Zulassungsprüfung Stochastik, 15.05.11

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.

Aufgabe 1 (14 Punkte)

(a) Seien $f_n : \Omega \longrightarrow [0, \infty)$ messbar. Zeigen Sie:

$f := \sum_{n=1}^{\infty} f_n$ ist messbar, und es gilt $\int f d\mu = \sum_{n=1}^{\infty} \int f_n d\mu$.

(b) Sei $b > 0$ und $f : (0, \infty) \longrightarrow \mathbb{R}$, $f(x) = \frac{1}{1 - e^{-bx}}$. Beweisen Sie

$$f(x) = \sum_{k=0}^{\infty} e^{-kbx}.$$

(c) Seien $a > 0$, $b > 0$ und $f : (0, \infty) \longrightarrow \mathbb{R}$, $f(x) = \frac{xe^{-ax}}{1 - e^{-bx}}$. Beweisen Sie

$$\int_{(0,\infty)} f(x) d\lambda(x) = \sum_{k=0}^{\infty} \frac{1}{(a + kb)^2}.$$

Lösung

Zu (a)

Die Funktionen $g_n := \sum_{k=1}^{n} f_k$ sind als endliche Summen messbarer Funktionen messbar. Wegen

$$\lim_{n\to\infty} g_n = \sum_{k=1}^{\infty} f_k$$

ist f als Grenzwert messbarer Funktionen messbar. Es gilt sogar

$$g_n \nearrow f \quad (n \to \infty)$$

also folgt mit dem Satz von der monotonen Konvergenz

$$\int f d\mu = \lim_{n\to\infty} \int g_n d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} \int f_k d\mu = \sum_{k=1}^{\infty} \int f_k d\mu.$$

Zu (b)

Es handelt sich um die geometrische Reihen $\sum_{k=0}^{\infty} \left(e^{-bx}\right)^k$. Wegen $b > 0$ gilt $e^{-bx} < 1$ für alle $x > 0$, also konvergiert obige Reihe für alle $x > 0$ gegen $\frac{1}{1 - e^{-bx}}$.

Zu (c)

Laut (b) gilt $f(x) = \sum_{k=0}^{\infty} xe^{-ax-kbx}$. Für $k \in \mathbb{N}$ und $R > 0$ gilt

$$\begin{aligned}
\int_0^R xe^{-ax-kbx}\,dx &= -\frac{xe^{-x(a+kb)}}{a+kb}\bigg|_0^R + \frac{1}{a+kb}\int_0^R e^{-x(a+kb)}\,dx = \\
&= -\frac{Re^{-R(a+kb)}}{a+kb} - \frac{1}{(a+kb)^2}e^{-ax-kbx}\bigg|_0^R \overset{R\to\infty}{\longrightarrow} \frac{1}{(a+kb)^2}
\end{aligned}$$

also

$$\int_0^{\infty} xe^{-ax-kbx}\,dx = \int_{(0,\infty)} xe^{-ax-kbx}\,d\lambda(x) = \frac{1}{(a+kb)^2}.$$

Die Behauptung folgt nun mit (a) und der obigen Reihendarstellung von f.

Aufgabe 2 (14 Punkte) Sei X eine Zufallsvariable mit Dichte $f(x) = \begin{cases} \frac{e^{-x}}{\sqrt{\pi x}} & x > 0 \\ 0 & \text{sonst} \end{cases}$.
Sei Y eine weitere Zufallsvariable, die bei gegebenem $X = x > 0$ normalverteilt ist mit Erwartungswert 0 und Varianz $\frac{1}{2x}$.

(a) Bestimmen Sie die gemeinsame Dichte f von Y und X.

(b) Beweisen Sie, dass für die bedingte Dichte $f_{X|Y}$ von X gegeben Y

$$f_{X|Y}(x|y) = (1+y^2)e^{-(1+y^2)x} \cdot 1_{(0,\infty)}(x)$$

gilt.

(c) Sind Y und X unabhängig?

(d) Bestimmen Sie $\mathbb{P}(X \leq x|Y)$ für $x > 0$. Um welche Verteilung handelt es sich?

(e) Bestimmen Sie $\mathbb{E}(X|Y)$ und $\mathrm{Var}(X|Y)$.

Lösung
Seien f_X und f_Y die Randdichten von (X, Y).
Zu (a)
Für die bedingte Dichte $f_{Y|X}$ von Y gegeben X gilt

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \text{ wenn } f_X(x) > 0.$$

Auflösen nach $f(x, y)$ ergibt für $x > 0$

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

Für $x \leq 0$ ist $f(x, y) = 0$.
Zu (b)

Aus (a) ergibt sich

$$f_Y(y) = \frac{1}{\pi}\int_0^\infty e^{-x(1+y^2)}\,dx = -\frac{1}{\pi(1+y^2)}e^{-x(1+y^2)}\Big|_0^\infty = \frac{1}{\pi(1+y^2)}.$$

Es gilt somit für $x > 0$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi}e^{-x(1+y^2)}}{\frac{1}{\pi(1+y^2)}} = e^{-x(1+y^2)}(1+y^2).$$

Für $x \leq 0$ ist $f_{X|Y}(x|y) = 0$.
Zu (c)
Die Zufallsvariablen X, Y sind nicht unabhängig, da die gemeinsame Dichte nicht Produkt der Randdichten ist.
Zu (d)
Es handelt sich um eine $\mathcal{E}(1+Y^2)$-Verteilung, somit gilt für $x > 0$

$$\mathbb{P}(X \leq x|Y) = 1 - e^{-x(1+Y^2)}.$$

Zu (e)
Wegen (d) kann man Erwartungswert und Varianz von $\mathcal{E}(1+y^2)$ verwenden:

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \int_0^\infty x f_{X|Y}(x|y)dx = \frac{1}{1+y^2} \Longrightarrow \mathbb{E}(X|Y) = \frac{1}{1+Y^2} \\ \mathrm{Var}(X|Y) &= \frac{1}{(1+Y^2)^2}.\end{aligned}$$

Aufgabe 3 (18 Punkte)
Sei X_{ij} die Anzahl der zahlenden Kunden in einem Warenhaus, die in der j-ten Minute der i-ten Stunde in der Zeit zwischen 16:00 und 20:00 zahlen. Es sei $\mathbb{E}(X_{ij}) = 15$, $\mathrm{Var}(X_{ij}) = 100$, die X_{ij} seien unabhängig. Sei X die Anzahl der zahlenden Kunden in diesen vier Stunden. Y sei die Anzahl der Kunden, die bargeldlos während dieser vier Stunden zahlen. Wir gehen im folgenden davon aus, dass X normalverteilt ist.

(a) Bestimmen Sie $\mathbb{E}(X)$ und $\mathrm{Var}(X)$.

(b) Bestimmen Sie die Wahrscheinlichkeit, dass nicht mehr als 3500 Kunden während der 4 Stunden bezahlen.

(c) 1/6 der Kunden zahlt bargeldlos. Bestimmen Sie Erwartungswert, Varianz und Verteilung von Y.

(d) Die Kapazität für bargeldloses Zahlen liegt bei 660. Bestimmen Sie die Wahrscheinlichkeit, dass diese Kapazität überschritten wird.

(e) Bei n beobachteten Tagen sei N die Anzahl der Tage, an denen die Kapazität für bargeldloses Zahlen überschritten wird. Bestimmen Sie die Verteilung von N.

(f) Diskutieren Sie die Annahme der Normalverteilung für X.

Lösung
Zu (a)
Laut Definition gilt $X = \sum\limits_{i=1}^{4} \sum\limits_{j=1}^{360} X_{ij}$. Somit gilt

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{4} \sum_{j=1}^{60} \mathbb{E}(X_{ij}) = 4 \cdot 15 \cdot 60 = 3600 \\
\mathrm{Var}(X) &= \sum_{i=1}^{4} \sum_{j=1}^{60} \mathrm{Var}(X_{ij}) = 24000.
\end{aligned}$$

Zu (b)
Da X normalverteilt ist mit den Parametern aus (a) folgt

$$\mathbb{P}(X > 3500) \quad = \quad 1 - \mathbb{P}(X \leq 3500) = \Phi\left(\frac{100}{\sqrt{24000}}\right) \approx \Phi(0,65) = 0,742.$$

Zu (c)
Laut Angabe gilt $Y = \frac{X}{6}$, damit folgt

$$\mathbb{E}(Y) = \frac{\mathbb{E}(X)}{6} = 600, \quad \mathrm{Var}(Y) = \frac{\mathrm{Var}(X)}{36} = \frac{2000}{3}$$

also $Y \sim N\left(600, \frac{2000}{3}\right)$.
Zu (d)
Aufgrund von (c) ergibt sich mit der Standardnormalverteilung

$$\mathbb{P}(Y > 660) = 1 - \mathbb{P}(Y \leq 660) = 1 - \Phi\left(\frac{60}{\sqrt{666,67}}\right) = 0,0102.$$

Zu (e)
Die Wahrscheinlichkeit, dass an einem Tag die Kapazität überschritten wird ist $p = 0,0102$. Somit ist $N \sim B(n,p)$ verteilt.
Zu (f)
X Summe einer großen Anzahl unabhängiger, identisch verteilter Zufallsvariablen mit endlicher Varianz. Mit dem zentralen Grenzwertsatz kann man daher davon ausgehen, dass X normalverteilt ist.

Aufgabe 4 (20 Punkte)
Für $x \in \mathbb{R}$ sei $\lfloor x \rfloor$ die größte ganze Zahl, die kleiner als x ist, z.B. $\lfloor 3,1 \rfloor = 3, \lfloor 3,8 \rfloor = 3$.

Sei $X \sim \mathrm{Exp}(\lambda)$ und die diskrete Zufallsvariable $Y := \lfloor X \rfloor$.

(a) Beweisen Sie, dass für die Wahrscheinlichkeitsfunktion von Y gilt:

$$\mathbb{P}(Y = y) = e^{-\lambda y}\left(1 - e^{-\lambda}\right), \quad y \in \mathbb{N}_0.$$

(b) Um welche bekannte Verteilung handelt es sich in (a)? Bestimmen Sie den Erwartungswert und die Varianz von Y.

(c) Bestimmen Sie einen Maximum Likelihood Schätzer für λ aus einer Stichprobe von unabhängig und identisch verteilten $Y_1, \dots Y_n \sim Y$.

(d) Gegeben sei eine Stichprobe mit Stichprobenlänge 15, Stichprobenmittel 1,6 und Stichprobenvarianz 4. Bestimmen Sie den Maximum Likelihood Schätzwert für λ.

Lösung
Zu (a)

$$\mathbb{P}(Y = y) = \mathbb{P}(y \leq X < y+1) = \mathbb{P}(y < X \leq y+1) = \mathbb{P}(X \leq y+1) - \mathbb{P}(X \leq y).$$

Für die Verteilungsfunktion von gilt

$$\mathbb{P}(X \leq x) = 1 - e^{-\lambda x}$$

und somit folgt weiter

$$\mathbb{P}(Y = y) = \left(1 - e^{-\lambda(y+1)}\right) - \left(1 - e^{-\lambda y}\right) = e^{-\lambda y} - e^{-\lambda(y+1)} = e^{-\lambda y}\left(1 - e^{-\lambda}\right).$$

Zu (b)
Es handelt sich um die geometrische Verteilung, $Geo(1-e^{-\lambda}) = NB(1, 1-e^{-\lambda})$. Damit folgt

$$\mathbb{E}(Y) = \frac{e^{-\lambda}}{1 - e^{-\lambda}}, \quad \mathrm{Var}(Y) = \frac{e^{-\lambda}}{(1 - e^{-\lambda})^2}.$$

Zu (c)
Aus der Likelihood L ergibt sich

$$\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} e^{-\lambda y_i}\left(1 - e^{-\lambda}\right) = \left(1 - e^{-\lambda}\right)^n \prod_{i=1}^{n} e^{-\lambda y_i} \\
\ell(\lambda) &= \ln L(\lambda) = n \ln\left(1 - e^{-\lambda}\right) - \sum_{i=1}^{n} \lambda y_i \\
\ell'(\lambda) &= \frac{n e^{-\lambda}}{1 - e^{-\lambda}} - \sum_{i=1}^{n} y_i = \frac{n}{e^{\lambda} - 1} - \sum_{i=1}^{n} y_i \\
\ell''(\lambda) &= -\frac{n e^{\lambda}}{(1 - e^{\lambda})^2} < 0
\end{aligned}$$

Aus $\ell'(\lambda) = 0$ ergibt sich wegen $\ell''(\lambda) < 0$ der ML-Schätzer

$$\hat{\lambda} = \ln\left(\frac{n + \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i}\right) = \ln\left(\frac{1 + \bar{y}}{\bar{y}}\right) \text{ mit } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Zu (d)
Laut (c) erhält man den Schätzwert $\hat{\lambda} = \ln\left(\frac{1+\bar{y}}{\bar{y}}\right) = \ln\left(\frac{1+1{,}6}{1{,}6}\right) = 0,49$.

Aufgabe 5 (24 Punkte)
In der folgenden Tabelle sind die relativen Sterbehäufigkeiten in ‰ eines homogenen Kollektivs der letzten 10 Jahre aufgezeichnet.

Jahr	1	2	3	4	5	6	7	8	9	10
rel. Häuf. (‰)	3,51	3,46	3,46	3,38	3,36	3,31	3,07	2,98	2,81	2,75

Bezeichnet man mit x_i das Jahr und mit y_i die relative Häufigkeit, so wird für die Zufallsvariablen Y_i das Modell

$$Y_i = a + bx_i + \varepsilon_i,\ i = 1, \dots, 10$$

untersucht. Die ε_i seien unabhängig und $N(0, \sigma^2)$-verteilt. Sie können verwenden:

$$\overline{x} = 5,5;\ \sum_{i=1}^{10}(x_i - \overline{x})^2 = 82,5$$

$$\overline{y} = 3,209;\ \sum_{i=1}^{10}(y_i - \overline{y})^2 = 0,725;\ \sum_{i=1}^{10}(y_i - \overline{y})(x_i - \overline{x}) = -7,407.$$

Stützen die Daten die Behauptung, dass die Sterblichkeit

(a) vom Beobachtungsjahr abhängt?

(b) sich jedes Jahr um 0,1 ‰-Punkte verringert?

Beide Aussagen sollen zu einem Niveau von $\alpha = 5$ % erfolgen.

Lösung
Zu (a)
Zu untersuchen ist die Hypothese $H_0 : b = 0$. Es ergeben sich folgende Schätzwerte:

$$\begin{aligned}
\hat{b} &= \frac{-7,407}{82,5} = -0,09 \\
\hat{\sigma}^2 &= \frac{1}{8}\left(0,725 - \frac{7,407^2}{82,5}\right) = 0,0075 \\
\mathrm{se}(\hat{b}) &= \sqrt{\frac{0,0075}{82,5}} = 0,0095 \\
T &= \frac{\hat{b} - b_0}{\mathrm{se}(\hat{b})} = \frac{-0,09}{0,0095} = 9,439.
\end{aligned}$$

H_0 wird abgelehnt, denn $|T| > t_{8;0,975} = 2,308$. Somit wird die Hypothese, dass die Sterblichkeit nicht vom Beobachtungsjahr abhängt, verworfen.
Zu (b) Zu untersuchen ist nun die Hypothese $H_0 : b = -0,1$. Mit den Werten aus (a) ergibt sich nun die Testgröße

$$T = \frac{\hat{b} - b_0}{\mathrm{se}(\hat{b})} = \frac{-0,09 + 0,1}{0,0095} = 1,053.$$

Die Hypothese, dass sich die Sterblichkeit jähhrlich um 0,1 ‰ verbessert wird also nicht verworfen.