Tagen im Jahr zu spät.
(iv) Die Wahrscheinlichkeit q, dass Max sich verspätet, soll so bestimmt werden, dass $150q \le 2$ gilt, also $q \le \frac{1}{75}$. Sei t_0 die Reisezeit, die mit einer Wahrscheinlichkeit von maximal $\frac{1}{75}$ überschritten wird, also

$$P(T > t_0) = \frac{1}{75}.$$

Wegen

$$P(T > t_0) = 1 - P(T \le t_0) = 1 - \Phi\left(\frac{t_0 - 25}{4}\right)$$

ergibt sich zunächst

$$\Phi\left(\frac{t_0 - 25}{4}\right) = \frac{74}{75} \approx 0,9867$$

und damit (näherungsweise)

$$t_0 = 25 + 4u_{0,\frac{74}{75}} \approx 25 + 4 \cdot 2,23 = 33,92 \approx 34.$$

Max sollte also um 07:26 starten.
Zu (b)
(i) Zu testen ist die Nullhypothese $H_0 : \mu \leq \mu_0$ mit $\mu_0 := 25$. Es handelt sich um einen einseitigen t-Test, Testgröße ist

$$T = \frac{\overline{x} - \mu_0}{s} \sqrt{n}$$

und besitzt bei den vorliegenden Daten den Wert

$$\frac{20 - 25}{3,5} \sqrt{10} = -4,52.$$

Die Hypothese wird im Fall $T > t_{9;0,95} = 1,833$ abgelehnt. Also wird H_0 nicht verworfen.
(ii) Das Schätzintervall lautet mit $t_{9;0,975} = 2,262$, $s = \sqrt{12,25} = 3,5$:

$$\left[\overline{x} - \frac{s \cdot t_{9;0,975}}{\sqrt{10}}, \overline{x} + \frac{s \cdot t_{9;0,975}}{\sqrt{10}} \right] = [17,5; 22,5] \, .$$

Zu (c)
Unter der Annahme der Normalverteilung gilt $P(T < 0) > 0$. Die Annahme der Normalverteilung kann daher nur näherungsweise erfüllt sein.