Zulassungsprüfung Stochastik, 05.05.12

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 (14 Punkte) Sei $g : \Omega \longrightarrow \Omega$ messbar. Das Bildmaß μ_g von μ bezüglich g sei invariant, also

$$\forall A \in \mathcal{A} : \quad \mu_g(A) = \mu(A).$$

Beweisen Sie:

(a) Für alle $A \in \mathcal{A}$ gilt $1_A \circ g = 1_{g^{-1}(A)}$.
Hierbei wird mit 1_A die Indikatorfunktion von A bezeichnet.

(b) Sei $f : \Omega \longrightarrow [0, \infty)$ messbar. Dann gilt $\displaystyle\int_\Omega f\, d\mu = \int_\Omega f \circ g\, d\mu$.

Lösung
Zu (a)
Es gilt für alle $A \in \mathcal{A}$ und $\omega \in \Omega$

$$1_A(g(\omega)) = 1 \Longleftrightarrow g(\omega) \in A \Longleftrightarrow \omega \in g^{-1}(A) \Longleftrightarrow 1_{g^{-1}(A)}(\omega) = 1.$$

Zu (b)
Sei $f = \sum_{i=1}^n \alpha_i 1_{A_i}$, $\alpha_i \geq 0$, $A_i \in \mathcal{A}$, $i = 1, \ldots, n$ eine positive Treppenfunktion. Dann gilt wegen (a)

$$f \circ g = \sum_{i=1}^n \alpha_i 1_{A_i} \circ g = \sum_{i=1}^n \alpha_i 1_{g^{-1}(A_i)}$$

und somit folgt aus

$$\begin{aligned}
\int_\Omega f \circ g\, d\mu &= \sum_{i=1}^n \alpha_i \int_\Omega 1_{g^{-1}(A_i)}\, d\mu = \sum_{i=1}^n \alpha_i \mu(g^{-1}(A_i)) = \sum_{i=1}^n \alpha_i \mu_g(A_i) \\
&= \sum_{i=1}^n \alpha_i \mu(A_i) = \sum_{i=1}^n \alpha_i \int_{A_i} d\mu = \int_\Omega \sum_{i=1}^n \alpha_i 1_{A_i} d\mu = \int_\Omega f\, d\mu
\end{aligned}$$

die Behauptung für positive Treppenfunktionen.
Sei nun $f \geq 0$. Es gibt eine Folge $(f_n)_{n\in\mathbb{N}}$ von positiven Treppenfunktionen mit $f_n \geq 0$ und $f_n \nearrow f$. Da f_n eine Treppenfunktion ist, gilt wie oben gezeigt

$$\int_\Omega f_n\, d\mu = \int_\Omega f_n \circ g\, d\mu$$

Aus $f_n \nearrow f$ folgt $f_n \circ g \nearrow f \circ g$ und mit dem Satz von der monotonen Konvergenz folgt für die beiden Seiten

$$\begin{aligned}
\lim_{n\to\infty} \int_\Omega f_n \circ g\, d\mu &= \int_\Omega f \circ g\, d\mu, \\
\lim_{n\to\infty} \int_\Omega f_n\, d\mu &= \int_\Omega f\, d\mu
\end{aligned}$$

und somit insgesamt :

$$\int_\Omega f \circ g \, d\mu = \int_\Omega f \, d\mu.$$

Aufgabe 2 (18 Punkte)
Seien $X, Y \sim U(0,1)$ unabhängig, und sei $V := \max\{X, Y\}$, $W := \min\{X, Y\}$.

(a) Bestimmen Sie die Verteilungsfunktionen von V und W.
[Ergebnis: $F_V(v) = v^2$, $F_W(w) = 2w - w^2$]

(b) Bestimmen Sie die Erwartungswerte von V und W.

(c) Beweisen Sie $VW = XY$.

(d) Bestimmen Sie $\mathrm{Cov}(V, W)$.

(e) Sind V und W unabhängig?

Lösung
Zu (a)
Die Verteilungsfunktion F der Gleichverteilung auf $(0,1)$ ist gegeben durch $F : \mathbb{R} \longrightarrow [0,1]$, $F(x) = \begin{cases} x & \text{falls } x \in [0,1] \\ 1 & \text{falls } x > 1 \\ 0 & \text{sonst} \end{cases}$. Sei $v, w \in \mathbb{R}$. Es gilt wegen der Unabhängigkeit von X, Y

$$\begin{aligned}
\mathbb{P}(V \leq v) &= \mathbb{P}(\max\{X,Y\} \leq v) = \mathbb{P}(X \leq v, Y \leq v) = \mathbb{P}(X \leq v)\mathbb{P}(Y \leq v) \\
&= F(v)^2 \\
\mathbb{P}(W \leq w) &= \mathbb{P}(\min\{X,Y\} \leq w) = 1 - \mathbb{P}(\min\{X,Y\} > w) \\
&= 1 - \mathbb{P}(X > w, Y > w) = 1 - \mathbb{P}(X > w)\mathbb{P}(Y > w) \\
&= 1 - (1 - F(w))^2 = 1 - 1 + 2F(w) - F(w)^2 = 2F(w) - F(w)^2.
\end{aligned}$$

Damit ergeben sich die Verteilungsfunktionen F_V, F_W wie folgt:

$$F_V(v) = \begin{cases} v^2 & \text{falls } v \in [0,1] \\ 1 & \text{falls } v > 1 \\ 0 & \text{sonst} \end{cases} \quad \text{und } F_W(w) = \begin{cases} 2w - w^2 & \text{falls } w \in [0,1] \\ 1 & \text{falls } w > 1 \\ 0 & \text{sonst} \end{cases}$$

Zu (b)
Aus (a) erhält man die Dichten f_V und f_W von V, W:

$$\begin{aligned}
f_V(v) &= 2v1_{(0,1)} \\
f_W(w) &= 2(1-w)1_{(0,1)}.
\end{aligned}$$

Damit folgt

$$\begin{aligned}
\mathbb{E}(V) &= \int_0^1 2v^2 \, dv = \frac{2}{3} \\
\mathbb{E}(W) &= \int_0^1 2w - 2w^2 \, dv = 1 - \frac{2}{3} = \frac{1}{3}.
\end{aligned}$$

Zu (c)
Sei $\omega \in \Omega$.
Im Fall $X(\omega) \leq Y(\omega)$ gilt $V(\omega) = Y(\omega)$ und $W(\omega) = X(\omega)$, und damit $X(\omega)Y(\omega) = V(\omega)W(\omega)$.
Im Fall $X(\omega) \geq Y(\omega)$ ergibt sich analog $X(\omega)Y(\omega) = V(\omega)W(\omega)$.

Zu (d)
Wegen (b), (c) und der Unabhängigkeit von X, Y gilt

$$\begin{aligned} \mathrm{Cov}(V,W) &= \mathbb{E}(VW) - \mathbb{E}(V)\mathbb{E}(W) = \mathbb{E}(XY) - \frac{2}{9} = \mathbb{E}(X)\mathbb{E}(Y) - \frac{2}{9} \\ &= \frac{1}{4} - \frac{2}{9} = \frac{1}{36}. \end{aligned}$$

Zu (e)
V und W sind nicht unabhängig, da sonst die Kovarianz verschwinden würde.

Aufgabe 3 (16 Punkte)
Die Zufallsvariable A sei $U[a,b]$ verteilt und die Zufallsvariable X sei unter A bedingt Pareto–verteilt, d.h., für $A = \alpha > 0$ besitzt X die Verteilungsfunktion

$$F_X(x) = \begin{cases} 1 - \dfrac{1}{x^\alpha} & x \geq 1 \\ 0 & \text{sonst} \end{cases}.$$

(a) Bestimmen Sie den bedingten Erwartungswert $\mathbb{E}(X|A = \alpha)$. Für welche α existiert er?

[Kontrollergebnis $\mathbb{E}(X|A) = 1 + \dfrac{1}{A-1}$]

(b) Bestimmen Sie $\mathbb{E}(X)$.

(c) Sind die Zufallsvariablen X und A unabhängig?

(d) Bestimmen Sie die Wahrscheinlichkeit $\mathbb{P}(X \geq x)$.

Lösung
Zu a)
Die bedingte Dichte von X gegeben $A = \alpha$ erhält man durch Ableitung von F_X nach x:

$$f(x|\alpha) = \begin{cases} \alpha x^{-\alpha-1} & x > 1 \\ 0 & \text{sonst} \end{cases}.$$

Daraus folgt

$$\mathbb{E}(X|A=\alpha) = \int_1^\infty \alpha x x^{-\alpha-1}\,dx \quad = \quad \alpha \int_1^\infty x^{-\alpha}\,dx.$$

Das uneigentliche Integral existiert nur für $\alpha > 1$ und in diesem Fall gilt

$$\mathbb{E}(X|A=\alpha) = \frac{\alpha}{\alpha-1} = 1 + \frac{1}{\alpha-1}.$$

Zu (b)

Mit dem Satz vom iterierten Erwartungswert gilt

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|A)) = \mathbb{E}\left(1+\frac{1}{A-1}\right) = \frac{1}{b-a}\int_a^b 1+\frac{1}{\alpha-1}\,d\alpha \\ &= \frac{1}{b-a}\left((b-a)+\ln\frac{b-1}{a-1}\right) = 1+\frac{1}{b-a}\ln\frac{b-1}{a-1}.\end{aligned}$$

Zu (c)
Die Zufallsvariablen X und A sind nicht unabhängig, da sonst $E(X|A) = E(X)$ gelten würde, was offensichtlich nicht der Fall ist.
Zu (d)

$$\begin{aligned}\mathbb{P}(X>x) &= \frac{1}{b-a}\int_a^b \mathbb{P}(X>x|A=\alpha)\,d\alpha = \frac{1}{b-a}\int_a^b x^{-\alpha}\,d\alpha \\ &= \frac{1}{b-a}\int_a^b e^{-\alpha\ln x}\,d\alpha = -\frac{1}{b-a}\frac{1}{\ln x}e^{-\alpha\ln x}\Big|_a^b = -\frac{1}{b-a}\frac{1}{\ln x}x^{-\alpha}\Big|_a^b \\ &= \frac{1}{b-a}\frac{1}{\ln x}(x^{-a}-x^{-b})\end{aligned}$$

Aufgabe 4 (18 Punkte)
Seien $X \sim N\left(1,\sigma^2\right)$ mit $\sigma^2 > 0$. Es sei eine Stichprobe vom Umfang n durch unabhängige Realisierungen $x_1,\ldots,x_n$ gegeben. Bestimmen Sie einen Maximum-Likelihood-Schätzer für den Parameter σ^2.
Lösung
Wegen der Unabhängigkeit ist die Likelihood gegeben durch

$$\begin{aligned}L(\sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{(x_i-1)^2}{\sigma^2}\right) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^n\frac{(x_i-1)^2}{\sigma^2}\right).\end{aligned}$$

Somit folgt für $\ell := \ln L$

$$\begin{aligned}\ell(\sigma^2) &= -\frac{n}{2}\left(\ln(2\pi)+\ln(\sigma^2)\right) - \frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-1)^2 \\ \ell'(\sigma^2) &= -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^n(x_i-1)^2 \\ \ell''(\sigma^2) &= \frac{n}{2(\sigma^2)^2}-\frac{1}{(\sigma^2)^3}\sum_{i=1}^n(x_i-1)^2\end{aligned}$$

Aus der zweiten Gleichung ergibt sich durch Nullsetzen und Auflösen

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n(x_i-1)^2$$

und wegen

$$\ell''(\hat{\sigma}^2) = \frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3}\sum_{i=1}^{n}(x_i-1)^2 = \frac{n}{2(\hat{\sigma}^2)^2} - \frac{n}{(\hat{\sigma}^2)^2} = -\frac{n}{2(\hat{\sigma}^2)^2} < 0$$

folgt, dass in $\hat{\sigma}^2$ ein Maximum vorliegt. Der gesuchte Schätzer ist

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-1)^2.$$

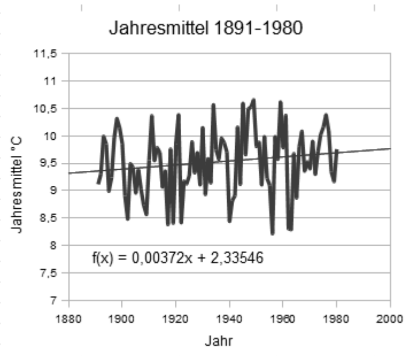
Aufgabe 5 (24 Punkte)
Es liegen 120 Daten vor, die Temperatur-Jahresmittelwerte der Jahre 1891-2010. Folgende Zusammenfassung wird geliefert:

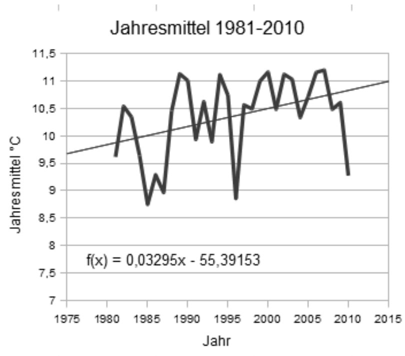
Zeitraum	Mittelwert	empirische Standardabweichung
1891-1980	$9,5$ °C	0,60 °C
1981-2010	$10,4$ °C	0,75 °C
1891-2010	$9,7$ °C	0,73 °C

Wir gehen davon aus, dass die Jahresmitteltemperatur eines Jahres normalverteilt ist und die Jahresmitteltemperaturen jeweils unabhängig voneinander sind.

(a) Wie beurteilen Sie die Behauptung, dass die Jahresmitteltemperatur in den Jahren 1981-2010 gegenüber den Jahren 1891-1980 nicht gestiegen ist? Geben Sie das Ergebnis zu einem Signifikanzniveau von 5 % an. Welche Annahmen treffen Sie? Begründen Sie ihr Vorgehen.

(b) In den folgenden Graphiken sind die Jahresmittel der Jahre 1891-1980 und 1981-2010 dargestellt zusammen mit den Ausgleichsgeraden.





Eine Statistiksoftware liefert noch die folgenden Werte:

$$\sum_{i=1891}^{1980}(y_i-\hat{y}_i)^2 = 31,63, \qquad \sum_{j=1891}^{1980}(j-1935,5)^2 = 60742,5$$

$$\sum_{i=1981}^{2010}(y_i-\hat{y}_i)^2 = 13,71, \qquad \sum_{j=1981}^{2010}(j-1995,5)^2 = 2247,5$$

Wie beurteilen Sie aufgrund dieser Daten mit dem Modell der einfachen linearen Regression die Behauptung, dass das Jahresmittel der Jahre

(i) 1891-1980 (ii) 1981-2010

nicht vom Jahr abhängt? Verwenden Sie das Signifikanzniveau von 5 %. Beschreiben Sie das von Ihnen verwendete Modell.

(c) Ist Ihr Vorgehen in (a) im Lichte Ihrer Ergebnisse von (b) gerechtfertigt?

(d) Untersuchen Sie den linearen Trend im Zeitraum 1981-2010. Stützen die Daten die Behauptung, dass er

(i) 0,3 °C (ii) 0,2 °C (iii) 0,4°C

pro Jahrzehnt beträgt? Verwenden Sie das Signifikanzniveau von 5 %.

(e) Eine Klimainitiative stellt folgende Behauptungen auf:

(1) Die Temperatur der Jahre 1981-2010 ist im Vergleich zu den Jahren 1891-1980 angestiegen.

(2) Der lineare Erwärmungstrend über die vergangenen 30 Jahre (1981-2010) liegt bei 0,3 °C pro Jahrzehnt.

Kommentieren Sie kurz die Behauptungen der Klimainitiative?

Lösung

Zu (a)

Sei t_i das Jahresmittel des Jahres $i = 1891, \dots, 2010$. Dann wird angenommen, dass die t_i unabhängige Realisierungen von Zufallsvariablen T_i sind mit

$$\begin{aligned} T_i &\overset{\text{iid}}{\sim} X \text{ wobei } X \sim N(\mu_1, \sigma_1^2),\ i = 1891, \dots, 1980 \\ T_i &\overset{\text{iid}}{\sim} Y \text{ wobei } Y \sim N(\mu_2, \sigma_2^2),\ i = 1981, \dots, 2010. \end{aligned}$$

Nullhypothese ist hier

$H_0 : \mu_1 \geq \mu_2$.

Es wird der einseitige t-Test verwendet, da die Varianzen unbekannt sind. Die Testgröße ist

$$T = \frac{\overline{y} - \overline{x}}{s} \sqrt{\frac{mn}{m+n}}, \quad s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{m+n-2}.$$

Konkret ergibt sich mit $\overline{y} = 10,4$, $\overline{x} = 9,5$, $n = 90$, $m = 30$

$$\begin{aligned} T &= \frac{0,9}{s} \sqrt{\frac{2700}{120}} \qquad s = \sqrt{\frac{89 \cdot 0,6^2 + 29 \cdot 0,75^2}{118}} = 0,64 \\ &= 6,67. \end{aligned}$$

Die Hypothese wird wegen $6,67 > t_{118,0.95} \approx 1,65$ verworfen.

Bemerkung zu H_0: Die Wahrscheinlichkeit fälschlicherweise anzunehmen, dass die mittlere Temperatur gestiegen ist beträgt bei diesem Vorgehen 5 %. Würde man $\mu_1 \leq \mu_2$ testen, wäre keine Aussage über diese Wahrscheinlichkeit möglich.

Zu (b)

Es werden die Modelle $T_i = a + bx_i + \varepsilon_i$ für $i = 1891, \dots, 1980$ bzw. $i = 1981, \dots, 2010$ getrennt untersucht. Die T_i sind die Jahresmittelwerte, x_i die

Jahre und ε_i sind in jedem der beiden Modelle jeweils unabhängig, identisch normalverteilt mit Erwartungswert 0. Die Nullhypothese lautet jeweils
$H_0 : b = 0$.
Zu (i) Die Testgröße lautet

$$\begin{aligned}\frac{\hat{b}}{\mathrm{se}(b)} &= \frac{0,00372}{0,00243} \text{ mit } \mathrm{se}(b) = \sqrt{\frac{31,63/88}{60742,5}} = 0,00243 \\ &= 1,53.\end{aligned}$$

Die Hypothese wird wegen $t_{88,0.95} \approx 1,96$ nicht verworfen.
Zu (ii) Die Testgröße lautet

$$\begin{aligned}\frac{\hat{b}}{\mathrm{se}(b)} &= \frac{0,03295}{0,01476} \text{ mit } \mathrm{se}(b) = \sqrt{\frac{13,71/28}{2247,5}} = 0,01476 \\ &= 2,23.\end{aligned}$$

Die Hypothese wird wegen $t_{28,0.95} \approx 2,048$ verworfen.
Zu (c)
Das Vorgehen in (a) ist nicht korrekt, da die Annahme, dass die Jahresmittel der Jahre 1981-2010 identisch verteilt sind, verletzt ist: der Erwartungswert ist über die Jahre nicht konstant.
Zu (d)
Zu untersuchen sind die Nullhypothesen
(i) $H_0 : b = 0,03$ (ii) $H_0 : b = 0,02$ (iii) $H_0 : b = 0,04$
Mit den Zahlen aus (b) also $\mathrm{se}(b) = 0,01476$ ergeben sich die Testgrößen

$$\frac{0,00295}{0,01476} = 0,2 \qquad \text{(i)}$$

$$\frac{0,01295}{0,01476} = 0,88 \qquad \text{(ii)}$$

$$-\frac{0,00705}{0,01476} = -0,48 \qquad \text{(iii)}$$

Somit kann keine der angegebenen Hypothesen verworfen werden.
Zu (e)
Es fehlt die Angabe des Signifikanzniveaus. Ferner wäre bei (2) die Angabe eines Schätzintervalls sinnvoller gewesen. Alle drei Behauptungen können aber bei einem Niveau von 5 % nicht verworfen werden.