Zulassungsprüfung Stochastik, 14.10.11

Wir gehen stets von einem Maßraum $(\Omega, \mathcal{A}, \mu)$ bzw. einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ 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.

Sollte Ihnen bei Teilaufgaben ein Ergebnis fehlen, dann treffen Sie eine plausible Annahme dafür.

Aufgabe 1 (8 Punkte)

(a) Gegeben sei eine monoton fallende Folge $(f_n)_{n\in\mathbb{N}}$ messbarer Funktionen $f_n : \Omega \longrightarrow [0, \infty)$ mit Grenzwert $f := \lim_{n\to\infty} f_n$. Ferner Sei f_1 integrierbar. Beweisen Sie: f ist integrierbar und es gilt

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

(b) Wurde aus der Wertung genommen, da die Aufgabe wegen eines Schreibfehlers nicht lösbar war.

Lösung
Zu (a)
Wegen der Monotonie konvergiert die Funktionenfolge $(f_n)_{n\in\mathbb{N}}$ punktweise. Ferner folgt wegen $f_n \geq 0$ und der Monotonie für alle $n \in \mathbb{N}$

$$0 \leq f_n \leq f_1$$

wobei f_1 integrierbar ist. Damit folgt mit dem Satz von der dominierten Konvergenz die Integrierbarkeit von f und die Gleichung

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Aufgabe 2 (18 Punkte) Seien X, Y Zufallsvariablen mit endlichem Erwartungswert und endlicher Varianz mit $\mathrm{Var}(X) > 0$. Beweisen Sie:

(a) $\mathrm{Cov}(X, Y|X) = 0$

(b) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, \mathbb{E}(Y|X))$

(c) $\mathbb{E}(XY) = \mathbb{E}(X\mathbb{E}(Y|X))$

(d) Gibt es reelle Zahlen $a, b \in \mathbb{R}$ mit $\mathbb{E}(Y|X) = aX + b$ fast sicher, dann folgt $a = \dfrac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$.

Lösung
Zu (a)
Die Zufallsvariable X ist klarerweise $\sigma(X)$-messbar, also gilt $\mathbb{E}(XY|X) = X\mathbb{E}(Y|X)$, $\mathbb{E}(X|X) = X$ und

$$\mathrm{Cov}(X, Y|X) = \mathbb{E}(XY|X) - \mathbb{E}(X|X)\mathbb{E}(Y|X) = X\mathbb{E}(Y|X) - X\mathbb{E}(Y|X) = 0.$$

Zu (b)
Durch Bedingen nach X ergibt sich mit (a)

$$\operatorname{Cov}(X,Y) = \operatorname{Cov}\left(\underbrace{\mathbb{E}(X|X)}_{=X}, \mathbb{E}(Y|X)\right) + \mathbb{E}\left(\underbrace{\operatorname{Cov}(X,Y|X)}_{=0}\right) = \operatorname{Cov}(X, \mathbb{E}(Y|X))$$

Zu (c)
Laut (b) gilt

$$\begin{aligned}\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) &= \operatorname{Cov}(X,Y) \\ &= \operatorname{Cov}(X, \mathbb{E}(Y|X)) \\ &= \mathbb{E}(X\mathbb{E}(Y|X)) - \mathbb{E}(X)\mathbb{E}(\mathbb{E}(Y|X)) \\ &= \mathbb{E}(X\mathbb{E}(Y|X)) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

Zu (d)
Mit (b) gilt

$$\operatorname{Cov}(X,Y) = \operatorname{Cov}(X, \mathbb{E}(Y|X)) = \operatorname{Cov}(X, aX+b) = a\operatorname{Cov}(X,X) = a\operatorname{Var}(X).$$

Wegen $\operatorname{Var}(X) > 0$ folgt das Gewünschte.

Aufgabe 3 (18 Punkte)
Die jährliche Niederschlagsmenge in Entendorf sei näherungsweise normalverteilt mit Erwartungswert 1020 mm und einer Standardabweichung von 221 mm.

(a) Wie hoch ist die Wahrscheinlichkeit, dass die Niederschlagsmenge in einem Jahr über 1067mm beträgt?

(b) Wie hoch ist die Wahrscheinlichkeit, dass die gesamte Niederschlagsmenge in zwei aufeinanderfolgenden Jahren mehr als 2134 mm beträgt (also mehr als doppelt so hoch wie in (a))

(c) Welche Annahmen über die Unabhängigkeit treffen Sie in (b)?

(d) Bestimmen Sie unter geeigneten Voraussetzungen das kleinste n so, dass die Wahrscheinlichkeit, dass die gesamte Niederschlagsmenge in n aufeinanderfolgenden Jahren mehr als $n \cdot 1067$ mm überschreitet, maximal 5 % beträgt.

(e) Wie hoch ist die Wahrscheinlichkeit, dass die Niederschlagsmenge in zwei aufeinanderfolgenden Jahren jeweils mehr als 1067 mm beträgt. Vergleichen Sie das Ergebnis mit (b) und erklären Sie die von Ihnen gemachte Beobachtung.

Lösung
Zu (a)
Sei X die Niederschlagsmenge in einem Jahr. Zu bestimmen ist $\mathbb{P}(X > 1067)$:

$$\begin{aligned}\mathbb{P}(X > 1067) &= 1 - \mathbb{P}(X \leq 1067) = 1 - \Phi\left(\frac{1067-1020}{221}\right) \approx 1 - \Phi(0,21) \\ &\approx 0,4168.\end{aligned}$$

Zu (b)
Seien X_1, X_2 die Niederschlagsmengen zweier aufeinanderfolgender Jahre. Dann gilt $X_1 + X_2 \sim N(2040, 2 \cdot 221^2)$. Damit gilt

$$\mathbb{P}(X_1 + X_2 > 2134) \quad = \quad 1 - \Phi\left(\frac{94}{221\sqrt{2}}\right) \approx 1 - \Phi(0,3) = 0,3821.$$

Zu (c)
Es wird die Unabhängigkeit von X_1, X_2 vorausgesetzt.
Zu (d)
Seien $X_1, \ldots, X_n$ die Niederschlagsmengen von n aufeinanderfolgenden Jahre, diese seien unabhängig. Dann gilt $\sum_{i=1}^{n} X_i \sim N(1020n, 221^2 n)$. Gesucht ist als n mit

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > n \cdot 1067\right) \quad \leq \quad 0,05.$$

Es gilt

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > n \cdot 1067\right) \quad = \quad 1 - \Phi\left(\frac{1067n - 1020n}{221\sqrt{n}}\right) = 1 - \Phi\left(\frac{221}{47}\sqrt{n}\right).$$

Auflösen der obigen Ungleichung nach n ergibt dann

$$\sqrt{n} \geq u_{0.95}\frac{47}{221} \approx 7,71 \Longrightarrow n = 60.$$

Zu (e)
Mit den Bezeichnungen von (b) gilt

$$\mathbb{P}(X_1 > 1067, X_2 > 1067) = \mathbb{P}(X_1 > 1067)\mathbb{P}(X_2 > 1067) \approx 0,4368^2 \approx 0,19.$$

Diese Wahrscheinlichkeit muss kleiner sein als die Wahrscheinlichkeit in (b) wegen

$$\{X_1 > 1067, X_2 > 1067\} \subset \{X_1 + X_2 > 2134\}.$$

Aufgabe 4 (18 Punkte)
Sei $X \sim \Gamma(2, 1/\vartheta)$ mit $\vartheta > 0$.

(a) Sei $X_1, \ldots X_n \sim X$ eine Stichprobe von unabhängig und identisch verteilten Zufallsvariablen. Bestimmen Sie einen Maximum Likelihood Schätzer für ϑ.

(b) Ist der Schätzer aus (a) erwartungstreu?

(c) Bestimmen den m.s.e (**m**ean **s**quared **e**rror, d.h. den mittleren quadratischen Fehler) des Schätzers aus (a).

(d) Gegeben sei eine Stichprobe mit Stichprobenlänge 15 und Stichprobenmittel 1,6. Bestimmen Sie den Maximum Likelihood Schätzwert für ϑ.

Lösung
Zu (a)
Aus der gemeinsamen Dichte der Zufallsvariablen ergibt sich

$$\begin{aligned} L(\vartheta) &= \prod_{i=1}^{n}\left(\frac{1}{\vartheta^2}x_i e^{-x_i/\vartheta}\right) = \vartheta^{-2n} e^{-1/\vartheta \sum\limits_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i \\ \ell(\vartheta) &= -2n\ln\vartheta - \frac{1}{\vartheta}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i \\ \ell'(\vartheta) &= -2n\frac{1}{\vartheta} + \frac{1}{\vartheta^2}\sum_{i=1}^{n} x_i \\ \ell''(\vartheta) &= 2n\frac{1}{\vartheta^2} - \frac{2}{\vartheta^3}\sum_{i=1}^{n} x_i. \end{aligned}$$

Nullsetzen der erste Ableitung und Auflösen nach $\hat{\vartheta}$ führt zur einzigen Nullstelle von ℓ'

$$\hat{\vartheta} = \frac{1}{2n}\sum_{i=1}^{n} x_i = \frac{\overline{x}}{2} \text{ mit } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

In $\hat{\vartheta}$ liegt ein Maximum von ℓ vor, denn

$$\ell''(\hat{\vartheta}) = \frac{8n}{\overline{x}^2} - \frac{16}{\overline{x}^3}\cdot n\overline{x} = -\frac{8n}{\overline{x}^2} < 0.$$

Damit ist $T_n(X_1,\ldots,X_n) = \frac{2}{n}\sum\limits_{i=1}^{n} X_i$ der Maximum Likelihood Schätzer.
Zu (b)

$$\mathbb{E}\left(T_n(X_1,\ldots,X_n)\right) = \mathbb{E}\left(\frac{1}{2n}\sum_{i=1}^{n} X_i\right) = \frac{1}{2n}\sum_{i=1}^{n}\underbrace{\mathbb{E}\left(X_i\right)}_{=2\vartheta} = \vartheta.$$

Ja, der Schätzer ist erwartungstreu.
Zu (c)
Da der Schätzer erwartungstreu ist, gilt

$$\text{m.s.e}(T_n) = \text{Var}(T_n).$$

Wegen der Unabhängigkeitsannahme folgt weiter

$$\text{Var}\left(T_n(X_1,\ldots,X_n)\right) = \text{Var}\left(\frac{1}{2n}\sum_{i=1}^{n} X_i\right) = \frac{1}{4n^2}\sum_{i=1}^{n}\underbrace{\text{Var}\left(X_i\right)}_{=2\vartheta^2} = \frac{\vartheta^2}{2n}.$$

Zu (d) $\hat{\vartheta} = {}^{1{,}6}/_{2} = 0{,}8.$

Aufgabe 5 (18 Punkte)
Seien $X_1,\ldots,X_n,X_{n+1} \sim N(\mu,\sigma^2)$ unabhängige Zufallsvariablen. Ziel ist es aus den Beobachtungen $X_1,\ldots,X_n$ ein Prognoseintervall für X_{n+1} zu konstruieren, wobei μ als unbekannt und σ^2 als bekannt vorausgesetzt wird. Sei $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.

(a) Bestimmen Sie die Verteilung von $\overline{X}$ und $\overline{X} - X_{n+1}$.

(b) Beweisen Sie, dass $\dfrac{\overline{X} - X_{n+1}}{\sigma\sqrt{1 + {}^1\!/\!_n}}$ standardnormaverteilt ist.

(c) Bestimmen Sie ein Prognoseintervall für X_{n+1}, das symmetrisch um $\overline{X}$ ist zum Niveau $1 - \alpha$ für $\alpha \in (0, 1)$.

(d) Sie können davon ausgehen, dass die mittlere Temperatur des November normalverteilt ist und die mittleren Temperaturen des Monats November jeweils unabhängig voneinander und identisch verteilt sind. Die Standardabweichung wird mit $0,4°$ C als bekannt vorausgesetzt.

Temperaturmessungen ergaben in den Jahren 1981-2010 eine mittlere Temperatur von $8,9°$ C für den Monat November. Bestimmen Sie hierfür das Prognoseintervall aus (c) zu einem Konfidenzniveau von 90 % für die mittlere Temperatur im November 2011. (Genauigkeit: in ° C, eine Nachkkommastelle.)

Lösung
Zu (a)
$\overline{X}$ und $\overline{X} - X_{n+1}$ sind Linearkombinationen von unabhängigen, normalverteilten Zufallsvariablen und somit auch normalverteilt. Es gilt

$$\begin{aligned}
\mathbb{E}\left(\overline{X}\right) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \mu \\
\mathrm{Var}\left(\overline{X}\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n} \quad \text{da } (X_i)_{i=1,\ldots,n} \text{ unabhängig} \\
\mathbb{E}\left(\overline{X} - X_{n+1}\right) &= \mathbb{E}\left(\overline{X}\right) - \mathbb{E}\left(X_{n+1}\right) = \mu - \mu = 0 \\
\mathrm{Var}\left(\overline{X} - X_{n+1}\right) &= \mathrm{Var}\left(\overline{X}\right) + \mathrm{Var}\left(-X_{n+1}\right) = \frac{\sigma^2}{n} + \sigma^2 \\
&= \sigma^2\left(1 + \frac{1}{n}\right).
\end{aligned}$$

Zu (b)
Die Zufallsvariable

$$\frac{\overline{X} - X_{n+1}}{\sigma\sqrt{1 + {}^1\!/\!_n}}$$

ist eine Linearkombination normalverteilter Zuvallsvariablen. Wegen (a) ist der Erwartungswert 0 und die Varianz 1.
Zu (c)
Gesucht ist $c > 0$ so, dass

$$\mathbb{P}\left(\overline{X} - c \leq X_{n+1} \leq \overline{X} + c\right) = 1 - \alpha$$

gilt. Für die linke Seite erhält man

$$\begin{aligned}
\mathbb{P}\left(\overline{X} - c \leq X_{n+1} \leq \overline{X} + c\right) &= \mathbb{P}\left(-c \leq X_{n+1} - \overline{X} \leq c\right) \\
&= \mathbb{P}\left(-\frac{c}{\sigma\sqrt{1+1\frac{1}{n}}} \leq \frac{X_{n+1} - \overline{X}}{\sigma\sqrt{1+1\frac{1}{n}}} \leq \frac{c}{\sigma\sqrt{1+1\frac{1}{n}}}\right) \\
&= 2\Phi\left(\frac{c}{\sigma\sqrt{1+1\frac{1}{n}}}\right) - 1.
\end{aligned}$$

Löst man die sich ergebende Gleichung nach c auf ergibt sich

$$c = u_{1-\alpha/2}\sigma\sqrt{1+\frac{1}{n}}.$$

Zu (d)
Mit $\alpha = 0,1$, $n = 30$ ergibt sich $c = 1,64 \cdot 0,4\sqrt{1 + 1/30} \approx 0,7°$ C, also das Prognoseintervall $[8.2, 9.6]$.