

Musterlösung

1. (8 Punkte)

Anna isst gerne Joghurts und hat jederzeit viele Joghurts im Kühlschrank stehen. Manchmal verliert sie den Überblick und ein paar Joghurts überschreiten das Haltbarkeitsdatum um wenige Tage. Aus Erfahrung passiert ihr dies in 10% der Fällen ($\mathcal{P}(H^c) = 0.1$). Da Anna aber nicht gerne Lebensmittel wegwirft, möchte sie die Situation statistisch anschauen. Sie fragt die Joghurtfirma um Auskunft, welche ihr verspricht, dass beim Einhalten des Haltbarkeitsdatums sich in 99.5% der Fällen auch keine Anzeichen von Schimmel bilden ($\mathcal{P}(S^c | H) = 0.995$). Anna weiss aus Erfahrung auch, dass Joghurts, welche bei ihr im Kühlschrank das Haltbarkeitsdatum überschreiten, oft noch keinen Schimmel gebildet haben. Sie rechnet mit 70% ($\mathcal{P}(S^c | H^c) = 0.7$).

Wir definieren folgende Ereignisse:

H := Haltbarkeitsdatum nicht überschritten

H^c := Haltbarkeitsdatum überschritten

S := Schimmelbildung (Joghurt ungeniessbar)

S^c := keine Anzeichen von Schimmel (Joghurt geniessbar)

- a) (1 Punkt) Wie hoch ist die Wahrscheinlichkeit, dass bei einem Joghurt mit eingehaltenem Haltbarkeitsdatum sich trotzdem Schimmel bildet?

Wir kennen die Wahrscheinlichkeit des Komplements:

$$\mathcal{P}[S | H] = 1 - \mathcal{P}[S^c | H] = 1 - 0.995 = 0.005$$

- b) (1 Punkt) Wie hoch ist die Wahrscheinlichkeit, dass Anna einen Joghurt aus ihrem Kühlschrank herauszieht, welcher die Haltbarkeit nicht überschreitet und keinen Schimmel hat?

$$\mathcal{P}(H \cap S^c) = \mathcal{P}(S^c | H)\mathcal{P}(H) = 0.995(1 - 0.1) = 0.8955 \approx 0.9$$

- c) (2 Punkte) Thomas ist bei Anna zu Besuch und möchte einen Joghurt essen. Mit welcher Wahrscheinlichkeit nimmt Thomas einen schimmeligen Joghurt aus Annas Kühlschrank? (Ersatzresultat: 0.04)

Wir zerlegen in disjunkte Ereignisse und benutzen dann die Definition der bedingten Wahrscheinlichkeit:

$$\mathcal{P}[S] = \mathcal{P}[S \cap H] + \mathcal{P}[S \cap H^c] = \mathcal{P}[S | H]\mathcal{P}[H] + \mathcal{P}[S | H^c]\mathcal{P}[H^c] = 0.005 * 0.9 + (1 - 0.7) * 0.1 = 0.0345 \approx 3\%$$

Thomas muss sich nicht viele Sorgen machen. (1 Punkt Formel, 1 Punkt Ergebnis)

- d) (3 Punkte) Wie hoch ist die Wahrscheinlichkeit, dass ein schimmelfreier Joghurt in Annas Kühlschrank das Haltbarkeitsdatum einhält ($\mathcal{P}(H | S^c)$)? Sind H und S^c unabhängig?

Wir benutzen den Satz von Bayes:

$$\mathcal{P}[H | S^c] = \frac{\mathcal{P}[S^c | H]\mathcal{P}[H]}{\mathcal{P}[S^c]} = \frac{0.995 * 0.9}{1 - 0.0345} = 0.9274987. \text{ (mit Ersatzresultat } 0.9328125)$$

Somit sieht man, dass H und S^c abhängig sind ($\mathcal{P}(H | S^c) \neq \mathcal{P}(H)$). 1 Punkt Formel, 1 Punkt Ergebnis, 1 Punkt Unabhängigkeit.

- e) (1 Punkt) Was ist die Wahrscheinlichkeit $\mathcal{P}(H | H)$?

Per Definition ist das 1.

2. (8 Punkte)

Der Forscher Dominik versucht in seinen Ferien (zum Leidwesen seiner Frau) die Verteilung von Sandkorngrößen in der Nähe des Strandhauses zu untersuchen. Mit Hilfe eines Mikroskops hat er folgende kumulative Verteilungsfunktion bestimmt

$$F(x) := \frac{x^2}{x^2 + 1} \left(= \frac{1}{1 + \frac{1}{x^2}} \right), x > 0$$

welche den Durchmesser X (in mm) von Sandkörnern sinnvoll beschreibt.

- a) (1 Punkt) Bestimmen Sie die Dichte zur Verteilungsfunktion $F(x)$. Die Dichte kann man mit der Ableitung der Verteilungsfunktion erhalten:

$$f(x) = F'(x) = \frac{2x(x^2 + 1) - x^2 \cdot 2x}{(x^2 + 1)^2} = \frac{2x}{(x^2 + 1)^2}.$$

(1 Punkt richtige Antwort)

- b) (1.5 Punkte) Zeigen Sie, dass für das α -Quantil q_α gilt

$$q_\alpha = \sqrt{\frac{\alpha}{1 - \alpha}}.$$

Das α -Quantil ist definiert als derjenige Wert, welcher mit α Wahrscheinlichkeit unterschritten wird. Deshalb gilt für das α -Quantil per Definition:

$$P(X \leq q_\alpha) = F(q_\alpha) = \alpha.$$

(0.5 Punkte)

Mit Einsetzen in die gegebenen kumulativen Verteilungsfunktion haben wir:

$$\frac{q_\alpha^2}{q_\alpha^2 + 1} = \alpha,$$

Auflösen nach q_α ergibt

$$\begin{aligned} q_\alpha^2 &= \alpha(q_\alpha^2 + 1) \\ q_\alpha^2(1 - \alpha) &= \alpha \\ q_\alpha^2 &= \frac{\alpha}{1 - \alpha} \\ q_\alpha &= \sqrt{\frac{\alpha}{1 - \alpha}} \end{aligned}$$

(1 Punkt fürs richtige Auflösen)

- c) (1 Punkt) Wie hoch ist die Wahrscheinlichkeit, dass Dominik ein Sandkorn in die Hand nimmt, welches grösser als 1mm ist?

Die Wahrscheinlichkeit berechnet sich wie folgt

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - \frac{1}{1 + 1} = 1 - 0.5 = 0.5$$

Das heisst, 1mm ist der Median. (1 Punkt)

- d) (1 Punkt) Aus seinen Messungen hat Dominik beobachtet, dass etwas mehr als die Hälfte der Sandkörner einen Durchmesser zwischen 0.5mm und 2mm haben. Was ist die Wahrscheinlichkeit unter F , dass ein Sandkorn einen Durchmesser zwischen 0.5mm und 2mm hat?

Die Wahrscheinlich des Durchmessers eines Sandkorns zwischen 0.5mm und 2mm zu liegen ist

$$\begin{aligned} P(0.5 \leq X \leq 2) &= P(X \leq 2) - P(X \leq 0.5) \\ &= F(2) - F(0.5) = 0.8 - 0.2 = 0.6. \end{aligned}$$

Die Verteilung reflektiert also die Beobachtung. (1 Punkt)

- e) (1.5 Punkte) Angenommen wir beobachten die Durchmesser von zwei Sandkörnern (Unabhängigkeit ist gegeben). Wie hoch ist die Wahrscheinlichkeit, dass beide Durchmesser kleiner als 1mm sind? Wir definieren zwei Zufallsvariablen als $X_1 :=$ Durchmesser vom ersten Sandkorn und $X_2 :=$ Durchmesser von den zweiten Sandkorn. Um nun die gesuchte Wahrscheinlichkeit zu berechnen, benötigen wir die folgende Zufallsvariable

$$T := \max(X_1, X_2).$$

Beide Durchmesser sind kleiner als 1mm ist äquivalent zur Aussage, dass der maximale Durchmesser T kleiner als 1mm ist. Die gesuchte Wahrscheinlichkeit ist dann

$$\begin{aligned} P(T \leq t) &= P(X_1 \leq 1 \text{ und } X_2 \leq 1) \quad (0.5 \text{ Punkte für richtige Wahrscheinlichkeit}) \\ &= P(X_1 \leq 1) \cdot P(X_2 \leq 1) \quad \text{da unabhängig} \\ &= \left(\frac{1}{1 + 1}\right)^2 = 1/4. \end{aligned}$$

(1 Punkt für die richtige Antwort)

- f) (2 Punkte) Wie viele zufällige, unabhängige Sandkörner muss Dominik mindestens unter dem Mikroskop anschauen, um mit einer Wahrscheinlichkeit von mehr als 90% mindestens ein Sandkorn mit einem Durchmesser grösser als 2mm zu messen?

Angenommen wir haben n Durchmesser X_1, \dots, X_n von n unabhängigen Sandkörnern gemessen. Die Wahrscheinlichkeit, dass mindestens ein Durchmesser aus insgesamt n grösser ist als 2mm ist

$$\begin{aligned} P(\max(X_1, \dots, X_n) > 2\text{mm}) &= 1 - P(\max(X_1, \dots, X_n) \leq 2\text{mm}) \quad (0.5 \text{ Punkte}) \\ &= 1 - P(X_1 \leq 2) \cdot \dots \cdot P(X_n \leq 2) \quad \text{Unabhängigkeit} \quad (0.5 \text{ Punkte}) \\ &= 1 - (0.8)^n. \end{aligned}$$

Wir verlangen, dass diese Wahrscheinlichkeit grösser ist als 0.9, deshalb muss $1 - (0.8)^n > 0.9$ für ein minimales n . Falls nun Gleichheit herrscht haben wir

$$\begin{aligned} (0.8)^n &= 0.1 \\ n \log(0.8) &= \log(0.1) \\ n &= \frac{\log(0.1)}{\log(0.8)} \approx 10.3 \end{aligned}$$

Da n die minimale natürliche Zahl ist, für welche $0.1 > (0.8)^n$ gilt, braucht man also mindestens $n = 11$ Sandkörner um eine solche Wahrscheinlichkeit zu erhalten. (0.5 Punkte für 10.3 0.5 Punkte für aufrunden auf 11)

3. (7 Punkte)

Jan ist Versicherungsexperte. Um die Kosten X (in CHF) von Versicherungsfällen zu modellieren, verwendet er folgende Dichtefunktion:

$$f(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\theta\pi}} \exp\left(-\frac{x^2}{2\theta}\right) & \text{für } x > 0 \\ 0 & \text{für } x \leq 0, \end{cases}$$

wobei $\theta > 0$ ein unbekannter Modellparameter ist.

- a) (1 Punkt) Bestimmen Sie die Likelihoodfunktion basierend auf n unabhängigen, identisch verteilten Beobachtungen x_1, \dots, x_n einer Zufallsvariable mit obiger Dichte. Vereinfachen Sie die Terme jeweils so weit wie möglich. Die Likelihoodfunktion hat folgende Form:

$$L(\theta) = \begin{cases} \left(\frac{\sqrt{2}}{\sqrt{\theta\pi}}\right)^n \exp\left(-\frac{1}{2\theta} \sum_{i=1}^n x_i^2\right) & \text{für } x_i > 0, i = 1, \dots, n; \\ 0 & \text{sonst.} \end{cases}$$

Bei einem kleinen Fehler 1/2 Punkt Abzug. Ganzer Punkt auch wenn Fallunterscheidung vergessen. 1/2 Punkt für richtige Likelihoodfunktion, 1/2 Punkt für $\prod_{i=1}^n \frac{\sqrt{2}}{\sqrt{\theta\pi}} \exp\left(-\frac{x_i^2}{2\theta}\right)$.

- b) (2 Punkte) Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{\theta}_{\text{MLE}}$ für θ basierend auf n unabhängigen, identisch verteilten Beobachtungen x_1, \dots, x_n . Schreiben Sie die allgemeine Form auf. (Ersatzresultat: $\hat{\theta}_{\text{MLE}} = \frac{1}{n\pi} \sum_{i=1}^n x_i^2$) Für $x_i > 0, i = 1, \dots, n$ ist die Loglikelihoodfunktion

$$L(\theta) = \frac{n}{2} \log\left(\frac{2}{\theta\pi}\right) - \frac{1}{2\theta} \sum_{i=1}^n x_i^2.$$

Ableiten ergibt

$$L'(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2.$$

Da der Maximum-Likelihood-Schätzer die Loglikelihoodfunktion maximiert, erhalten wir die Bedingung

$$0 = L'(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2.$$

Durch Umformen erhalten wir folgenden Schätzer:

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Für kleine Fehler jeweils 1/2 Punkt Abzug. Falls falsche Likelihoodfunktion aus a) verwendet wurde, zählt dies in der Regel als Folgefehler. 1 Punkt für die Log-likelihoodfunktion, 1 Punkt für das Ableiten und Auflösen.

- c) (2 Punkte) Bestimmen Sie die Varianz von $\hat{\theta}_{\text{MLE}}$ für den Fall $\theta = 1$. Sie können verwenden, dass $\text{Var}(X^2) = 2\theta^2$.

$$\text{Var}(\hat{\theta}_{\text{MLE}}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n x_i^2\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i^2) = \frac{1}{n^2} \cdot n \cdot 2\theta^2 = \frac{2}{n}$$

Wir haben verwendet, dass x_i i.i.d. verteilte Zufallsvariablen sind. Für kleine Fehler jeweils 1/2 Punkt Abzug. 1 Punkt für $\frac{1}{n} \text{Var}(x_i^2)$.

- d) (2 Punkte) Berechnen Sie $\mathcal{E}(X)$ (Ersatzresultat: $\mathcal{E}(X) = \frac{\sqrt{2\theta}}{\sqrt{3}}$) und bestimmen Sie damit den Momentenschätzer $\hat{\theta}_{\text{Moment}}$ für θ basierend auf n unabhängigen, identisch verteilten Beobachtungen x_1, \dots, x_n .

Mit partieller Integration gilt

$$\mathcal{E}(X) = \int_0^\infty x \frac{\sqrt{2}}{\sqrt{\theta\pi}} \exp\left(-\frac{x^2}{2\theta}\right) dx = \left[-\frac{\sqrt{2\theta}}{\sqrt{\pi}} \exp\left(-\frac{x^2}{2\theta}\right) \right]_0^\infty = \frac{\sqrt{2\theta}}{\sqrt{\pi}}.$$

Folglich ist der Momentenschätzer:

$$\hat{\theta}_{\text{Moment}} = \frac{\pi}{2} \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2.$$

Für kleine Fehler innerhalb der Aufgabe jeweils 1/2 Punkt Abzug. 1 Punkt für Erwartungswert, 1 Punkt für das Auflösen nach θ .

4. (7 Punkte)

Ein Lebensmittelhersteller behauptet, dass ein neuentwickeltes Proteinbrot seiner Marke deutlich mehr Eiweiss enthält als das der Konkurrenz. Es soll 40g Eiweiss pro 100g Brot enthalten. Der Alpinist Christopher ist skeptisch und misst den Eiweissgehalt in $n = 30$ Broten nach. Die Brote haben jeweils eine Masse von 100g. Wir nehmen an, dass die Messungen unabhängig voneinander und identisch gemäss einer Normalverteilung verteilt sind. Mit X_i bezeichnen wir die i -te Messung des Eiweissgehaltes. Wir haben also $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Folgende Schätzer wurden ermittelt:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = 39.14$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 = 34.44$$

Wir wollen nun mit einem t -Test untersuchen, ob die Messergebnisse auf einen signifikant kleineren Wert für den Eiweissgehalt hinweisen als vom Hersteller angegeben. Verwenden Sie das Signifikanzniveau $\alpha = 0.05$.

- a) (2 Punkte) Schreiben Sie die Nullhypothese, die Alternativhypothese und die Teststatistik dazu auf. Um welche Art von Test handelt es sich? (Einstichprobentest, Zweistichprobentest, einseitig, zweiseitig, ...)

$$H_0 : \mu = \mu_0$$

$$H_A : \mu < \mu_0.$$

Es handelt sich um einen einseitigen Einstichprobentest. Die Teststatistik ist gegeben als

$$T = \sqrt{n} \frac{\hat{\mu} - \mu_0}{S_n}, \quad (1)$$

mit $S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2}$. (0.5 Punkte für richtige Null- und Alternativhypothese, 0.5 Punkte für einseitigen Einstichprobentest, 1 Punkt für richtige Teststatistik.)

- b) (1 Punkt) Führen Sie den entsprechenden t -Test durch: Was ist der Wert der Teststatistik? Was ist das Resultat des Tests? Zunächst berechnet man $s_n = \sqrt{\hat{\sigma}^2} = \sqrt{34.44} \approx 5.87$. Dann berechnet sich die Teststatistik als

$$T = \sqrt{30} \frac{39.14 - 40}{\sqrt{34.44}} \approx -0.8027 \quad (2)$$

Nun werden die Quantile in der Tabelle nachgeschaut ($n - 1 = 29$ Freiheitsgrade). Wir erhalten $t_{\alpha, n-1} = -1.7$. H_0 wird also nicht verworfen. (0.5 Punkte für Wert der Teststatistik und 0.5 Punkte für richtigen Testentscheid)

Nehmen Sie nun an, dass die Standardabweichung bekannt ist und verwenden Sie $\sigma = 5$. Wir wollen weiterhin untersuchen, ob die Messergebnisse auf einen signifikant kleineren Wert für den Eiweissgehalt hinweisen als vom Hersteller angegeben.

- c) (2 Punkte) Betrachten Sie nun den entsprechenden Z -Test mit $\alpha = 0.05$. Welchen Wert müssten wir für das arithmetische Mittel $\hat{\mu}$ beobachten, um die Nullhypothese gerade noch zu verwerfen? Runden Sie das Ergebnis auf zwei Nachkommastellen. Das 5% Quantil der Standardnormalverteilung ist $\Phi^{-1}(0.05) = -1.64$. Dann gilt: $z = \sqrt{n} \frac{\hat{\mu} - \mu_0}{\sigma} = -1.64$. Auflösen nach $\hat{\mu}$ ergibt $\hat{\mu} \approx 38.5$. Dies ist der grösste Wert für das arithmetische Mittel $\hat{\mu}$, bei dem wir die Nullhypothese gerade noch verwerfen würden. (1 Punkt für richtige Formel für $\hat{\mu}$ und 1 Punkt für richtigen Wert. Bei falschem Quantil Teilpunkte.)
- d) (2 Punkte) Berechnen Sie die Macht des Z -Tests unter der Alternativhypothese $\mu_A = 39$ mit $\sigma = 5$. Runden Sie das Ergebnis auf zwei Nachkommastellen. Sei $Y \sim \mathcal{N}(\mu_A, \sigma/\sqrt{n})$. Die Macht berechnet sich wie folgt: $\mathbb{P}(Y \leq 38.5) = \Phi(\sqrt{n} \frac{38.5 - \mu_A}{\sigma}) = \Phi(-0.55) \approx 0.29$. (1 Punkt für Ausdruck $\Phi(\sqrt{n} \frac{38.5 - \mu_A}{\sigma})$ und 1 Punkt für richtiges Ergebnis)

5. (10 Punkte) Entscheiden Sie für jede Aussage, ob sie richtig oder falsch ist. Markieren Sie Ihre Antwort auf dem separaten Multiple Choice Antwortblatt.

- a) (2 Punkte) R, R, R, R
- b) (2 Punkte) R, F, R, R
- c) (2 Punkte) F, F, R, R
- d) (2 Punkte) F, R, R, R
- e) (2 Punkte) F, R, R, R