Eingangsprüfung Stochastik, 5. Oktober 2007

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

Aufgabe 1 (6 Punkte)
Sind die folgenden auf $[-1, 1]$ definierten Funktionen messbar? Begründen Sie Ihre Antwort:

(a) $\operatorname{sgn}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0. \end{cases}$

(b) $f(x) = \begin{cases} \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0. \end{cases}$

(c) $\operatorname{sgn}(f(x))$.

Lösung
Zu (a)
Es gilt

$$\operatorname{sgn} = (-1) \cdot 1_{[-1,0)} + 1_{[0,1]}.$$

Also ist sgn eine Treppenfunktion und somit messbar.
Zu (b)
Für $g(x) = \begin{cases} \dfrac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$ gilt

$$\{g < \alpha\} \quad = \begin{cases} [-1, 0] \cup \left(\dfrac{1}{\alpha}, \infty\right) & \alpha > 0 \\ (-\infty, 0) & \alpha = 0 \\ \left(\dfrac{1}{\alpha}, 0\right) & \alpha < 0 \end{cases}$$

d.h. für alle α ist $\{g < \alpha\}$ ein Intervall oder die Vereinigung zweier Intervalle. Somit ist g messbar. Da stetige Funktionen messbar sind, ist sin messbar und somit ist $f = \sin \circ g$ als Verkettung messbarer Funktionen messbar.
Zu (c)
Wegen (a) und (b) ist $\operatorname{sgn} \circ f$ als Verkettung messbarer Funktionen messbar.
Aufgabe 2 (10 Punkte)
Gegeben sei eine eine Stichprobe $x_1, x_2, \ldots, x_n$, d.h. n unabhängige Realisierungen einer Zufallsvariablen $X : \Omega \longrightarrow \mathbb{R}$. Die x_i, $i = 1, \ldots, n$ seien paarweise verschieden. Wir betrachten die empirische Verteilungsfunktion $F_n : \mathbb{R} \longrightarrow [0, 1]$,

$$F_n(x) := \frac{1}{n} \sum_{j=1}^{n} 1_{(-\infty,x]}(x_j) = \frac{\# \{j = 1, \ldots, n | x_j \leq x\}}{n}.$$

Sei $Y : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable mit der Verteilungsfunktion F_n.

(a) Welche Werte nimmt Y mit welcher Wahrscheinlichkeit an? Begründen Sie Ihre Antwort.

(b) Bestimmen Sie Erwartungswert und Varianz von Y.

(c) Bestimmen Sie für die messbare Funktion $g : \mathbb{R} \longrightarrow \mathbb{R}$

$$\int_{\mathbb{R}} g(x) dF_n(x).$$

Lösung
Zu (a)
Es werden die Werte $x_1, \ldots, x_n$ mit der Wahrscheinlichkeit $\dfrac{1}{n}$ angenommen:

$$\mathbb{P}(Y = x_i) = F_n(x_i) - F_n(x_i - 0) = \frac{1}{n} \left(\#\{j | x_j \leq x_i\} - \#\{j | x_j < x_i\}\right) = \frac{1}{n}.$$

Zu (b)

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{i=1}^{n} x_i \mathbb{P}(Y = x_i) = \sum_{i=1}^{n} x_i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x} \\ \mathrm{Var}(Y) &= \sum_{i=1}^{n} (x_i - \overline{x})^2 \mathbb{P}(Y = x_i) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2. \end{aligned}$$

Zu (c)
Sei δ_{x_i} das Dirac Maß. Wegen

$$\delta_{x_i}(-\infty, x] = 1_{(-\infty, x]}(x_i), \quad x \in \mathbb{R}$$

folgt

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}(-\infty, x], \quad x \in \mathbb{R}.$$

Das durch F_n induzierte Maß ist also $\frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$. Somit gilt für $g : \mathbb{R} \longrightarrow \mathbb{R}$ messbar:

$$\int_{\mathbb{R}} g \, dF_n(x) = \int_{\mathbb{R}} g \, d\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}\right)(x) = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} g \, d\delta_{x_i}(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i).$$

Aufgabe 3 (20 Punkte)
Sei $B \subset \mathbb{R}^2$ das Innere des Dreiecks mit den Ecken $(0,0)$, $(0,1)$ und $(1,1)$ und $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$,

$$f(x, y) := \begin{cases} \frac{1}{2\sqrt{xy}} & (x, y) \in B \\ 0 & \text{sonst} \end{cases}$$

die gemeinsame Dichte zweier Zufallsvariablen $X, Y : \Omega \longrightarrow \mathbb{R}$.

(a) Weisen Sie nach, dass f eine Dichte ist.

(b) Weisen Sie nach, dass $f_X : \mathbb{R} \longrightarrow \mathbb{R}, f_Y : \mathbb{R} \longrightarrow \mathbb{R}$

$$f_X(x) := \begin{cases} \frac{1-\sqrt{x}}{\sqrt{x}} & x \in (0,1) \\ 0 & \text{sonst} \end{cases}, \qquad f_Y(y) := \begin{cases} 1 & y \in (0,1) \\ 0 & \text{sonst} \end{cases}$$

die Dichten von X, Y sind.

(c) Bestimmen Sie die bedingten Dichten $f(x|y)$ und $f(y|x)$ und den bedingten Erwartungswert $\mathbb{E}(X|Y)$.

(d) Sind X, Y unabhängig? Begründen Sie Ihre Antwort.

Lösung
Zu (a)

$$\begin{aligned}\int_{\mathbb{R}^2} f(x,y)\,dxdy &= \int_0^1 \int_x^1 \frac{1}{2\sqrt{xy}}\,dydx = \int_0^1 \frac{1}{\sqrt{x}}\sqrt{y}\Big|_x^1\,dx \\ &= \int_0^1 \frac{1}{\sqrt{x}}(1-\sqrt{x})\,dx = 2\sqrt{x}\Big|_0^1 - 1 = 1.\end{aligned}$$

Zu (b)
Die Randdichten erhält man durch Integration

$$\begin{aligned}f_X(x) &= \int_{\mathbb{R}} f(x,y)\,dy \\ f_Y(y) &= \int_{\mathbb{R}} f(x,y)\,dx.\end{aligned}$$

Für $x \in (0,1)$ bzw. $y \in (0,1)$ ergibt sich

$$\begin{aligned}f_X(x) &= \int_x^1 \frac{1}{2\sqrt{xy}}dy = \frac{1}{\sqrt{x}}\sqrt{y}\Big|_x^1 = \frac{1-\sqrt{x}}{\sqrt{x}} \\ f_Y(y) &= \int_0^y \frac{1}{2\sqrt{xy}}dx = \frac{1}{\sqrt{y}}\sqrt{x}\Big|_0^y = 1.\end{aligned}$$

Für $x \notin (0,1)$ bzw. $y \notin (0,1)$ gilt $f_X(x) = f_Y(y) = 0$.

Zu (c)
Für $(x,y) \in B$ gilt

$$\begin{aligned}f(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{xy}} \\ f(y|x) &= \frac{f(x,y)}{f_X(x)} = \frac{1}{2\sqrt{xy}} \cdot \frac{\sqrt{x}}{1-\sqrt{x}} = \frac{1}{2\sqrt{y}(1-\sqrt{x})}.\end{aligned}$$

Daraus folgt nun für $y \in (0,1)$

$$\mathbb{E}(X|Y=y) = \int_0^y x f(x|y)\,dx = \int_0^y x\frac{1}{2\sqrt{xy}}\,dx = \frac{1}{\sqrt{y}}\frac{1}{3}x^{3/2}\Big|_0^y = \frac{y}{3}$$

also

$$\mathbb{E}(X|Y) = \frac{Y}{3}.$$

Zu (d)
Wären X, Y unabhängig, dann würde

$$f_X(x)f_Y(y) = \frac{1-\sqrt{x}}{\sqrt{x}} = \frac{1}{2\sqrt{xy}} = f(x,y) \text{ f.ü.}$$

gelten. Auf B gilt aber

$$\frac{1-\sqrt{x}}{\sqrt{x}} = \frac{1}{2\sqrt{xy}} \Longleftrightarrow y = \frac{1}{4}\left(\frac{1+\sqrt{x}}{1-x}\right)^2$$

also $f_X(\cdot)f_Y(\cdot) \neq f$ auf $B \setminus N$ mit $N := \left\{(x,y) \in B | y = 4\left(\frac{1+\sqrt{x}}{1-x}\right)^2\right\}$. N ist Teilmenge eines Funktionsgraphen, also eine Nullmenge. Also sind X, Y nicht unabhängig.

Aufgabe 4 (26 Punkte)

In $n = 14$ Jahren, wurden folgende Schadenhäufigkeiten (SH) pro Tausend versicherter Kfz und die mittlere Anzahl der gefahrenen Tausend Kilometer pro Kfz bestimmt:

Nr.	x_i - Tkm	y_i - SH
1	12,154	109
2	13,877	105
3	14,150	103
4	13,355	98
5	13,280	94
6	13,244	91
7	13,163	86
8	13,118	83
9	13,217	83
10	13,378	85
11	13,060	78
12	12,989	75
13	13,016	72
14	12,939	69

,

$\sum x_i^2 = 2445,75$, $\sum y_i^2 = 110289$, $\overline{x} = 13,21$, $\overline{y} = 87,93$, $\sum x_i y_i = 16279,39$.
Unterstellt wird folgender Zusammenhang zwischen den gefahrenen Kilometern x und der Zufallsvariablen Schadenhäufigkeit Y:

$$Y_i = a x_i + \varepsilon_i, \quad i = 1, \ldots, 14$$

wobei $\varepsilon_i \sim N(0, \sigma^2)$ und unabhängig sind und σ^2 bekannt sei.

(a) Bestimmen Sie Erwartungswert und Varianz von Y_i.

(b) Sei

$$\hat{A} := \frac{\sum\limits_{i=1}^{n} Y_i x_i}{\sum\limits_{i=1}^{n} x_i^2}.$$

Weisen Sie nach, dass $\hat{A} \sim N\left(a, \frac{\sigma^2}{\sum x_i^2}\right)$ gilt.

(c) Weisen Sie nach, dass $\hat{A}$ ein ML-Schätzer für a ist.

(d) Bestimmen Sie mit obigen Werten einen Schätzwert für a.

(e) Bestimmen Sie mit $\sigma = 12$ aus den obigen Daten ein 95 % - Konfidenzintervall für a.

Hinweis: Man zeige $\mathbb{P}(a \in [\hat{A} - \delta, \hat{A} + \delta]) = \mathbb{P}(\hat{A} \in [a - \delta, a + \delta])$

Lösung
Zu (a)

$$\begin{aligned}
\mathbb{E}(Y_i) &= \mathbb{E}(ax_i + \varepsilon_i) = ax_i + \mathbb{E}(\varepsilon_i) = ax_i, \\
\mathrm{Var}(Y_i) &= \mathrm{Var}(ax_i + \varepsilon_i) = \mathrm{Var}(\varepsilon_i) = \sigma^2.
\end{aligned}$$

Zu (b)
Es gilt $Y_i \sim N(ax_i, \sigma^2)$, da Y_i Summe einer normalverteilten Zufallsvariablen und einer Konstanten ist. Somit ist $\hat{A}$ normalverteilt, da $\hat{A}$ eine Linearkombination normalverteilter Zufallsvariablen ist.

$$\mathbb{E}(\hat{A}) = \mathbb{E}\left(\frac{\sum_{i=1}^{n} Y_i x_i}{\sum_{i=1}^{n} x_i^2}\right) = \frac{1}{\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} \mathbb{E}(Y_i) x_i = \frac{1}{\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} a x_i^2 = a.$$

Da die ε_i unabhängig sind, trifft dies auch auf die $Y_i x_i$ zu, und es folgt

$$\begin{aligned}
\mathrm{Var}(\hat{A}) &= \mathrm{Var}\left(\frac{\sum_{i=1}^{n} Y_i x_i}{\sum_{i=1}^{n} x_i^2}\right) = \frac{\mathrm{Var}\left(\sum_{i=1}^{n} Y_i x_i\right)}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sum_{i=1}^{n} \mathrm{Var}\left(Y_i x_i\right)}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sum_{i=1}^{n} x_i^2 \sigma^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.
\end{aligned}$$

Zu (c)
Da die ε_i unabhängig sind, trifft dies auch auf die Y_i zu. Dann ist die Likelihood L durch

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

gegeben. Es folgt:

$$\begin{aligned}
\ell(a) &:= \ln L(a) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - ax_i)^2 \\
\ell'(a) &= \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i (y_i - ax_i) \\
\ell''(a) &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 < 0.
\end{aligned}$$

ℓ' besitzt also in

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

ein globales Maximum.
Zu (d)

$$\hat{a} = \frac{16279,39}{2445,75} = 6,656.$$

Zu (e)
Der Hinweis folgt so:

$$\begin{aligned}
\mathbb{P}(a \in [\hat{A}-\delta, \hat{A}+\delta]) &= \mathbb{P}(\hat{A}-\delta \leq a \leq \hat{A}+\delta) \\
&= \mathbb{P}(-a-\delta \leq -\hat{A} \leq -a+\delta) \\
&= \mathbb{P}(a-\delta \leq \hat{A} \leq a+\delta) \\
&= \mathbb{P}(\hat{A} \in [a-\delta, a+\delta])
\end{aligned}$$

Gesucht ist ein $\delta > 0$ mit

$$0,95 = \mathbb{P}\left(\hat{A} \in [a-\delta, a+\delta]\right).$$

Es gilt wegen $\hat{A} \sim N\left(a, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$

$$\begin{aligned}
\mathbb{P}\left(\hat{A} \in [a-\delta, a+\delta]\right) &= \mathbb{P}\left(\hat{A} \leq a+\delta\right) - \mathbb{P}\left(\hat{A} \leq a-\delta\right) \\
&= \Phi\left(\frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sigma}\,\delta\right) - \Phi\left(-\frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sigma}\,\delta\right) \\
&= 2\Phi\left(\frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sigma}\,\delta\right) - 1.
\end{aligned}$$

Somit reicht es

$$\frac{\left(\sum_{i=1}^n x_i^2\right)^{1/2}}{\sigma}\delta = u_{0,975} = 1,96 \Longrightarrow \delta = 1,96 \cdot \frac{\sigma}{\left(\sum_{i=1}^n x_i^2\right)^{1/2}} = 0,476$$

zu wählen, es ergibt sich das Intervall

$$[6,180,\ 7,132].$$