Prof. Dr. U. Kamps
C. Empacher, D. Rauwolf RWTH Aachen, SoSe 2022

14. Juli 2022

Einführung in die angewandte Stochastik

7. Globalübung - Lösungen

Aufgabe 32

(a)

Zugehörige Berechnungstabelle:

i	x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$
1	2.49	6.7	6.2001	44.89	16.683
2	2.68	6.2	7.1824	38.44	16.616
3	2.62	5.8	6.8644	33.64	15.196
4	2.51	6.2	6.3001	38.44	15.562
5	2.84	5.4	8.0656	29.16	15.336
6	2.65	6.5	7.0225	42.25	17.225
7	2.76	5.9	7.6176	34.81	16.284
Summe	18.55	42.7	49.2527	261.63	112.902

Mit Hilfe der Einträge in der letzten Zeile der o.a. Berechnungstabelle erhält man:

$$\overline{x} \;=\; \frac{1}{7}\sum_{i=1}^{7} x_i \;\overset{\text{Tab.}}{=}\; \frac{1}{7}\cdot 18.55 \;=\; 2.65\;, \tag{1}$$

$$\overline{y} \;=\; \frac{1}{7}\sum_{i=1}^{7} y_i \;\overset{\text{Tab.}}{=}\; \frac{1}{7}\cdot 42.7 \;=\; 6.1\;, \tag{2}$$

$$s_x^2 \;\overset{\text{A 3.32}}{=}\; \frac{1}{7}\sum_{i=1}^{7} x_i^2 - \overline{x}^2 \;\overset{\text{Tab.,(1)}}{=}\; \frac{1}{7}\cdot 49.2527 - 2.65^2 \;=\; 0.0136\;, \tag{3}$$

$$s_y^2 \;\overset{\text{A 3.32}}{=}\; \frac{1}{7}\sum_{i=1}^{7} y_i^2 - \overline{y}^2 \;\overset{\text{Tab.,(2)}}{=}\; \frac{1}{7}\cdot 261.63 - 6.1^2 \;\approx\; 0.1657\;, \tag{4}$$

$$\begin{aligned} s_{xy} \;&\overset{\text{A 7.25}}{=}\; \frac{1}{7}\sum_{i=1}^{7} x_i\,y_i - \overline{x}\,\overline{y} \;\overset{\text{Tab.,(1),(2)}}{=}\; \frac{1}{7}\cdot 112.902 - 2.65\cdot 6.1 \\ &\approx\; -0.0361\;. \end{aligned} \tag{5}$$

Hieraus ergeben sich gemäß A 8.4 bzw. D 7.4 die folgenden Kleinste-Quadrate-Schätzungen:

$$\hat{b} = \frac{s_{xy}}{s_x^2} \overset{(3),(5)}{\approx} \frac{-0.0361}{0.0136} = -2.654 \,, \tag{6}$$

$$\hat{a} = \overline{y} - \hat{b}\,\overline{x} \overset{(1),(2),(6)}{\approx} 6.1 + 2.654 \cdot 2.65 \approx 13.133 \,. \tag{7}$$

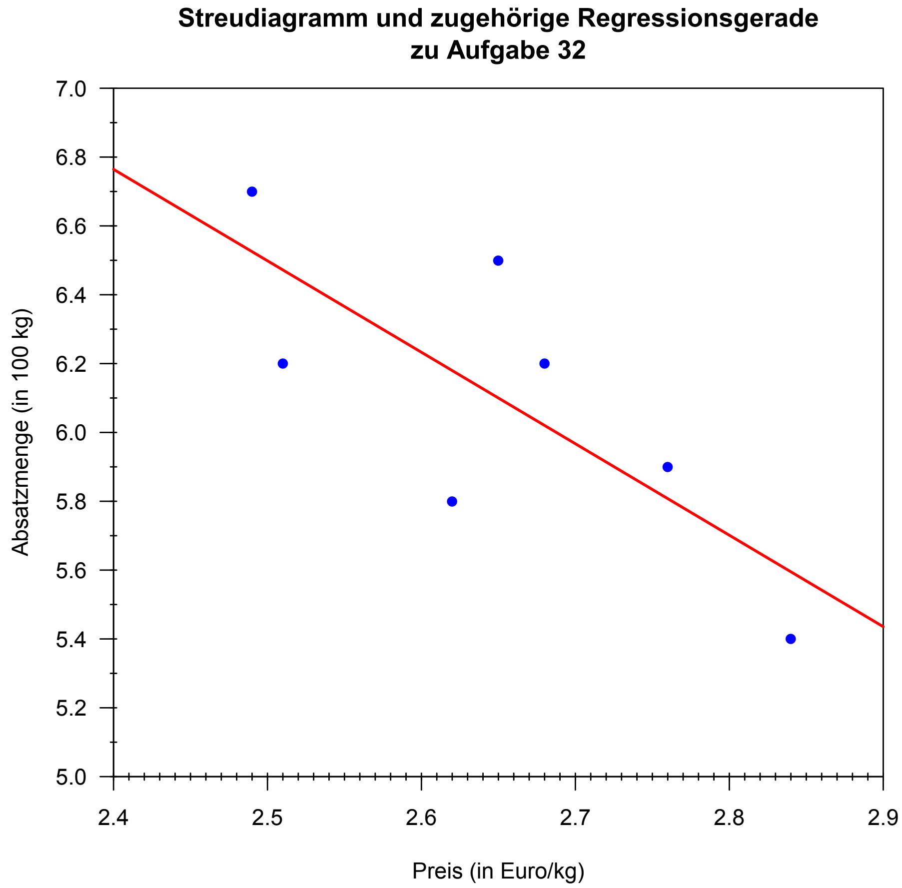
Die zugehörige *geschätzte* Regressionsgerade ist somit gegeben durch:

$$\hat{y}(x) = \hat{a} + \hat{b}\,x \overset{(6),(7)}{\approx} 13.133 - 2.654\,x \,. \tag{8}$$

Hiermit erhält man speziell für $x = 2.7$ (€/kg) die folgende geschätzte Absatzmenge (in 100 kg):

$$\hat{y}(2.7) \overset{(8)}{\approx} 13.133 - 2.654 \cdot 2.7 \approx 5.967 \,.$$

(b)



Wie man anhand der Grafik erkennen kann, gibt die Regressionsgerade den zugehörigen Trend an (höhere Angebotspreise korrespondieren mit geringeren Absatzmengen). Aufgrund der Streuung weichen die Datenpunkte jedoch teilweise deutlich von der Regressionsgeraden ab.

(c) Gemäß D 7.13 ist ein zweiseitiges Konfidenzintervall für die (unbekannte) Varianz σ^2 zum Konfidenzniveau $1 - \alpha$ gegeben durch

$$\mathcal{K}_{\sigma^2} = \left[\frac{(n-2)\,\hat{\sigma}^2}{\chi^2_{1-\alpha/2}(n-2)} \,, \frac{(n-2)\,\hat{\sigma}^2}{\chi^2_{\alpha/2}(n-2)} \right]$$

$$\text{mit} \quad \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \big(y_i - (\hat{a} + \hat{b}\,x_i)\big)^2$$

als (erwartungstreuer) Schätzung für σ^2 .

Weiter bezeichnen hierbei $\chi^2_{\alpha/2}(n-2)$ und $\chi^2_{1-\alpha/2}(n-2)$ das $\frac{\alpha}{2}$-Quantil bzw. $(1-\frac{\alpha}{2})$-Quantil der χ^2-Verteilung mit $n-2$ Freiheitsgraden.

Achtung: In der linearen Regression (Anpassung einer Geraden) jeweils $\mathbf{n-2}$ Freiheitsgrade!

Gemäß A 8.4 gilt folgende Darstellung:

$$\hat{\sigma}^2 \;=\; \frac{n}{n-2}\, s_y^2 \left(1 - r_{xy}^2\right) \;=\; \frac{n}{n-2} \left(s_y^2 - \frac{s_{xy}^2}{s_x^2} \right) . \tag{9}$$

Einsetzen der Zahlenwerte:

Mit $n = 7$ und $1-\alpha = 0.9 \iff \alpha = 0.1$ erhält man aus der Quantil-Tabelle der χ^2-Verteilung:

$$\chi^2_{\alpha/2}(n-2) \;=\; \chi^2_{0.05}(5) \overset{\text{Tab.}}{\approx} 1.15 \quad \text{und} \tag{10}$$

$$\chi^2_{1-\alpha/2}(n-2) \;=\; \chi^2_{0.95}(5) \overset{\text{Tab.}}{\approx} 11.07 \,. \tag{11}$$

Weiter folgt mit den Ergebnissen aus (a):

$$\begin{aligned} \hat{\sigma}^2 \;&\overset{(9)}{=}\; \frac{n}{n-2} \left(s_y^2 - \frac{s_{xy}^2}{s_x^2} \right) \\ &\overset{(3),(4),(5)}{\approx} \frac{7}{5} \cdot \left(0.1657 - \frac{(-0.0361)^2}{0.0136} \right) \approx 0.098 \,. \end{aligned} \tag{12}$$

Mit (10) – (12) erhält man das folgende zweiseitige 90%-Konfidenzintervall für σ^2:

$$\mathcal{K}_{\sigma^2} \;\approx\; \left[\frac{5 \cdot 0.098}{11.07} \,,\, \frac{5 \cdot 0.098}{1.15} \right] \;\approx\; \left[0.044 \,, 0.426 \right] .$$

(d) Gemäß D 7.12 ist ein zweiseitiges Konfidenzintervall für den (unbekannten) Regressionskoeffizienten b zum Konfidenzniveau $1-\alpha$ gegeben durch

$$\mathcal{K}_b \;=\; \left[\hat{b} - t_{1-\alpha/2}(n-2)\, \frac{\hat{\sigma}}{\sqrt{n\, s_x^2}} \,,\, \hat{b} + t_{1-\alpha/2}(n-2)\, \frac{\hat{\sigma}}{\sqrt{n\, s_x^2}} \right]$$

$$\text{mit} \quad \hat{\sigma} \;=\; \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} \left(y_i - (\hat{a} + \hat{b}\, x_i)\right)^2} \quad \text{(vgl. (c))} \,.$$

Weiter bezeichnet hierbei $t_{1-\alpha/2}(n-2)$ das $(1-\frac{\alpha}{2})$-Quantil der t-Verteilung mit $n-2$ Freiheitsgraden.

Achtung: In der linearen Regression (Anpassung einer Geraden) jeweils $\mathbf{n-2}$ Freiheitsgrade!

Einsetzen der Zahlenwerte:

Mit $n = 7$ und $1 - \alpha = 0.95 \iff \alpha = 0.05 \iff 1 - \alpha/2 = 0.975$ erhält man aus der Quantil-Tabelle der t-Verteilung:

$$t_{1-\alpha/2}(n-2) \;=\; t_{0.975}(5) \overset{\text{Tab.}}{\approx} 2.571\ . \tag{13}$$

Mit (3), (6), (12) und (13) erhält man schließlich das folgende zweiseitige 95%-Konfidenzintervall für b:

$$\begin{aligned}\mathcal{K}_b &\approx \left[\, -2.654 - 2.571 \cdot \frac{\sqrt{0.098}}{\sqrt{7 \cdot 0.0136}}\,,\; -2.654 + 2.571 \cdot \frac{\sqrt{0.098}}{\sqrt{7 \cdot 0.0136}} \,\right] \\ &\approx \left[\, -5.263\,,\; -0.045 \,\right]\ .\end{aligned}$$

Bemerkung:

Mit dem relativ geringen Stichprobenumfang $n = 7$, der bereits in Aufgabenteil (b) festgestellten hohen Streuung der Datenpunkte und dem relativ hohen geforderten Konfidenzniveau 95 % ergibt sich hier für den Steigungsparameter b ein recht großes Konfidenzintervall und damit (nur) eine sehr grobe Eingrenzung dieses Parameters.

Aufgabe 33

(a) Gemäß Voraussetzung und B 2.5 ist für $i \in \{1,\ldots,n\}$ und $\vartheta \in (0,\infty)$ die Zähldichte $f^{X_i\,|\,\theta=\vartheta} : \mathbb{N}_0 \longrightarrow [0,1]$ von X_i gegeben durch

$$f^{X_i\,|\,\theta=\vartheta}(k) \;=\; \frac{\vartheta^k}{k!}\, e^{-\vartheta}\ ,\ k \in \mathbb{N}_0\ . \tag{1}$$

Damit ist die gemeinsame Zähldichte $f^{X_1,\ldots,X_n\,|\,\theta=\vartheta} : \mathbb{N}_0^n \longrightarrow [0,1]$ von $X_1,\ldots,X_n$ für $\vartheta \in (0,\infty)$ gegeben durch

$$\begin{aligned} f^{X_1,\ldots,X_n\,|\,\theta=\vartheta}(x_1,\ldots,x_n) &\overset{\text{Unabh.}}{=} \prod_{i=1}^n f^{X_i\,|\,\theta=\vartheta}(x_i) \overset{(1)}{=} \prod_{i=1}^n \frac{\vartheta^{x_i}}{x_i!}\, e^{-\vartheta} \\ &= \left(\prod_{i=1}^n \frac{1}{x_i!}\right) \vartheta^{\sum_{i=1}^n x_i}\, e^{-n\vartheta} \;=\; \left(\prod_{i=1}^n \frac{1}{x_i!}\right) \vartheta^{n\,\overline{x}}\, e^{-n\vartheta} \end{aligned} \tag{2}$$

für $x = (x_1,\ldots,x_n) \in \mathbb{N}_0^n$.

Weiter ist nach Voraussetzung die a-priori Verteilung der Zufallsvariablen θ gegeben durch die Gamma-Verteilung $\Gamma(1,2)$. Gemäß B 3.9 ist damit eine Riemann-Dichte $f^\theta : \mathbb{R} \longrightarrow [0,\infty)$ von θ (die a-priori Dichte) gegeben durch

$$\begin{aligned} f^\theta(\vartheta) &= \begin{cases} \dfrac{1^2}{\Gamma(2)}\, \vartheta^{2-1}\, e^{-1\cdot\vartheta} & ,\quad \vartheta > 0\ , \\ 0 & ,\quad \vartheta \le 0\ . \end{cases} \\ &\overset{\Gamma(2)\,=\,1!\,=\,1}{=} \begin{cases} \vartheta\, e^{-\vartheta} & ,\quad \vartheta > 0\ , \\ 0 & ,\quad \vartheta \le 0\ . \end{cases} \end{aligned} \tag{3}$$

Gemäß der Bayes-Formel für Dichten (Formel (D.6) der Vorlesung auf Folie 557) ist dann für $x = (x_1,\ldots,x_n) \in \mathbb{N}_0^n$ eine Riemann-Dichte der zugehörigen a-posteriori Verteilung von θ (die a-posteriori Dichte) gegeben durch

$$f^{\theta\,|\,X_1,\dots,X_n}(\vartheta\,|\,x_1,\dots,x_n) \overset{\text{(D.6)}}{=} \frac{f^{X_1,\dots,X_n\,|\,\theta}(x_1,\dots,x_n\,|\,\vartheta)\cdot f^{\theta}(\vartheta)}{f^{X_1,\dots,X_n}(x_1,\dots,x_n)} \tag{4}$$

$$\overset{(2),(3)}{=} \frac{1}{f^{X_1,\dots,X_n}(x_1,\dots,x_n)}\left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)\vartheta^{n\,\overline{x}}\,e^{-n\vartheta}\,\vartheta\,e^{-\vartheta}\,\mathbb{1}_{(0,\infty)}(\vartheta)$$

$$= \underbrace{\frac{1}{f^{X_1,\dots,X_n}(x_1,\dots,x_n)}\left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)}_{\text{unabhängig von }\vartheta}\vartheta^{1+n\,\overline{x}}\,e^{-(1+n)\vartheta}\,\mathbb{1}_{(0,\infty)}(\vartheta)$$

$$\propto \vartheta^{2+n\,\overline{x}-1}\,e^{-(1+n)\vartheta}\,\mathbb{1}_{(0,\infty)}(\vartheta)\ ,\ \vartheta\in\mathbb{R}\ .$$

Der Vergleich der letzten Darstellung in (4) mit B 3.9 zeigt bereits, dass hierdurch – bis auf den Normierungsfaktor – eine Riemann-Dichte der Gamma-Verteilung

$$\Gamma\big(1+n, 2+n\,\overline{x}\big)$$

gegeben und damit die Aussage aus Aufgabenteil (a) bewiesen ist.

Dennoch wird nachfolgend der zugehörige Normierungsfaktor noch berechnet und damit dann die vollständige Darstellung der a-posteriori-Dichte $f^{\theta\,|\,X_1,\dots,X_n}$ angegeben.

Gemäß C 3.12 gilt für die Rand-Zähldichte $f^{X_1,\dots,X_n}:\ \mathbb{N}_0^n \longrightarrow [0,1]$ von $X_1,\dots,X_n$:

$$f^{X_1,\dots,X_n}(x_1,\dots,x_n) = \int_{-\infty}^{\infty} f^{X_1,\dots,X_n,\theta}(x_1,\dots,x_n,\vartheta)\,d\vartheta \tag{5}$$

$$\overset{\text{C 7.1}}{=} \int_{-\infty}^{\infty} f^{X_1,\dots,X_n\,|\,\theta}(x_1,\dots,x_n\,|\,\vartheta)\,f^{\theta}(\vartheta)\,d\vartheta \tag{6}$$

$$\overset{\text{vgl. (4)}}{=} \int_{-\infty}^{\infty} \underbrace{\left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)}_{\text{unabh. von }\vartheta}\vartheta^{1+n\,\overline{x}}\,e^{-(1+n)\vartheta}\,\mathbb{1}_{(0,\infty)}(\vartheta)\,d\vartheta$$

$$= \left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)\int_{0}^{\infty}\vartheta^{1+n\,\overline{x}}\,e^{-(1+n)\vartheta}\,d\vartheta$$

$$= \left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)\frac{\Gamma(2+n\,\overline{x})}{(1+n)^{2+n\,\overline{x}}}\underbrace{\int_{0}^{\infty}\underbrace{\frac{(1+n)^{2+n\,\overline{x}}}{\Gamma(2+n\,\overline{x})}\,\vartheta^{2+n\,\overline{x}-1}\,e^{-(1+n)\vartheta}}_{\substack{\text{Riemann-Dichte von }\Gamma(1+n,\,2+n\,\overline{x})\\ \text{an der Stelle }\vartheta}}\,d\vartheta}_{=1}$$

$$= \left(\prod_{i=1}^{n}\frac{1}{x_i!}\right)\frac{\Gamma(2+n\,\overline{x})}{(1+n)^{2+n\,\overline{x}}}$$

Hiermit folgt für $x = (x_1, \dots, x_n) \in \mathbb{N}_0^n$:

$$\begin{aligned}
f^{\theta \,|\, X_1,\dots,X_n}(\vartheta \,|\, x_1, \dots, x_n) &\overset{(4)}{=} \frac{1}{f^{X_1,\dots,X_n}(x_1, \dots, x_n)} \left(\prod_{i=1}^{n} \frac{1}{x_i!} \right) \vartheta^{1+n\,\overline{x}}\, e^{-(1+n)\vartheta}\, \mathbb{1}_{(0,\infty)}(\vartheta) \\
&\overset{(5)}{=} \frac{\left(\prod_{i=1}^{n} \frac{1}{x_i!} \right) \vartheta^{1+n\,\overline{x}}\, e^{-(1+n)\vartheta}\, \mathbb{1}_{(0,\infty)}(\vartheta)}{\left(\prod_{i=1}^{n} \frac{1}{x_i!} \right) \frac{\Gamma(2+n\,\overline{x})}{(1+n)^{2+n\,\overline{x}}}} \qquad (7) \\
&= \begin{cases} \dfrac{(1+n)^{2+n\,\overline{x}}}{\Gamma(2+n\,\overline{x})}\, \vartheta^{2+n\,\overline{x}-1}\, e^{-(1+n)\vartheta} & , \quad \vartheta > 0 \, , \\ 0 & , \quad \vartheta \le 0 \, . \end{cases}
\end{aligned}$$

Gemäß B 3.9 ist dies die Darstellung einer Riemann-Dichte zur Gamma-Verteilung $\Gamma\big(1+n, 2+n\,\overline{x}\big)$.

(b) Sei gemäß Aufgabenstellung nun $n = 9$ und $x = (x_1, \dots, x_9) \in \mathbb{N}_0^9$ eine Realisation von $X = (X_1, \dots, X_9)$ mit $\overline{x} = 2$.

Da hier die zugehörige quadratische Verlustfunktion zugrundegelegt wird, ist gemäß D 8.10 zu $x = (x_1, \dots, x_9)$ eine Bayes-Schätzung $\hat{\vartheta}_B(x)$ für den unbekannten Parameter $\vartheta \in (0,\infty)$ gegeben durch

$$\hat{\vartheta}_B(x) \;=\; \mathrm{E}\big(\theta \,\big|\, X = x\big) \, . \qquad (8)$$

Gemäß C 7.4 ist dies der bedingte Erwartungswert der Zufallsvariablen θ unter der Bedingung $X = x$, d.h. der Erwartungswert bzgl. der bedingten Verteilung der Zufallsvariablen θ unter der Bedingung $X = x$ und somit der Erwartungswert bzgl. der a-posteriori Verteilung

$$P^{\theta \,|\, X=x} \overset{(a)}{=} \Gamma\big(1+n, 2+n\,\overline{x}\big) \underset{\overline{x}=2}{\overset{n=9}{=}} \Gamma\big(10\,, 20\big) \qquad (9)$$

(wobei P das zugrundeliegende Wahrscheinlichkeitsmaß bezeichnet).

Gemäß C 5.2, (iv) ist (allgemein) der Erwartungswert einer Zufallsvariablen $Z \sim \Gamma(\alpha,\beta)$ mit $\alpha,\beta > 0$ gegeben durch

$$\mathrm{E}(Z) \;=\; \frac{\beta}{\alpha} \, . \qquad (10)$$

Damit erhält man insgesamt aus (7) – (9) die folgende Bayes-Schätzung für den unbekannten Parameter $\vartheta \in (0,\infty)$:

$$\hat{\vartheta}_B(x) \;=\; \frac{20}{10} \;=\; 2 \, , \qquad (11)$$

die hier mit der betreffenden Maximum-Likelihood-Schätzung

$$\hat{\vartheta}_{\mathrm{ML}} \;=\; \overline{x} \;=\; 2$$

übereinstimmt.