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

7. Juli 2022

Einführung in die angewandte Stochastik

6. Globalübung - Lösungen

Aufgabe 28

(a) Seien $x_1, \ldots, x_n \in (1, \infty)$ Realisationen der laut Aufgabenstellung gegebenen Zufallsvariablen $X_1, \ldots, X_n$.

Dann ist die zugehörige Likelihood-Funktion gemäß B 3.12 und D 3.2 gegeben durch:

$$\mathsf{L}\big(\alpha \,|\, x_1, \ldots, x_n\big) \overset{\text{Unabh.}}{=} \prod_{i=1}^{n} f_\alpha(x_i) \overset{\text{Def. } f_\alpha}{=} \prod_{i=1}^{n} \frac{\alpha}{x_i^{\alpha+1}} = \alpha^n \prod_{i=1}^{n} x_i^{-(\alpha+1)}, \quad \alpha \in (0, \infty). \tag{1}$$

Die zugehörige Log-Likelihood-Funktion ist gegeben durch:

$$\begin{aligned} \ell(\alpha) \;&=\; \ln\Big(L\big(\alpha \,|\, x_1, \ldots, x_n\big)\Big) \;\overset{(1)}{=}\; \ln\left(\alpha^n \prod_{i=1}^{n} x_i^{-(\alpha+1)}\right) \\ &=\; \ln\left(\alpha^n\right) + \sum_{i=1}^{n} \ln\left(x_i^{-(\alpha+1)}\right) \;=\; n\,\ln(\alpha) - (\alpha+1) \sum_{i=1}^{n} \ln(x_i), \quad \alpha \in (0, \infty). \end{aligned} \tag{2}$$

Gemäß (2) ist ℓ differenzierbar auf $(0, \infty)$ mit:

$$\begin{aligned} \ell'(\alpha) \;&= \frac{n}{\alpha} - \sum_{i=1}^{n} \ln(x_i) \;\left\{\begin{array}{c} > \\ = \\ < \end{array}\right\}\; 0 \\ &\iff \quad \frac{n}{\alpha} \left\{\begin{array}{c} > \\ = \\ < \end{array}\right\} \underbrace{\sum_{i=1}^{n} \ln(x_i)}_{>0 \text{ da } x_1, \ldots, x_n > 1} \quad \iff \quad \alpha \left\{\begin{array}{c} < \\ = \\ > \end{array}\right\} \underbrace{\frac{n}{\sum\limits_{i=1}^{n} \ln(x_i)}}_{=:\,\widehat{\alpha}}. \end{aligned} \tag{3}$$

Aus (3) folgt, dass ℓ streng monoton wachsend auf $(0, \widehat{\alpha}]$ und streng monoton fallend auf $[\widehat{\alpha}, \infty)$ ist.

Somit besitzt die Log-Likelihood-Funktion ℓ und damit auch die Likelihood-Funktion $\alpha \mapsto \mathsf{L}\big(\alpha \,|\, x_1, \ldots, x_n\big)$ selbst ein eindeutig bestimmtes globales Maximum in $\widehat{\alpha}$.

D.h.: Eine Maximum-Likelihood-Schätzung für den unbekannten Parameter α ist gegeben durch

$$\widehat{\alpha} = \frac{n}{\sum\limits_{i=1}^{n} \ln(x_i)}.$$

Ebenso wie in den Beispielen der Vorlesung ist die Maximum-Likelihood-Schätzung auch hier sogar eindeutig bestimmt, d.h. weitere Maximum-Likelihood-Schätzungen für α existieren nicht.

(b) Es bezeichnen $x_1, \ldots, x_{12}$ die $n = 12$ gegebenen Datenwerte. Einsetzen dieser Daten liefert folgenden Schätzwert für α:

$$\widehat{\alpha} \underset{n=12}{\overset{\text{(a)}}{=}} \frac{12}{\sum\limits_{i=1}^{12} \ln(x_i)} \approx 0{,}903\,.$$

(c) Mittels Ersetzen des unbekannten Parameters α durch den in (b) berechneten Schätzwert $\widehat{\alpha} \approx 0.903$ erhält man folgende Wahrscheinlichkeit:

$$P_{\widehat{\alpha}}(X_1 > 2) \;=\; 1 - P_{\widehat{\alpha}}(X_1 \leq 2) \;=\; 1 - F_{\widehat{\alpha}}(2) \overset{\text{Eins.}}{\approx} 1 - \left(1 - \frac{1}{2^{0.903}}\right) \approx 0{,}535\,.$$

Wirtschaftswissenschaftlicher Hintergrund zu Aufgabe 28

Der Schweizer Wirtschaftswissenschaftler *Vilfredo Pareto* formulierte 1897 die empirisch begründete These, dass der Anteil der Personen in einer Volkswirtschaft, die ein Einkommen von mindestens x haben, (ungefähr) $\big(\beta/x\big)^\alpha$ beträgt, wobei der Parameter β hierbei als Mindesteinkommen interpretiert wird und $\alpha > 0$ die sogenannte *Pareto-Konstante* ist.

Wird bei gegebenen Parametern α,β das Einkommen einer zufällig aus der Volkswirtschaft ausgewählten Person durch die Zufallsvariable X mit (stetiger) Verteilungsfunktion F beschrieben, gilt gemäß dieses Ansatzes:

$$P(X \geq x) = \begin{cases} \left(\dfrac{\beta}{x}\right)^\alpha, & x \geq \beta, \\ 1, & x < \beta, \end{cases} \iff F(x) = P(X \leq x) = \begin{cases} 1 - \left(\dfrac{\beta}{x}\right)^\alpha, & x \geq \beta, \\ 0, & x < \beta. \end{cases}$$

Normiert man das Mindesteinkommen zu 1, d.h. für $\beta = 1$ ergibt sich gerade die in der Aufgabenstellung zu Aufgabe 28 angegebene Pareto-Verteilung mit Verteilungsfunktion

$$F_\alpha(x) = \begin{cases} 1 - \dfrac{1}{x^\alpha}, & x \geq 1, \\ 0, & x < 1, \end{cases}$$

in Abhängigkeit vom (dann einzigen) Parameter $\alpha > 0$.

Vor diesem Hintergrund sind die in Aufgabe 28(b) angegebenen Daten als Vielfache des Mindesteinkommens zu interpretieren. Legt man beispielsweise ein Mindesteinkommen von 500 € zugrunde, ergeben sich aus diesen Daten die folgenden 12 Einkommenswerte (in €):

825 , 1485 , 1760 , 1535 , 1030 , 1205 , 1920 , 2740 , 930 , 2450 , 3680 , 790 .

In Aufgabe 28(c) wird (mittels der Parameterschätzung $\widehat{\alpha} = 0.903$) somit die Wahrscheinlichkeit berechnet, mit der das Einkommen einer zufällig ausgewählten Person das Doppelte des Mindesteinkommens übersteigt.

Aufgabe 29

Modell:

Die Auslastungen der Flüge können beschrieben werden durch stochastisch unabhängige, identisch verteilte Zufallsvariablen $X_1, \ldots, X_n$ mit folgender Interpretation für $i \in \{1, \ldots, n\}$:

$$X_i = \begin{cases} 1, & \text{falls der } i\text{-te Passagier den Flug } \textbf{nicht} \text{ antritt,} \\ 0, & \text{falls der } i\text{-te Passagier den Flug antritt.} \end{cases}$$

(Hierbei wird angenommen, dass die Entscheidungen der Passagiere, den Flug anzutreten/nicht anzutreten, unabhängig voneinander erfolgen.)

Mit dieser Festlegung der Zufallsvariablen $X_1, \ldots, X_n$ gilt dann:

$$X_i \sim \mathrm{bin}(1, p) \quad \text{für} \quad i \in \{1, \ldots, n\},$$

wobei $p \in (0, 1)$ die unbekannte Wahrscheinlichkeit bezeichnet, mit der ein zufällig ausgewählter Passagier seinen Flug **nicht** antritt.

Gemäß D 4.5 ist in dieser Situation ein zweiseitiges (symmetrisches) *approximatives* Konfidenzintervall für den Parameter p zum Konfidenzniveau $1 - \alpha$ gegeben durch:

$$K = \left[\widehat{p} - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\widehat{p}\,(1 - \widehat{p})} \,,\, \widehat{p} + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\widehat{p}\,(1 - \widehat{p})} \right].$$

Hierbei bezeichnen (mit $x_1, \ldots, x_n$ als Realisationen von $X_1, \ldots, X_n$)

$$\widehat{p} = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

die (erwartungstreue) Maximum-Likelihood-Schätzung für den unbekannten Parameter p und $u_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$ das $(1 - \frac{\alpha}{2})$-Quantil zu $\mathrm{N}(0, 1)$.

Einsetzen der Zahlenwerte:

Mit $1 - \alpha = 0{,}9 \iff \alpha = 0{,}1 \iff 1 - \frac{\alpha}{2} = 0{,}95$ erhält man aus der Quantiltabelle der Standardnormalverteilung:

$$u_{1-\frac{\alpha}{2}} = u_{0{,}95} = \Phi^{-1}(0{,}95) \overset{\text{Tab.}}{\approx} 1{,}64\,.$$

Dies ergibt zusammen mit $n = 1000$ und

$$\widehat{p} = \overline{x} \overset{\text{Vor.}}{=} \frac{74}{1000} = 0{,}074$$

das folgende zweiseitige approximative 90%-Konfidenzintervall für p:

$$K \approx \left[0{,}074 - 1{,}64 \cdot \sqrt{\frac{0{,}074 \cdot 0{,}926}{1000}} \,,\, 0{,}074 + 1{,}64 \cdot \sqrt{\frac{0{,}074 \cdot 0{,}926}{1000}} \right] \approx \big[0{,}060 \,, 0{,}088\big].$$

Aufgabe 30

(a) Gemäß D 5.8 ist unter der vorausgesetzten Normalverteilungsannahme mit *unbekannter* Varianz σ^2 ein zweiseitiges (symmetrisches) Konfidenzintervall für den Erwartungswert μ zum Konfidenzniveau $1-\alpha$ gegeben durch:

$$K_1 = \left[\overline{x} - t_{1-\frac{\alpha}{2}}(n-1)\,\frac{\widehat{\sigma}}{\sqrt{n}}\,,\ \overline{x} + t_{1-\frac{\alpha}{2}}(n-1)\,\frac{\widehat{\sigma}}{\sqrt{n}}\right].$$

Hierbei bezeichnen

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2\right)$$

die Stichprobenvarianz zu den Daten $x_1,\dots,x_n$ und $t_{1-\frac{\alpha}{2}}(n-1)$ das $(1-\frac{\alpha}{2})$-Quantil zur t-Verteilung mit $n-1$ Freiheitsgraden.

Einsetzen der Zahlenwerte:

Mit $n = 12$ und $1-\alpha = 0{,}9 \iff \alpha = 0{,}1 \iff 1-\frac{\alpha}{2} = 0{,}95$ erhält man aus der Quantiltabelle der t-Verteilung:

$$t_{1-\frac{\alpha}{2}}(n-1) = t_{0{,}95}(11) \overset{\text{Tab.}}{\approx} 1{,}796\,.$$

Weiter berechnet man aus den gegebenen Daten $x_1,\dots,x_{12}$:

$$\overline{x} = \frac{1}{12}\sum_{i=1}^{12} x_i \approx 3{,}528 \quad\text{und}\quad \widehat{\sigma} = \sqrt{\frac{1}{11}\left(\sum_{i=1}^{12} x_i^2 - 12\cdot\overline{x}^2\right)} \approx 0{,}185\,.$$

Hiermit erhält man das folgende zweiseitige 90%-Konfidenzintervall für μ bei *unbekannter* Varianz σ^2:

$$K_1 \approx \left[3{,}528 - 1{,}796\cdot\frac{0{,}185}{\sqrt{12}}\,,\ 3{,}528 + 1{,}796\cdot\frac{0{,}185}{\sqrt{12}}\right] \approx \left[3{,}432\,,\ 3{,}624\right].$$

(b) Gemäß D 5.6 ist unter der vorausgesetzten Normalverteilungsannahme mit *bekannter* Varianz σ^2 ein zweiseitiges (symmetrisches) Konfidenzintervall für den Erwartungswert μ zum Konfidenzniveau $1-\alpha$ gegeben durch:

$$K_2 = \left[\overline{x} - u_{1-\frac{\alpha}{2}}\,\frac{\sigma}{\sqrt{n}}\,,\ \overline{x} + u_{1-\frac{\alpha}{2}}\,\frac{\sigma}{\sqrt{n}}\right].$$

Hierbei bezeichnet $u_{1-\frac{\alpha}{2}} = \Phi^{-1}(1-\frac{\alpha}{2})$ das $(1-\frac{\alpha}{2})$-Quantil zur Standardnormalverteilung $\mathrm{N}(0,1)$.

Einsetzen der Zahlenwerte:

Mit $1 - \frac{\alpha}{2} \overset{(a)}{=} 0{,}95$ erhält man aus der Quantiltabelle der Standardnormalverteilung:

$$u_{1-\frac{\alpha}{2}} = u_{0{,}95} = \Phi^{-1}(0{,}95) \overset{\text{Tab.}}{\approx} 1{,}64\,.$$

Dies ergibt zusammen mit $n = 12$, $\overline{x} \overset{(a)}{\approx} 3{,}528$ und

$$\sigma = \sqrt{\sigma^2} \overset{\text{Vor.}}{=} \sqrt{0{,}0324} = 0{,}18$$

bei *bekannter* Varianz $\sigma^2 = 0{,}0324$ das folgende zweiseitige 90%-Konfidenzintervall für μ:

$$K_2 \approx \left[3{,}528 - 1{,}64 \cdot \frac{0{,}18}{\sqrt{12}}\,,\ 3{,}528 + 1{,}64 \cdot \frac{0{,}18}{\sqrt{12}}\right] \approx \left[3{,}443\,,\ 3{,}613\right].$$

Bemerkung zu Aufgabe 30 (a) und (b)

Da hier $\widehat{\sigma} \approx \sigma$ und $t_{0{,}95}(11) \approx u_{0{,}95}$ gilt, unterscheiden sich die beiden berechneten Konfidenzintervalle K_1 und K_2 nur geringfügig.

(c) Wie bereits in (b) angegeben, ist unter der vorausgesetzten Normalverteilungsannahme mit *bekannter* Varianz σ^2 ein zweiseitiges (symmetrisches) Konfidenzintervall für den Erwartungswert μ zum Konfidenzniveau $1 - \alpha$ für $n \in \mathbb{N}$ Beobachtungen gegeben durch:

$$K(n) = \left[\overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\,,\ \overline{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]. \tag{1}$$

Sei $l > 0$ eine vorgegebene maximale Intervalllänge. Dann gilt für die Intervalllänge $L(K(n))$ des Konfidenzintervalls $K(n)$:

$$\begin{aligned} L(K(n)) &\overset{(1)}{=} \overline{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \left(\overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 2\,u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le l \\ &\iff \quad \sqrt{n} \ge \frac{2\,u_{1-\frac{\alpha}{2}}\,\sigma}{l} \\ &\iff \quad n \ge \left(\frac{2\,u_{1-\frac{\alpha}{2}}\,\sigma}{l}\right)^2 = \frac{4\,\sigma^2\,u^2_{1-\frac{\alpha}{2}}}{l^2}. \end{aligned} \tag{2}$$

Einsetzen der Zahlenwerte:

Mit $l = 0{,}1$, $\sigma^2 = 0{,}0324$ und $u_{1-\frac{\alpha}{2}} = u_{0{,}95} \overset{(b)}{\approx} 1{,}64$ erhält man aus (2):

$$n \ge \frac{4 \cdot 0{,}0324 \cdot 1{,}64^2}{0{,}1^2} \approx 34{,}86\,.$$

Wegen $n \in \mathbb{N}$ müssen somit mindestens $n = 35$ Versuchsflächen vorhanden sein, um bei bekannter Varianz $\sigma^2 = 0{,}0324$ ein zweiseitiges 90%-Konfidenzintervall für den Erwartungswert μ mit einer maximalen Länge von $l = 0{,}1$ (t) angeben zu können.

Aufgabe 31

Gemäß D 5.13 ist ein zweiseitiges (symmetrisches) Konfidenzintervall für $d = \mu_1 - \mu_2$ zum Konfidenzniveau $1-\alpha$ bei unbekannter (gleicher) Varianz σ^2 gegeben durch

$$K = \left[\widehat{\Delta} - t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\,\widehat{\sigma}_{\text{pool}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,,\ \widehat{\Delta} + t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\,\widehat{\sigma}_{\text{pool}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right].$$

Hierbei bezeichnen $\widehat{\Delta} = \overline{x} - \overline{y}$ die erwartungstreue Punktschätzung für $d = \mu_1 - \mu_2$,

$$\begin{aligned}
\widehat{\sigma}^2_{\text{pool}} &= \frac{1}{n_1+n_2-2}\left(\underbrace{\sum_{i=1}^{n_1}(x_i-\overline{x})^2}_{=(n_1-1)\,\widehat{\sigma}_1^2} + \underbrace{\sum_{j=1}^{n_2}(y_j-\overline{y})^2}_{=(n_2-1)\,\widehat{\sigma}_2^2} \right) \\
&= \frac{(n_1-1)\,\widehat{\sigma}_1^2 + (n_2-1)\,\widehat{\sigma}_2^2}{n_1+n_2-2}
\end{aligned}$$

den erwartungstreuen Punktschätzer für σ^2 kombiniert aus beiden vorliegenden Stichproben und $t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$ das $(1-\frac{\alpha}{2})$-Quantil der t-Verteilung mit n_1+n_2-2 Freiheitsgraden.

Einsetzen der Zahlenwerte:

Mit $n_1 = 12\,,\ n_2 = 8,\ 1-\alpha = 0{,}95 \iff 1-\frac{\alpha}{2} = 0{,}975$ erhält man aus der Quantiltabelle der t-Verteilung:

$$t_{1-\frac{\alpha}{2}}(n_1+n_2-2) = t_{0{,}975}(18) \overset{\text{Tab.}}{\approx} 2{,}101\ .$$

Aus den Messwerten $x_1,\ldots,x_{12}$ und $y_1,\ldots,y_8$ berechnet man:

$$\overline{x} = \frac{1}{12}\sum_{i=1}^{12} x_i = 50{,}4\,,\quad \overline{y} = \frac{1}{8}\sum_{i=1}^{8} y_i = 50{,}6\,,\quad \widehat{\Delta} = \overline{x}-\overline{y} = -0{,}2\,,$$

$$\widehat{\sigma}_1^2 = \frac{1}{11}\sum_{i=1}^{12}(x_i-\overline{x})^2 \approx 4{,}433\,,\quad \widehat{\sigma}_2^2 = \frac{1}{7}\sum_{j=1}^{8}(y_j-\overline{y})^2 \approx 3{,}326\,,$$

$$\widehat{\sigma}_{\text{pool}} \approx \sqrt{\frac{11\cdot 4{,}433 + 7\cdot 3{,}326}{18}} \approx 2{,}0006\,.$$

Hiermit erhält man:

$$\begin{aligned}
K &\approx \left[-0{,}2 - 2{,}101\cdot 2{,}0006\cdot\sqrt{\frac{1}{12}+\frac{1}{8}}\,,\ -0{,}2 + 2{,}101\cdot 2{,}0006\cdot\sqrt{\frac{1}{12}+\frac{1}{8}} \right] \\
&\approx \big[-2{,}119\,,\ 1{,}719 \big].
\end{aligned}$$