Eingangsprüfung Stochastik, 15.05.10

Wir gehen stets von 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.

Aufgabe 1 (11 Punkte)
Sei $x \in \mathbb{R}$ fest. Gegeben sei das Maß $\mu : \mathcal{B}^1 \longrightarrow \{0,1\}$, $\mu(A) = 1_A(x)$. Beweisen Sie, dass für alle Borel-messbaren Funktionen $f : \mathbb{R} \longrightarrow \mathbb{R}$

$$\int_{\mathbb{R}} f d\mu = f(x)$$

gilt.
Lösung
Ist $f \geq 0$ eine Treppenfunktion, d.h. es gibt $\alpha_i \geq 0$, $A_i \in \mathcal{B}^1$, $i = 1, \ldots, n$ mit $f = \sum_{i=1}^{n} \alpha_i 1_{A_i}$, dann gilt

$$\int_{\mathbb{R}} f\, d\mu = \sum_{i=1}^{n} \alpha_i \underbrace{\mu(A_i)}_{=1_{A_i}(x)} = \sum_{i=1}^{n} \alpha_i 1_{A_i}(x) = f(x).$$

Ist nun $f \geq 0$, dann sei $(f_n)_{n\in\mathbb{N}}$ eine Folge von Treppenfunktionen mit $f_n \uparrow f$, $f_n \geq 0, n \in \mathbb{N}$. Mit dem Satz von der Monotonen Konvergenz folgt

$$\int_{\mathbb{R}} f\, d\mu = \lim_{n\to\infty} \underbrace{\int_{\mathbb{R}} f_n\, d\mu}_{=f_n(x)} = \lim_{n\to\infty} f_n(x) = f(x).$$

Im allgemeinen Fall sei $f_{\pm} := \max(0, \pm f) \geq 0$. f_+ und f_- sind messbar, somit gilt

$$\int_{\mathbb{R}} f_+\, d\mu = f_+(x) \in \mathbb{R} \text{ und } \int_{\mathbb{R}} f_-\, d\mu = f_-(x) \in \mathbb{R}.$$

Daraus folgt wegen $|f| = f_+ + f_-$, dass f integrierbar ist und aus $f = f_+ - f_-$ schließlich

$$\int_{\mathbb{R}} f\, d\mu = \int_{\mathbb{R}} f_+\, d\mu - \int_{\mathbb{R}} f_-\, d\mu = f_+(x) - f_-(x) = f(x).$$

Aufgabe 2 (25 Punkte) Sei $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable mit Dichte

$$f(x) = \begin{cases} p & \text{falls } x \in (0,1] \\ q & \text{falls } x \in (1,2] \\ 0 & \text{sonst} \end{cases}.$$

Wir betrachten ferner Zufallsvariablen $X_1, \ldots, X_n$ unabhängig und identisch wie X verteilt.

(a) Welche Bedingung müssen p und q erfüllen, damit f eine Wahrscheinlichkeitsdichte ist?

(b) Bestimmen Sie die Verteilungsfunktion, den Erwartungswert und die Varianz von X in Abhängigkeit von p.

(c) Sei

$$\hat{P} = \frac{3}{2} - \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Beweisen Sie, dass $\hat{P}$ ein erwartungstreuer Schätzer von p ist, und bestimmen Sie den quadratischen Fehler $\operatorname{mse}\big(\hat{P}\big)$ von $\hat{P}$, d.h.

$$\operatorname{mse}\big(\hat{P}\big) = \mathbb{E}((\hat{P} - p)^2) = \operatorname{Var}(\hat{P}) + (\mathbb{E}(\hat{P}) - p)^2.$$

(d) Für $i = 1, \ldots, n$ sei $K_i : \Omega \longrightarrow \{0, 1\}$, $K_i(\omega) = \begin{cases} 1 & \text{falls } X_i(\omega) \leq 1 \\ 0 & \text{sonst} \end{cases}$ und $K := \sum_{i=1}^{n} K_i$.

(i) Bestimmen Sie die Verteilung von K_i und von K in Abhängigkeit von p.

(ii) Zeigen Sie, dass $\hat{\hat{P}} := \frac{K}{n}$ ein erwartungstreuer Schätzer von p ist, und bestimmen Sie $\operatorname{mse}\Big(\hat{\hat{P}}\Big)$.

(e) Vergleichen Sie $\hat{\hat{P}}$ und $\hat{P}$.

Lösung

Zu (a) Es muss gelten

$$1 = \int_0^1 p\,dx + \int_1^2 q\,dx = p + q \Longrightarrow q = 1 - p$$

und $p, q \geq 0$.

Zu (b)

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^1 px\,dx + \int_1^2 qx\,dx = \frac{p}{2} + (1-p)\frac{x^2}{2}\bigg|_1^2 = \frac{p}{2} + (1-p)\frac{3}{2} \\
&= \frac{3}{2} - p \\
\mathbb{E}(X^2) &= \int_0^1 px^2\,dx + \int_1^2 qx^2\,dx = \frac{p}{3} + (1-p)\frac{x^3}{3}\bigg|_1^2 \\
&= \frac{p}{3} + (1-p)\frac{7}{3} = \frac{7}{3} - 2p \\
\operatorname{Var}(X) &= \frac{7}{3} - 2p - \left(\frac{3}{2} - p\right)^2 = \frac{7}{3} - 2p - \frac{9}{4} + 3p - p^2 \\
&= \frac{1}{12} + p - p^2 = \frac{1}{12} + p(1-p).
\end{aligned}$$

Sei F die Verteilungsfunktion von X. Es gilt $F(x) = 0$ für $x \leq 0$,

$$\begin{aligned}
x \in (0,1] &\implies F(x) = \int_0^x p dt = px \\
x \in (1,2] &\implies F(x) = \int_0^1 p dt + \int_1^x (1-p) dt = p + (1-p)(x-1) \\
x \in [2,\infty) &\implies F(x) = 1.
\end{aligned}$$

Zu (c)

$$\begin{aligned}
\mathbb{E}\left(\hat{P}\right) &= \frac{3}{2} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X) = \frac{3}{2} - \mathbb{E}(X) = p \\
\operatorname{Var}\left(\hat{P}\right) &= \operatorname{Var}\left(\frac{3}{2} - \frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) \\
&= \frac{1}{n}\operatorname{Var}(X) = \frac{1}{12n} + \frac{p(1-p)}{n}.
\end{aligned}$$

Da $\hat{P}$ erwartungstreu ist, gilt $\operatorname{mse}(\hat{P}) = \operatorname{Var}(\hat{P})$.
Zu (d)
(i)

$$\mathbb{P}(K_i = 1) = \mathbb{P}(X_i \leq 1) = \mathbb{P}(X \leq 1) = p \text{ für alle } i = 1, \ldots, n.$$

Damit ist K_i Bernoulli-verteilt mit Parameter p. Wegen $K_i = 1_{(-\infty,1]}(X_i)$ und der Unabhängigkeit der X_i sind die Zufallsvariablen K_i ebenfalls unabhängig. Damit gilt $K \sim B(n,p)$.
(ii) Es folgt

$$\begin{aligned}
\mathbb{E}(\hat{\hat{P}}) &= \frac{1}{n}\mathbb{E}(K) = \frac{1}{n} \cdot np = p \\
\operatorname{Var}(\hat{\hat{P}}) &= \frac{1}{n^2}\operatorname{Var}(K) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}
\end{aligned}$$

Da $\hat{\hat{P}}$ erwartungtreu ist, gilt $\operatorname{mse}(\hat{\hat{P}}) = \operatorname{Var}(\hat{\hat{P}})$.
Zu (e)
Beide Schätzer sind erwartungstreu. Wegen

$$\operatorname{mse}(\hat{P}) - \operatorname{mse}(\hat{\hat{P}}) = \frac{1}{12n} + \frac{p(1-p)}{n} - \frac{p(1-p)}{n} = \frac{1}{12n} > 0$$

ist der Schätzer $\hat{\hat{P}}$ vorzuziehen, da er den kleineren Standardfehler aufweist.

Aufgabe 3 (16 Punkte)
Wir setzen als bekannt voraus, dass für identisch und unabhängig verteilte Zufallsvariablen X_i, $i = 1, \ldots, n$, $X_i \sim \operatorname{Exp}\left(\frac{1}{2}\right)$

$$\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

gilt.

Gegeben seien $X_1, \dots, X_n$ unabhängige Exp(λ)-verteilte Zufallsvariablen, mit unbekanntem $\lambda > 0$. Für den Erwartungswert $\mu := \mathbb{E}(X)$ soll die Vermutung $\mu = 2$ untersucht werden. Es wird der Schätzer $\overline{X} := \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ verwendet.

(a) Angenommen es gilt $\lambda = \dfrac{1}{2}$. Bestimmen Sie für $\alpha \in (0,1)$ Zahlen a, b mit der Eigenschaft

$$\mathbb{P}(\overline{X} \leq a) = \frac{\alpha}{2}, \quad \mathbb{P}(\overline{X} \leq b) = 1 - \frac{\alpha}{2}$$

(b) Wie beurteilen Sie für die Daten

1,3 6,1 0,7 3,1 2,7 0,5 0,4 0,4 2,4 0,6

die Hypothese $H_0 : \mu = 2$ zum Niveau 5 %?

Lösung
Zu (a)
Für $z > 0$ gilt

$$\mathbb{P}(\overline{X} \leq z) \quad = \quad \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq z\right) = \mathbb{P}\left(\sum_{i=1}^{n} X_i \leq nz\right) = \chi^2_{2n}(nz)$$

Somit sind a und b so zu bestimmen, dass $na = \chi^2_{2n;\alpha/2}$ und $nb = \chi^2_{2n;1-\alpha/2}$ gilt, also ergibt sich

$$a = \frac{\chi^2_{2n;\alpha/2}}{n}, \quad b = \frac{\chi^2_{2n;1-\alpha/2}}{n}.$$

Zu (b)
$H_0 : \mu = 2$ wird zum Niveau α abgelehnt, wenn unter H_0 der Mittelwert der Stichprobe zu klein bzw. zu groß ist, bzw. präziser, wenn $\overline{X}$ stark von 2 abweicht, d.h. wenn

$$\mathbb{P}\left(\overline{X} \leq a\right) + \mathbb{P}(\overline{X} \geq b) = \alpha.$$

Wählt man a, b wie in (a) ist diese Gleichung erfüllt. Somit ergibt sich für $\alpha = 0,05$

$$a = \frac{\chi^2_{20;0,025}}{10} = 0,9591, \quad b = \frac{\chi^2_{20;0,975}}{10} = 3,417.$$

Der Mittelwert der Stichprobe beträgt 1,82. Wegen $a < 1,82 < b$ wird H_0 nicht abgelehnt.

Aufgabe 4 (12 Punkte)
Seien X, Y Zufallsvariablen mit Verteilungsfunktionen F_X, F_Y und $I \sim B(1,p)$ und X, Y, I seien unabhängig. Sei

$$Z := IX + (1-I)Y.$$

(a) Zeigen Sie $F_Z = pF_X + (1-p)F_Y$.

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

(c) Geben Sie ein Beispiel für die Verteilungen von X, Y an, so dass Z keine λ^1-Dichte besitzt.

Lösung
Zu (a)

$$\begin{aligned}
\mathbb{P}(Z \le z) &= \mathbb{P}(Z \le z|I = 1)\mathbb{P}(I = 1) + \mathbb{P}(Z \le z|I = 0)\mathbb{P}(I = 0) \\
&= \mathbb{P}(X \le z|I = 1)p + \mathbb{P}(Y \le z|I = 0)(1-p) \\
&= \mathbb{P}(X \le z)p + \mathbb{P}(Y \le z)(1-p) \\
&= pF_X(z) + (1-p)F_Y(z).
\end{aligned}$$

Die dritte Zeile folgt aus der Unabhängigkeit von I, X, Y.
Zu (b)

$$\begin{aligned}
\mathbb{E}(Z|I) &= \mathbb{E}(IX|I) + \mathbb{E}((1-I)Y|I) = I\mathbb{E}(X|I) + (1-I)\mathbb{E}(Y|I) \\
&= I\mathbb{E}(X) + (1-I)\mathbb{E}(Y).
\end{aligned}$$

Die letzte Gleichung folgt aus der Unabhängigkeit von I und X bzw. Y.

$$\begin{aligned}
\mathrm{Var}(Z|I) &= \mathbb{E}(Z^2|I) - \mathbb{E}(Z|I)^2 \\
&= \mathbb{E}(IX^2 + (1-I)Y^2|I) - (I\mathbb{E}(X)^2 + (1-I)\mathbb{E}(Y)^2) \\
&= \mathbb{E}(IX^2|I) + \mathbb{E}((1-I)Y^2|I) - I\mathbb{E}(X)^2 - (1-I)\mathbb{E}(Y)^2 \\
&= I\mathbb{E}(X^2|I) + (1-I)\mathbb{E}(Y^2|I) - I\mathbb{E}(X)^2 - (1-I)\mathbb{E}(Y)^2 \\
&= I\mathbb{E}(X^2) + (1-I)\mathbb{E}(Y^2) - I\mathbb{E}(X)^2 - (1-I)\mathbb{E}(Y)^2 \\
&= I\mathrm{Var}(X) + (1-I)\mathrm{Var}(Y).
\end{aligned}$$

Zu (c)
Für $X, Y \sim B(1,p)$ gilt $Z \sim B(1,p)$ wegen (a).

Aufgabe 5 (9 Punkte)
Sei (X, Y) eine zweidimensionale auf $[0,1] \times [0,1]$ gleichverteilte Zufallsvariable, also mit λ^2-Dichte $f = 1_{[0,1]\times[0,1]}$.

(a) Beweisen Sie $1_{[0,1]\times[0,1]}(x,y) = 1_{[0,1]}(x) \cdot 1_{[0,1]}(y)$ für alle $x, y \in \mathbb{R}^2$.

(b) Bestimmen Sie die Randverteilungen, also die Verteilung von X und Y.

(c) Sind X und Y unabhängig?

Lösung
Zu (a)

$$\begin{aligned}
1_{[0,1]\times[0,1]}(x,y) = 1 &\iff (x,y) \in [0,1] \times [0,1] \iff x \in [0,1], y \in [0,1] \\
&\iff 1_{[0,1]}(x) = 1_{[0,1]}(y) = 1 \iff 1_{[0,1]}(x) \cdot 1_{[0,1]}(y) = 1.
\end{aligned}$$

Zu (b)
Die Randdichte f_X ist gegeben durch

$$f_X(x) \quad = \quad \int_{\mathbb{R}} f(x,y)\,dy = 1_{[0,1]}(x) \int_{\mathbb{R}} 1_{[0,1]}(y)\,dy = 1_{[0,1]}(x)$$

und analog $f_Y(y) = 1_{[0,1]}(y)$. Somit sind X, Y jeweils gleichverteilt auf $[0,1]$.
Zu (c)
X und Y sind unabhängig, da wegen (a) die gemeinsame Dichte das Produkt der Dichten f_X, f_Y ist.

Aufgabe 6 (16 Punkte)
Sei X die Bearbeitungszeit für einen Schaden. Es wird angenommen, dass näherungsweise $X \sim N(\mu, 100)$ gilt. Es werden die Bearbeitungszeiten $x_1, \ldots, x_n$ voneinander unabhängiger Schäden aufgezeichnet.

(a) Bestimmen Sie dafür einen Maximum Likelihood Schätzer für μ.

(b) Für $n = 10$ ergebe die Summe der Bearbeitungszeiten 500. Geben Sie ein Schätzintervall zum Signifikanzniveau von 95% für μ an.

(c) Wie groß muss die Stichprobe sein, damit die Länge des Konfidenzintervalls maximal 2 beträgt?

Lösung
Zu (a)
Seien $x_1, \ldots, x_n$ die Daten. Dann ist eine Likelihood L bzw. eine Loglikelihood $\ell = \ln L$ gegeben durch

$$\begin{aligned}
L(\mu) &= \left(\frac{1}{2\pi \cdot 100}\right)^{n/2} \exp\left(-\frac{1}{200}\sum_{i=1}^{n}(x_i-\mu)^2\right) \\
\ell(\mu) &= -\frac{n}{2}\ln(2\pi \cdot 100) - \frac{1}{200}\sum_{i=1}^{n}(x_i-\mu)^2 \\
\ell'(\mu) &= \frac{1}{100}\sum_{i=1}^{n}(x_i-\mu) \\
\ell''(\mu) &= -\frac{n}{100} < 0.
\end{aligned}$$

Daraus ergibt sich durch Nullsetzen von ℓ' der Maximum Likelihood Schätzer $\hat{\mu} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
Zu (b)
Es handelt sich um den Fall einer Normalverteilung mit bekannter Varianz. Somit ergibt sich mit $u_{0,975}\dfrac{\sigma}{\sqrt{n}} = 1,96 \cdot \dfrac{10}{\sqrt{n}} = 6,20$ das Schätzintervall

$$[43,80; 56,20]$$

Zu (c)

Die Länge des Konfidenzintervalls beträgt $2 \cdot u_{0,975} \frac{\sigma}{\sqrt{n}} = \frac{39,2}{\sqrt{n}}$. Damit muss gelten

$$\frac{39,2}{\sqrt{n}} \leq 2 \Longleftrightarrow n \geq 384,16.$$

Die Stichprobe muss mindestens $n = 385$ betragen.