Zulassungsprüfung Stochastik, 12.10.12

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.

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

Aufgabe 1 (18 Punkte)

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

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

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

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

Aufgabe 2 (18 Punkte)

Die Zuckersüß AG füllt Zucker 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 „relativer 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 eine geeignete Approximation.

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

Aufgabe 3 (18 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 $\mathbb{E}\left(\frac{1}{\Theta}\right)$ und $\mathbb{E}\left(\frac{1}{\Theta^2}\right)$ $\left[\text{Ergebnis: } \frac{1}{b-a} \ln\left(\frac{b}{a}\right),\ \frac{1}{ab}\right]$

(b) Bestimmen Sie $\mathbb{E}(N|\Theta)$ und $\mathrm{Var}(N|\Theta)$.

(c) Bestimmen Sie $\mathbb{E}(N)$ und $\mathrm{Var}(N)$.

(d) Bestimmen Sie $\mathbb{P}(N = 0)$.

Aufgabe 4 (18 Punkte)
Seien $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ und $W_1, \ldots, W_n$ unabhängige Zufallsvariablen mit

$$X_i \overset{\text{iid}}{\sim} N(\mu_1, 1),\ Y_i \overset{\text{iid}}{\sim} N(\mu_2, 1),\ W_i \overset{\text{iid}}{\sim} N(\mu_1 + \mu_2, 1)$$

mit $\mu_1, \mu_2 \in \mathbb{R}$ und Realisierungen $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ und $w_1, \ldots, w_n$.

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

$$\begin{aligned} 2n\hat{\mu}_1 + n\hat{\mu}_2 &= \sum_{i=1}^n x_i + \sum_{i=1}^n w_i \\ n\hat{\mu}_1 + 2n\hat{\mu}_2 &= \sum_{i=1}^n y_i + \sum_{i=1}^n w_i \end{aligned}$$

erfüllen.

(b) Bestimmen Sie die ML-Schätzer für μ_1 und μ_2. Sie können ohne Beweis ihre Existenz voraussetzen.

Aufgabe 5 (18 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) Bisher ging man bei Nichtrauchern von einem Blutdruck von 120 aus. Wird diese Beobachtung bei einem Niveau von $0,05$ widerlegt?

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

Lösungsvorschläge

Aufgabe 1
Zu (a)
Für $n \in \mathbb{N}$ sei $f_n : [1,\infty) \longrightarrow \mathbb{R}$, $f_n(x) := \sum\limits_{k=1}^{n} ke^{-kx}$. 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)
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)
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

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

Aufgabe 2
Zu (a)

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

Zu (b)

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

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}\left(\frac{N}{n}\right) = \frac{1}{n}\mathbb{E}(N) = \frac{1}{n}np = p = 0,0968 \\
\mathrm{Var}(Y) &= \mathrm{Var}\left(\frac{N}{n}\right) = \frac{1}{n^2}\mathrm{Var}(N) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n} = \frac{0,08743}{10000} \\
&= 0,8743 \cdot 10^{-6}.
\end{aligned}$$

Zu (c)
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

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

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

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

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

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

Aufgabe 3
Zu (a) Wegen $\Theta \sim U[a,b]$ folgt

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

Zu (b)

$$\begin{aligned}
\mathbb{E}(N|\Theta = \theta) &= \frac{1-\theta}{\theta} = \frac{1}{\theta} - 1 \Longrightarrow \mathbb{E}(N|\Theta) = \frac{1}{\Theta} - 1 \\
\mathrm{Var}(N|\Theta = \theta) &= \frac{1-\theta}{\theta^2} \Longrightarrow \mathrm{Var}(N|\Theta) = \frac{1-\Theta}{\Theta^2} = \frac{1}{\Theta^2} - \frac{1}{\Theta}.
\end{aligned}$$

Zu (c)

Mit dem Satz von der iterierten Erwartung folgt

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

Damit folgt mit (a)

$$\begin{aligned}
\mathbb{E}(N) &= \frac{1}{b-a} \ln\left(\frac{b}{a}\right) - 1 \\
\mathrm{Var}(N) &= \frac{2}{ab} - \frac{1}{(b-a)^2} \ln^2\left(\frac{b}{a}\right) - \frac{1}{b-a} \ln\left(\frac{b}{a}\right).
\end{aligned}$$

Zu (d)

$$\mathbb{P}(N=0) = \frac{1}{b-a} \int_a^b \mathbb{P}(N=0|\Theta=\theta)\, d\theta = \frac{1}{b-a} \int_a^b \theta\, d\theta = \frac{b+a}{2}.$$

Aufgabe 4
Zu (a)
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

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

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)
Mit $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$ erfüllen die ML Schätzwerte $\hat{\mu}_1$, $\hat{\mu}_2$ das lineare Gleichungssystem

$$2\hat{\mu}_1 + \hat{\mu}_2 = \overline{x} + \overline{w} \quad \text{(I)}$$
$$\hat{\mu}_1 + 2\hat{\mu}_2 = \overline{y} + \overline{w}. \quad \text{(II)}$$

Auflösen ergibt

$$\begin{aligned}
-3\hat{\mu}_2 &= \overline{x} - \overline{w} - 2\overline{y} && \text{(I $-2\cdot$ II)}\\
\hat{\mu}_2 &= -\frac{1}{3}\overline{x} + \frac{1}{3}\overline{w} + \frac{2}{3}\overline{y}\\
\hat{\mu}_1 &= \frac{2}{3}\overline{x} - \frac{2}{3}\overline{w} - \frac{4}{3}\overline{y} + \overline{y} + \overline{w} = -\frac{1}{3}\overline{y} + \frac{1}{3}\overline{w} + \frac{2}{3}\overline{x}. && \text{(in II eingesetzt)}
\end{aligned}$$

Damit ergeben sich die ML-Schätzer

$$\begin{aligned}
\hat{\mu}_1 &= -\frac{1}{3}\overline{Y} + \frac{1}{3}\overline{W} + \frac{2}{3}\overline{X}\\
\hat{\mu}_2 &= -\frac{1}{3}\overline{X} + \frac{1}{3}\overline{W} + \frac{2}{3}\overline{Y}
\end{aligned}$$

mit $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, $\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i$.

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.
(a) 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

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

(b) Beim t-Test ergibt sich die Testgröße $t = \dfrac{123-120}{5,7}\sqrt{14} = 1,96$. Die Hypothese wird nicht verworfen wegen $t_{13;0,975} = 2.160$ und $|t| < t_{13;0,975}$.
(c) Bei unbekannter Varianz und zwei unabhängigen Stichproben wird der t-Test verwendet. Zu prüfen ist

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

Testgröße ist

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

Wegen $T(x,y) \notin [t_{23;1-\alpha/2}, t_{23;1-\alpha/2}] = [-2,069; 2,069]$ wird H_0 verworfen.