Aufgaben und Lösungsvorschlag

1. Multiple choice **[7 Punkte]**

(a) **[1 Punkt]** Gilt $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$? (Alle Antworten sind so zu verstehen, dass nur Zufallsvariablen X, Y betrachtet werden, deren Erwartungswert existiert und endlich ist.)

i. Ja, $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ gilt für alle Zufallsvariablen X, Y.
ii. $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ gilt für alle unabhängigen Zufallsvariablen X, Y. Es gibt nicht unabhängige Zufallsvariablen für die $\mathbb{E}[XY] \neq \mathbb{E}[X]\,\mathbb{E}[Y] \in \mathbb{R}$ gilt.
iii. Nein, es gibt unabhängige sowie nicht unabhängige Zufallsvariablen für die $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y] \in \mathbb{R}$ gilt.
iv. Nein, für alle Zufallsvariablen X, Y gilt $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$.

Lösung:

1.(a)ii: Theorem 3.14 aus dem Skript beweist **1.**(a)ii im Fall von stetigen Zufallsvariablen, die Aussage gilt aber auch allgemeiner (auch für diskrete Zufallsvariablen) solange $\mathbb{E}[X] \in \mathbb{R}$ und $\mathbb{E}[Y] \in \mathbb{R}$ wohldefiniert und endlich sind.

(b) **[1 Punkt]** Was bedeutet es, wenn ein statistischer Test ein Signifikanz-Niveau (relevance level) α hat?

i. Die Nullhypothese H_0 ist mit einer Wahrscheinlichkeit von α falsch, wenn der Test die Nullhypotehse verwirft.
ii. Wenn die Nullhypothese H_0 wahr wäre, dann würde mit der Wahrscheinlichkeit von α die Nullhypothese vom Test verworfen werden.
iii. Die Alternativhypothese H_1 ist mit einer Wahrscheinlichkeit von α falsch, wenn der Test die Nullhypotehse verwirft.
iv. Wenn die Alternativhypothese H_1 wahr wäre, dann würde mit der Wahrscheinlichkeit von α die Nullhypothese vom Test verworfen werden.

Lösung:

1.(b)ii

(c) **[2.5 Punkte]** Sei $X_1, X_2, \ldots$ eine Folge unabhängiger Bernoulli Zufallsvariablen, wobei X_t eine Erfolgswahrscheinlichkeit $\mathbb{P}[X_t = 1] = 1 - e^{-\frac{1}{2^t}}$ hat. Wie hoch ist die Wahrscheinlichkeit, dass niemals ein Erfolg erzielt wird?

i. 0
ii. e^{-4}
iii. 2^{-e}
iv. e^{-1}
v. $1 - e^{-\frac{1}{2}}$
vi. $\log(2)$
vii. 1
viii. ∞

Lösung:

1.(c)iv: $\mathbb{P}[\forall t \in \mathbb{N} : X_t \neq 1] = \mathbb{P}\left[\bigcap_{t\in\mathbb{N}}\{X_t \neq 1\}\right]$. Wegen der Stetigkeit des Masses (Proposition 1.12 im Skript) und weil $B_n := \bigcap_{t\in\{1,\ldots,n\}}\{X_t \neq 1\} \supseteq \bigcap_{t\in\{1,\ldots,n+1\}}\{X_t \neq 1\}$, kann die Wahrscheinlichkeit des Durchschnitts durch den Limes der Wahrscheinlichkeiten ersetzt werden:

$$\mathbb{P}\left[\bigcap_{t\in\mathbb{N}}\{X_t \neq 1\}\right] = \lim_{n\to\infty} \mathbb{P}\left[\bigcap_{t\in\{1,\ldots,n\}}\{X_t \neq 1\}\right].$$

Jetzt kann die Unabhängigkeit der X_i verwendet werden um die Wahrscheinlichkeit des endlichen Durchschnitts zu berechnen:

$$\begin{aligned}\mathbb{P}\left[\bigcap_{t\in\{1,\ldots,n\}}\{X_t \neq 1\}\right] &= \prod_{t\in\{1,\ldots,n\}} \mathbb{P}[\{X_t \neq 1\}] \\ &= \prod_{t\in\{1,\ldots,n\}} e^{-\frac{1}{2^t}} \\ &= e^{\sum_{t\in\{1,\ldots,n\}} -\frac{1}{2^t}} \\ &= e^{-\sum_{t\in\{1,\ldots,n\}} \frac{1}{2^t}}.\end{aligned}$$

Wenn wir all diese Resultate kombinieren erhalten wir

$$\begin{aligned}\mathbb{P}\left[\bigcap_{t\in\mathbb{N}}\{X_t \neq 1\}\right] &= \lim_{n\to\infty} e^{-\sum_{t\in\{1,\ldots,n\}} \frac{1}{2^t}} \\ &= e^{-\lim_{n\to\infty}\sum_{t\in\{1,\ldots,n\}} \frac{1}{2^t}} \\ &= e^{-\sum_{t=1}^{\infty}\left(\frac{1}{2}\right)^t} \\ &= e^{-\frac{\frac{1}{2}}{1-\frac{1}{2}}} \\ &= e^{-1}.\end{aligned}$$

(d) **[2.5 Punkte]** Sei $Y_1, Y_2, \ldots, Y_n$ eine Folge von i.i.d. Messungen einer unbekannten Größe m. Die Verteilung einer Messung ist $\mathcal{N}(m, \sigma^2)$ (weil wir annehmen, dass der Messfehler $\mathcal{N}(0, \sigma^2)$ verteilt ist) mit bekanntem $\sigma = 0.1$. Wir betrachten das Konfidenzintervall $I = [\frac{1}{n}\sum_{i=1}^n Y_i - a, \frac{1}{n}\sum_{i=1}^n Y_i + a]$. Wähle das grösste z sodass I ein z%-Konfidenzintervall ist.

i. $100\left(2\Phi\left(\frac{a}{\sqrt{n}\sigma}\right) - 1\right)$
ii. $100\left(2\Phi\left(\frac{\sigma a}{\sqrt{n}}\right) - 1\right)$
iii. $100\left(2\Phi\left(\frac{\sqrt{n}a}{\sigma}\right) - 1\right)$
iv. $100\left(2\Phi\left(a\sqrt{n}\sigma\right) - 1\right)$

Lösung:

1.(d)iii: Wir verwenden die Kurzschreibweise $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
Aufgrund der „Properties of normal random variables“ auf Seite 50 des Skripts erhalten wir $\bar{Y}_n \sim \mathcal{N}(m, \sigma^2/n)$ und somit $Z := \frac{\sqrt{n}(\bar{Y}_n - m)}{\sigma} \sim \mathcal{N}(0,1)$.
$\mathbb{P}\left[\bar{Y}_n - a \leq m\right] = \mathbb{P}\left[\bar{Y}_n \leq m + a\right] = \mathbb{P}\left[\bar{Y}_n - m \leq a\right] = \mathbb{P}\left[\frac{\sqrt{n}(\bar{Y}_n - m)}{\sigma} \leq \frac{\sqrt{n}a}{\sigma}\right] = \mathbb{P}\left[Z \leq \frac{\sqrt{n}a}{\sigma}\right] = \Phi\left(\frac{\sqrt{n}a}{\sigma}\right)$ und somit $\mathbb{P}\left[\bar{Y}_n - a > m\right] = 1 - \Phi\left(\frac{\sqrt{n}a}{\sigma}\right)$. Analog erhält man $\mathbb{P}\left[\bar{Y}_n + a \geq m\right] = \mathbb{P}\left[\bar{Y}_n \geq m - a\right] = \mathbb{P}\left[\bar{Y}_n - m \geq -a\right] = \mathbb{P}\left[\frac{\sqrt{n}(\bar{Y}_n - m)}{\sigma} \geq \frac{-\sqrt{n}a}{\sigma}\right] = \mathbb{P}\left[Z \geq \frac{-\sqrt{n}a}{\sigma}\right] = 1 - \Phi\left(-\frac{\sqrt{n}a}{\sigma}\right) = \Phi\left(\frac{\sqrt{n}a}{\sigma}\right)$, wobei die letzten beiden Schritte wegen Stetigkeit beziehungsweise Symmetrie gelten. Somit gilt $\mathbb{P}\left[\bar{Y}_n + a < m\right] = 1 - \Phi\left(\frac{\sqrt{n}a}{\sigma}\right)$.
$\mathbb{P}\left[m \in [\bar{Y}_n - a, \bar{Y}_n + a]\right] = \frac{z}{100}$ ist äquivalent zu $\frac{z}{100} = 1 - \mathbb{P}\left[m \notin [\bar{Y}_n - a, \bar{Y}_n + a]\right]$. Da die Ereignisse $\left\{\bar{Y}_n - a > m\right\}$ und $\left\{\bar{Y}_n + a < m\right\}$ disjunkt sind können wir die Wahrscheinlichkeit der Vereinigung als Summe der Wahrscheinlichkeiten schreiben: $\frac{z}{100} = 1 - \mathbb{P}\left[\left\{\bar{Y}_n - a > m\right\} \cup \left\{\bar{Y}_n + a < m\right\}\right] = 1 - \left(\mathbb{P}\left[\bar{Y}_n - a > m\right] + \mathbb{P}\left[\bar{Y}_n + a < m\right]\right) = 1 - 2\left(1 - \Phi\left(\frac{\sqrt{n}a}{\sigma}\right)\right) = 2\Phi\left(\frac{\sqrt{n}a}{\sigma}\right) - 1$.

2. Arbeitsweg **[8 Punkte]**

Frau Huber fährt jeden Tag mit Wahrscheinlichkeit $\frac{1}{3}$ mit dem Velo und mit Wahrscheinlichkeit $\frac{2}{3}$ mit dem Bus zur Arbeit. Nimmt sie das Velo, so ist ihre (zufällige) Fahrzeit in Minuten gleichverteilt auf $[1, 3]$; nimmt sie den Bus, so ist sie gleichverteilt auf $[2, 6]$. Wir bezeichnen die Fahrzeit von Frau Huber in Minuten mit T.

(a) [**3 Punkte**] Gegeben, dass T zwischen 2 und 5 liegt, mit welcher Wahrscheinlichkeit ist Frau Huber dann mit dem Velo gefahren?

(b) [**1 Punkt**] Gegeben, dass T höchstens 1.5 beträgt, mit welcher Wahrscheinlichkeit ist Frau Huber dann mit dem Bus gefahren?

(c) [**4 Punkte**] Berechnen Sie die Verteilungsfunktion von T.

Lösung:

Wir definieren die Ereignisse

$$\begin{aligned} V &:= \{\text{Huber fährt mit dem Velo}\},\\ B &:= \{\text{Huber fährt mit dem Bus}\}. \end{aligned}$$

(a) Nach der Formel von Bayes gilt

$$\begin{aligned} \mathbb{P}\left[V \mid T \in [2,5]\right] &= \frac{\mathbb{P}\left[T \in [2,5] \cap V\right]}{\mathbb{P}\left[T \in [2,5]\right]}\\ &= \frac{\mathbb{P}\left[T \in [2,5] \mid V\right]\mathbb{P}\left[V\right]}{\mathbb{P}\left[T \in [2,5] \mid V\right]\mathbb{P}\left[V\right] + \mathbb{P}\left[T \in [2,5] \mid B\right]\mathbb{P}\left[B\right]}\\ &= \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3} + \frac{3}{4}\cdot\frac{2}{3}} = \frac{\frac{1}{6}}{\frac{1}{6}+\frac{3}{6}} = \frac{\frac{1}{6}}{\frac{4}{6}} = \frac{1}{4}. \end{aligned}$$

(b) Falls $T \leq 1.5$, dann kann Frau Huber nicht mit dem Bus gefahren sein, da in diesem Fall $T \in [2, 6]$ gilt. Also $\mathbb{P}\left[B \mid T \leq 1.5\right] = 0$.

(c) Nach der Formel von der totalen Wahrscheinlichkeit gilt für $x \in \mathbb{R}$

$$\mathbb{P}\left[T \leq x\right] = \mathbb{P}\left[T \leq x \mid V\right]\mathbb{P}\left[V\right] + \mathbb{P}\left[T \leq x \mid B\right]\mathbb{P}\left[B\right] =$$

$$= \begin{cases} 0 + 0 & = 0 \qquad (x < 1)\\ \frac{1}{3}\cdot\frac{1}{2}(x-1) & = \frac{1}{6}(x-1) \qquad (x \in [1,2))\\ \frac{1}{3}\cdot\frac{1}{2}(x-1) + \frac{2}{3}\cdot\frac{1}{4}(x-2) & = \frac{1}{3}x - \frac{1}{2} \qquad (x \in [2,3))\\ \frac{1}{3}\cdot 1 + \frac{2}{3}\cdot\frac{1}{4}(x-2) & = \frac{1}{6}x \qquad (x \in [3,6))\\ \frac{1}{3} + \frac{2}{3} & = 1 \qquad (x \geq 6). \end{cases}$$

3. T **und** U **[7 Punkte]**

Seien T und U zwei unabhängige Zufallsvariablen, wobei T exponentialverteilt ist mit Parameter $\lambda = 2$ (somit ist die Dichte $f_T(t) = \lambda e^{-\lambda t}$) und U gleichverteilt ist auf dem Intervall $[0, 2]$. Wir definieren $X := 2T - \frac{1}{2}U - 3$.

(a) **[3 Punkte]** Berechnen Sie den Erwartungswert und die Varianz von X.

(b) **[4 Punkte]** Welche Wahrscheinlichkeit hat das Ereignis $\{X \leq -3\}$?

Lösung:

(a)

$$\mathbb{E}[X] = 2\mathbb{E}[T] - \frac{1}{2}\mathbb{E}[U] - \mathbb{E}[3] = 2 \times \frac{1}{2} - \frac{1}{2} \times 1 - 3 = -\frac{5}{2} = -2.5$$

$$\sigma_X^2 = 2^2\sigma_T^2 + \left(-\frac{1}{2}\right)^2 \sigma_U^2 + \sigma_{-3}^2 = 2^2\frac{1}{2^2} + \frac{1}{2^2}\frac{2^2}{12} + 0 = 1 + \frac{1}{12} = \frac{13}{12}.$$

(b)

$$\begin{aligned}
\mathbb{P}[X \leq -3] &= \mathbb{P}\left[2T - \frac{1}{2}U - 3 \leq -3\right] \\
&= \mathbb{P}[T \leq U/4] \\
&= \int_0^2 \int_0^{u/4} 2e^{-2t} dt \, \frac{1}{2} du \\
&= \int_0^2 \mathbb{P}[T \leq u/4] \frac{1}{2} du \\
&= \int_0^2 (1 - e^{-2(u/4)}) \frac{1}{2} du \\
&= \int_0^2 (1 - e^{-u/2}) \frac{1}{2} du \\
&= 1 + \frac{1}{2}\left(2e^{-u/2}\Big|_0^2\right) \\
&= 1 + e^{-1} - e^{-0/2} = e^{-1} = \frac{1}{e}.
\end{aligned}$$

Für ein allgemeines $\lambda \in \mathbb{R}$ wäre die Lösung: $1 - 2/\lambda + 2/\lambda \cdot e^{-\lambda/2}$.

4. Einkommen in Zürich **[5 Punkte]**

Die Zufallsvariable X gibt das Einkommen eines zufällig ausgewählten Einwohners Zürich an. Zur Modellierung von X können wir eine sogenannte verallgemeinerte Pareto-Verteilung mit Dichte

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta}(1+x)^{-(1+\frac{1}{\theta})}, & \text{falls } x > 0, \\ 0, & \text{falls } x \leq 0 \end{cases}$$

verwenden. Dabei ist $\theta > 0$ ein unbekannter Parameter, der auf der Basis von Daten $x_1, \ldots, x_n \in \mathbb{R}_{>0}$ (die das Einkommen von n zufällig ausgewählten Einwohnern angeben) geschätzt werden soll; diese Daten werden wie üblich als Realisierungen von Zufallsvariablen $X_1, \ldots, X_n$ aufgefasst, die unter $\mathbb{P}_\theta$ i.i.d. mit Dichte $f_X(x;\theta)$ sind, für jede Wahl des Parameters θ.

(a) Sei $\mathbf{X} = (X_1, \ldots, X_n)$. Berechne den Maximum-Likelihood-Schätzer für θ.

Hinweis 4.1: Erinnere dich, dass $\arg\max_{\theta\in\mathbb{R}_{>0}} h(g(\theta)) = \arg\max_{\theta\in\mathbb{R}_{>0}} g(\theta)$ für jede streng monoton steigende Funktion h, wie zum Beispiel log.

Lösung:

(a) Die log-Likelihood-Funktion ist gegeben durch

$$\begin{aligned} \log L(x_1, \ldots, x_n; \theta) &= \log\left(\frac{1}{\theta^n}\prod_{i=1}^n (1+x_i)^{-(1+\frac{1}{\theta})}\right) \\ &= -n\log\theta - (1+\frac{1}{\theta})\sum_{i=1}^n \log(1+x_i). \end{aligned}$$

Die Ableitung nach θ ist

$$\frac{\partial}{\partial\theta}\log L(x_1, \ldots, x_n; \theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n \log(1+x_i),$$

und das ist 0 für

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^n \log(1+x_i).$$

Um sicher zu gehen, dass es sich bei $\hat{\theta}$ um ein Maximum handelt betrachten wir die zweite Ableitung:

$$\frac{\partial^2}{\partial\theta^2}\log L(x_1, \ldots, x_n; \theta) = +\frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n \log(1+x_i).$$

Wenn wir $\hat{\theta}$ einsetzen erhalten wir

$$\frac{\partial^2}{\partial\theta^2}\log L(x_1, \ldots, x_n; \theta)\big|_{\theta=\hat{\theta}} = +\frac{n}{\hat{\theta}^2} - \frac{2}{\hat{\theta}^3}n\hat{\theta} = -\frac{n}{\hat{\theta}^2} < 0.$$

Somit maximiert $\hat{\theta}$ tatsächlich $\log L$ und somit auch L, weil der log streng monoton steigend ist. Der ML-Schätzer für θ ist also gegeben durch

$$T_n = \frac{1}{n}\sum_{i=1}^{n} \log(1 + X_i).$$

5. Rote und schwarze Kugeln **[3 Punkte]**

Wir betrachten eine Urne mit 20 roten und 80 schwarzen Kugeln. Wir ziehen 3 Mal mit zurücklegen (nach jedem Zug wird die gezogene Kugel zurück in die Urne gelegt und die Kugeln gemischt bevor erneut gezogen wird).

(a) [**2 Punkte**] Definiere den einfachsten (kleinsten) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ der dieses Experiment beschreibt, wobei wir nur an den Farben der Kugeln interessiert sind und in welcher Reihenfolge die Farben gezogen werden.

(b) [**0.5 Punkte**] Wie viele Elemente hat Ω?

(c) [**0.5 Punkte**] Wie viele Elemente hat $\mathcal{F}$?

Lösung:

(a) i. $\Omega := \{0, 1\}^3$, wobei 0 für rot und 1 für schwarz steht.

ii. $\mathcal{F} := \mathcal{P}(\Omega)$

iii. $\mathbb{P} : \mathcal{F} \to [0, 1], A \mapsto \mathbb{P}[A] := \sum_{\omega \in A} \prod_{i=1}^{3} 0.2^{1-\omega_i} 0.8^{\omega_i}$

Alternative falsche Lösung:

i. $\tilde{\Omega} := \{1, 2, 3, \ldots, 100\}^3$, wobei die die Zahlen 1 bis 20 den 20 roten Kugeln und die Zahlen 21 bis 100 den 80 schwarzen Kugeln entsprechen.

ii. $\tilde{\mathcal{F}} := \mathcal{P}\left(\tilde{\Omega}\right)$

iii. $\tilde{\mathbb{P}} : \mathcal{F} \to [0, 1], A \mapsto \tilde{\mathbb{P}}[A] := \frac{|A|}{|\Omega|}$

$(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ ist zwar auch ein Wahrscheinlichkeitsraum der das Experiment beschreibt, aber $\tilde{\Omega}$ enthält viele mehr Elemente als Ω, somit ist $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ nicht der einfachste mögliche Wahrscheinlichkeitsraum.

(b) $|\Omega| = 2^3 = 8$.

Alternative falsche Lösung: $|\tilde{\Omega}| = 100^3 = 10^6 = 1\,000\,000$.

(c) $|\mathcal{F}| = 2^8 = 256$.

Alternative falsche Lösung: $|\tilde{\mathcal{F}}| = 2^{1\,000\,000}$.