

Musterlösung

1. (10 Punkte)

a) (1 Pt)

$$P(E^c|B^+) = 1 - P(E|B^+) = 1 - 0.35 = 0.65$$

b) (1.5 Pt)

Abhängig (0.5 Pt).

$$P(E|E^c) = 0 \quad (1\text{Pt})$$

c) (3 Pt)

$$P(B^+|E) = \frac{P(E|B^+) \cdot P(B^+)}{P(E)} = \frac{0.35 \cdot 0.2}{0.1} = 0.7$$

$$P(B^-|E) = 1 - P(B^+|E) = 1 - 0.7 = 0.3$$

$$P(E^c \cap B^-) = P(B^-) - P(B^- \cap E) = P(B^-) - P(B^-|E) \cdot P(E) = 0.8 - 0.3 \cdot 0.1 = 0.77$$

oder auch

$$\begin{aligned} P(E^c \cap B^-) &= P(E^c|B^-) \cdot P(B^-) = \left(1 - P(E|B^-)\right) \cdot P(B^-) \\ &= \left(1 - \frac{P(B^c|E)P(E)}{P(B^-)}\right) \cdot P(B^-) = \left(1 - \frac{0.3 \cdot 0.1}{0.8}\right) \cdot 0.8 = 0.77 \end{aligned}$$

d) (1.5 Pt)

$$\begin{aligned} P(\text{Falsch}) &= P(E \cap B^+) + P(E^c \cap B^-) = P(E|B) \cdot P(B) + P(E^c \cap B^-) \\ &= 0.35 \cdot 0.2 + 0.77 = 0.07 + 0.77 = 0.84 \end{aligned}$$

(0.5 Pt) für die 1. Gleichung (auch mit Zahlen). (0.5 Pt) für $P(E \cap B^+)$ und (0.5 Pt) für das RICHTIGE Ergebnis.

e) (3 Pt)

$$\begin{aligned} \mathbf{E}[\text{Gewinn}_X] &= P(E^+) \cdot 25000 + P(E^-) \cdot (-5000) \\ &= 0.1 \cdot 25000 + 0.9 \cdot (-5000) = -2000 \text{ CHF} \quad (1\text{Pt}) \end{aligned}$$

$$\begin{aligned} \mathbf{E}[\text{Gewinn}_Y] &= P(E^+ \cap B^+) \cdot 25000 + P(E^- \cap B^+) \cdot (-5000) + P(B^-) \cdot 0 - 500 \\ &= 0.07 \cdot 25000 - (0.2 - 0.07) \cdot 5000 - 500 = 1750 - 650 - 500 = 600 \text{ CHF} \quad (2\text{Pt}) \end{aligned}$$

Firma X verliert (in Erwartung) 2000 CHF pro Produkt. Firma Y gewinnt 600 CHF, deswegen ist die Strategie von Firma Y besser.

2. (10 Punkte)

Diese Prüfung dauert maximal 2 Stunden (= 120 Minuten). Es werden aber vielleicht nur 108 Minuten verstreichen, bis Sie Ihre Prüfung abgeben. Ihr Banknachbar hingegen benötigt zum Beispiel 114 Minuten. Die benötigte Zeit bis zur Prüfungsabgabe ist also eine Zufallsvariable. Wir nehmen an, dass wir diese Zeit durch folgende kumulative Verteilungsfunktion beschreiben können:

$$F(x) = \begin{cases} 1 - \left(1 - \left(\frac{x}{120}\right)^a\right)^b & \text{für } 0 \leq x \leq 120 \\ 0 & \text{für } x < 0 \\ 1 & \text{für } x > 120 \end{cases}$$

wobei x die benötigte Zeit in Minuten und $a > 0, b > 0$ zwei reelle Parameter sind.

- a) (2 Punkte) Berechnen Sie die Dichte von obiger Verteilungsfunktion $F(x)$.
Die Dichte berechnet sich als die Ableitung der Kumulativen Verteilungsfunktion, für $0 \leq x \leq 120$ gilt

$$\begin{aligned} f(x) &= \frac{dF(x)}{dx} \\ &= \frac{ab}{120} \left(\frac{x}{120}\right)^{a-1} \left(1 - \left(\frac{x}{120}\right)^a\right)^{b-1}. \end{aligned}$$

Für $x < 0$ und $x > 120$ gilt $f(x) = 0$.

(1 Punkt für Dichte=Ableitung Verteilungsfunktion, 1 Punkt für richtige Antwort)

- b) (1 Punkt) Berechnen Sie die Wahrscheinlichkeit α zum Quantil mit Wert $q_\alpha = 100$, ausgedrückt in den Variablen a und b .

Die Wahrscheinlichkeit zu $q_\alpha = 100$ ist:

$$\alpha = F(100) = 1 - \left(1 - \left(\frac{5}{6}\right)^a\right)^b.$$

(0.5 Punkte für $\alpha = F(100)$ und 0.5 Punkte für Antwort)

- c) (1 Punkt) Nehmen Sie nun an, dass die Parameter $a = 14$ und $b = 1.1$ sind. Was ist die Wahrscheinlichkeit, dass Sie für die Abgabe der Prüfung länger als 90 Minuten brauchen?

Diese Wahrscheinlichkeit kann man mit der Verteilungsfunktion leicht berechnen:

$$P[X > 90] = 1 - P[X \leq 90] = 1 - F(90) = \left(1 - \left(\frac{90}{120}\right)^{14}\right)^{1.1} \approx 0.98.$$

(0.5 Punkte für korrekte Umformung 1-..., 0.5 Punkte für Antwort)

- d) (3 Punkte) Nehmen Sie für $F(x)$ wieder an, dass $a = 14$ und $b = 1.1$. Der Erwartungswert dieser Verteilung ist in etwa 111.5, und die Varianz 57.7. Was ist approximativ die Wahrscheinlichkeit, dass das arithmetische Mittel von 20 Studenten, welche i.i.d. mit $F(x)$ verteilt sind, grösser als 109?

Hier ist $\mu = 111.5$, $\sigma = \sqrt{57.7}$ und $n = 20$. Mit Hilfe des zentralen Grenzwertsatzes erhalten wir

$$\begin{aligned} P\left[\frac{1}{n} \sum_{i=1}^n X_i > 109\right] &= P\left[\frac{\sqrt{n}\left(\frac{1}{n} \sum_{i=1}^n X_i - \mu\right)}{\sigma} > \frac{\sqrt{n}(109 - \mu)}{\sigma}\right] \\ &= 1 - \Phi\left(\frac{\sqrt{n}(109 - \mu)}{\sigma}\right) \\ &\approx 1 - \Phi(-1.47) \\ &\approx 0.93 \end{aligned}$$

(1 Punkt Idee für Verteilung, 1 Punkt korrekte Umformung zur Normalverteilung mit ZGWS, 1 Punkt Antwort)

- e) (1 Punkt) Hätte man Teilaufgabe 2d) auch berechnen können, falls Sie nicht gewusst hätten, dass $a = 14$ und $b = 1.1$? Begründen Sie.

Hinweis: Nichts rechnen. Ja. Der zentrale Grenzwertsatz kann man auch anwenden ohne die genaue Verteilung zu kennen, man braucht nur i.i.d. verteilte Zufallsvariablen bei welchen man den Erwartungswert und die Varianz kennt.

(1 Punkt Begründung)

- f) (2 Punkte) Was ist die Wahrscheinlichkeit, dass bei 20 Studenten die längste Prüfungsdauer höchstens 118 Minuten ist? Verwenden Sie $a = 14$, $b = 1.1$. Nehmen Sie an, dass die Studenten unabhängig voneinander abgeben.

Ein Lösungsweg ist wie folgt:

$$\begin{aligned} P[\max(X_1, \dots, X_{20}) \leq 118] &= P[X_1 \leq 118, \dots, X_{20} \leq 118] \\ &= \prod_{i=1}^{20} P[X_i \leq 118] \\ &= (F(118))^{20} \\ &\approx 0.02. \end{aligned}$$

Ein anderer Lösungsweg kann man mit der Binomialverteilung berechnen. Da die Studenten unabhängig von einander abgeben, können wir diese Aufgabe auch als Binomialverteilung auffassen. Wir berechnen die Wahrscheinlichkeit, dass wir bei 20 "Versuchen", 20 "Erfolge" haben, mit Erfolgswahrscheinlichkeit $p = F(118)$. Ein Erfolg tritt hier ein, falls ein Student höchstens 118 Minuten braucht.

$$\begin{aligned} P[\max(X_1, \dots, X_{20}) \leq 118] &= \text{Binom}(x = 20, n = 20, p = F(118)) \\ &= \binom{20}{20} \cdot p^{20}(1-p)^0 \\ &= p^{20} = (F(118))^{20} \\ &\approx 0.02. \end{aligned}$$

(0.5 Punkte Maximum Verteilung oder Binomial, **1 Punkt** iid richtig verwenden, **0.5 Punkte** Antwort)

3. (7 Punkte)

Die Inkubationszeit eines Virus ist definiert als die Zeitspanne (gemessen in Tagen) von der Ansteckung bis zum Ausbruch der Krankheit. Diese Inkubationszeit wird mit folgender Dichte beschrieben:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{(\log(x)-1)^2}{2\sigma^2}} & x > 0 \end{cases}$$

wobei wir mit $\log(\cdot)$ den natürlichen Logarithmus bezeichnen. Hier ist $\sigma > 0$ der unbekannte Parameter, welchen es zu schätzen gilt. Wir möchten nun den Parameter der Verteilung kennen, dazu nehmen wir eine Stichprobe. Das Ziel ist, diesen unbekannt Parameter daraus zu schätzen, um somit zum Beispiel die Dauer von Quarantänemaßnahmen für ein neuartiges Virus abschätzen zu können. Wir betrachten die Daten (in Tagen):

x_1	x_2	x_3	x_4	x_5
1.62	5.29	3.93	1.49	1.57

Hinweis: Für eine Zufallsvariable X mit obiger Dichte gilt $\mathcal{E}[X] = e^{1+\sigma^2/2}$.

- a) (3 Punkte) Bestimmen Sie die Likelihood- und die log-Likelihoodfunktion basierend auf n unabhängigen identisch verteilten Beobachtungen x_1, \dots, x_n einer Zufallsvariable mit obiger Dichte. Berechnen Sie daraus den Maximum-Likelihood Schätzer für den unbekannt Parameter σ . Die konkreten Werte brauchen Sie nicht einzusetzen.

Die Likelihoodfunktion ist gegeben durch

$$\ell(\sigma; x_1, \dots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \left(\prod_{i=1}^n x_i\right)^{-1} \cdot \exp\left(-\sum_{i=1}^n \frac{(\log(x_i) - 1)^2}{2\sigma^2}\right).$$

Die log-Likelihoodfunktion ist dann

$$\log \ell(\sigma; x_1, \dots, x_n) = \log(1) - \frac{n}{2} \log(2\pi) - n \log(\sigma) - \sum_{i=1}^n \log(x_i) - \frac{1}{2\sigma^2} \sum_{i=1}^n (\log(x_i) - 1)^2.$$

Um nun den Maximum Likelihood Schätzer für σ zu erhalten leiten wir die log-Likelihoodfunktion nach σ ab und setzen das Resultat gleich null.

$$\left. \frac{d\ell}{d\sigma} \right|_{\hat{\sigma}} = -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum_{i=1}^n (\log(x_i) - 1)^2 \stackrel{!}{=} 0.$$

Wenn wir das auflösen nach $\hat{\sigma}$ erhalten wir den Schätzer

$$\hat{\sigma}_{MLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(x_i) - 1)^2}.$$

(1 Punkt Likelihood, **1 Punkt** log-Likelihood, **1 Punkt** Maximum Likelihood Schätzer, jeweils **-0.5 Punkte** für Fehler)

- b) (2 Punkte) Bestimmen Sie den Momentenschätzer für σ , sowohl basierend auf n unabhängigen Beobachtungen x_1, \dots, x_n , wie auch den realisierten Wert für die gegebene Stichprobe. Mit dem Tipp $\mathbf{E}[X] = e^{1+\sigma^2/2}$ erhalten wir durch Gleichsetzen der Momente $\mathbf{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i$ und auflösen nach σ den Momentenschätzer

$$\hat{\sigma}_{MoM} = \sqrt{\log\left(\left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right) - 2.}$$

Der Schätzer nimmt mit den Daten approximativ den Wert 0.2119 an.

(1 Punkt für richtige Formel und 1 Punkt für korrekten Wert)

- c) (1 Punkt) Nehmen wir für diese Teilaufgabe an, dass wir folgende Daten beobachtet haben:

x_1	x_2	x_3
2.12	1.63	4.21

Welcher Schätzer (Momentenschätzer oder Maximum-Likelihood Schätzer) ist bei diesen Daten zu bevorzugen? Begründen Sie.

Der Maximum Likelihood Schätzer ist zu bevorzugen, da dieser in allen Situationen definiert ist. Der Momentenschätzer hingegen ist nicht definiert für die gegebenen Daten, da wir haben, dass:

$$\log\left(\left(\frac{1}{3} \sum_{i=1}^3 x_i\right)^2\right) < 2.$$

Der Term unter der Wurzel wird negativ.

(0.5 Punkt Antwort, 0.5 Punkt Begründung)

- d) (1 Punkt) Einen weiteren Momentenschätzer kann man auch berechnen, wenn man nur das zweite Moment $\mathcal{E}[X^2] = e^{2+2\sigma^2}$ betrachtet. Berechnen Sie diesen Schätzer basierend auf n unabhängigen Beobachtungen x_1, \dots, x_n .

Den Ausdruck für das zweite Moment setzen wir mit dem zweiten empirischen Moment gleich, $\mathbf{E}[X^2] = \frac{1}{n} \sum_{i=1}^n x_i^2$. Danach können wir nach σ auflösen und erhalten einen anderen Momentenschätzer

$$\hat{\sigma}_{MoM2} = \sqrt{\log\left(\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - 1.}$$

(1 Punkt Antwort)

4. a) t -Test für X_i i.i.d. $\sim \mathcal{N}(\mu_X, \sigma^2)$ und X_j i.i.d. $\sim \mathcal{N}(\mu_Y, \sigma^2)$:
- $H_0 : \mu_X = \mu_Y, \quad H_A : \mu_X \neq \mu_Y$ (0.5 Pt)
 - Teststatistik $T = \frac{\bar{X}_n - \bar{Y}_m}{s_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} = t_{120}$ unter H_0 (0.5 Pt)
 - $s_{pool} = \sqrt{\frac{1}{n+m-2} ((n-1)s_X^2 + (m-1)s_Y^2)} = 0.75$ (0.5 Pt)
 - $t = -1.253$ (0.5 Pt)
 - Verwerfungsbereich $K = (-\infty, -t_{120,0.975}] \cup [t_{120,0.975}, \infty) = (-\infty, -1.98] \cup [1.98, \infty)$ (0.5 Pt)
 - Testentscheid: $t \notin K$, also wird H_0 nicht verworfen. (0.5 Pt)
- b) Der p -Wert ist das Niveau, bei dem gerade noch verworfen würde. Also das p , für das gilt:

$$|t| = 1.253 = t_{120, 1-p/2}$$

Aus der t -Quantil-Tabelle sehen wir, dass $t_{120,0.8} = 0.842$ und $t_{120,0.9} = 1.282$. Somit muss der p -Wert im Intervall $[0.2, 0.4]$ liegen.

- c) Das zweiseitige Vertrauensintervall für den t -Test ist gegeben durch

$$I = \bar{x}_n - \bar{y}_m \pm s_{pool} \sqrt{\frac{1}{n} + \frac{1}{m}} \cdot t_{n+m-2, 1-\frac{\alpha}{2}} \quad (0.5 Pt).$$

Also $I = [-0.44, 0.1]$ (0.5 Pt).

- d) $E[X_i] = E[Y_i]$ ist die einzige Annahme, die nicht notwendig ist.

- e) Die Macht ist die Wahrscheinlichkeit, dass H_0 verworfen wird, wenn H_A stimmt. Hierfür berechnen wir zuerst den Verwerfungsbereich (1 Pt):

$$K = (-\infty, -z_{0.975}] \cup [z_{0.975}, \infty) = (-\infty, -1.96] \cup [1.96, \infty).$$

Die Macht ist nun gegeben durch (1 Pt):

$$\begin{aligned} P_{H_A}(K) &= P_{\mathcal{N}(1,1)}((-\infty, -1.96] \cup [1.96, \infty)) \\ &= P_{\mathcal{N}(0,1)}((-\infty, -2.96] \cup [0.96, \infty)) \\ &= \Phi(-2.96) + 1 - \Phi(0.96) \\ &= 1 - \Phi(2.96) + 1 - \Phi(0.96) \\ &= 2 - 0.9985 - 0.8315 \\ &= 0.17. \end{aligned}$$

5. (12 Punkte)

- a) F,R,F,R
- b) F,R,R,R
- c) R,R,F,R
- d) R,F,R,R

e) F,R,R,R

f) F,F,R,R