Einführung in die Wahrscheinlichkeitstheorie und Statistik
Prof. Dr. Jan Johannes
Sergio Brenner Miguel
Wintersemester 2020/21

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

5. Übungsblatt - Lösungsskizzen

Aufgabe 17 (Neyman-Pearson-Lemma für stetige Verteilungen, 4 Punkte).
Beweisen Sie das Neyman-Pearson-Lemma aus der Vorlesung:
Sei $(\mathbb{R}^n, \mathscr{B}^n, (\mathbb{P}_0, \mathbb{P}_1))$ ein (binäres) stetiges statistisches Experiment. Für das Testproblem der einfachen Nullhypothese $H_0 : \{\mathbb{P}_0\}$ gegen die einfache Alternativhypothese $H_1 : \{\mathbb{P}_1\}$ ist jeder Neyman-Pearson-Test $\varphi = \mathbb{1}_{A_k}$ mit kritischem Wert $k \in \mathbb{R}^+$ und Ablehnbereich

$$A_k = \{x \in \mathbb{R}^n \mid \mathbb{f}_1(x) \geq k\mathbb{f}_0(x)\}$$

ein **bester Test zum Niveau** $\mathbb{P}_0(A_k) \in [0,1]$.

Lösung 17.
Dass der Neyman-Pearson-Test $\varphi = \mathbb{1}_{A_k}$ das Niveau $\alpha = \mathbb{P}_0(A_k)$ einhält ist klar. Es bleibt noch zu zeigen, dass der Test den Fehler zweiter Art unter allen α-Tests minimiert. D.h. es ist zu zeigen:
Für jeden anderen Test $\tilde{\varphi}$ mit $\mathbb{P}_0(\tilde{\varphi} = 1) \leq \alpha$ gilt:

$$\mathbb{P}_1(\tilde{\varphi} = 0) \geq \mathbb{P}_1(\varphi = 0).$$

Seien nun A bzw. $\tilde{A}$ die Ablehnbereiche von φ bzw. $\tilde{\varphi}$.
Es gilt nach Definition des Neyman-Pearson-Tests φ:

$$x \in A \iff \mathbb{f}_1(x) - k\mathbb{f}_0(x) \geq 0$$

Also folgt (wir können die Integrale hinschreiben, weil A und $\tilde{A}$ messbare Mengen sind):

$$\begin{aligned}
& \int_A \mathbb{f}_1(x) - k\mathbb{f}_0(x)dx && \geq \int_{\tilde{A}} \mathbb{f}_1(x) - k\mathbb{f}_0(x)dx \\
\implies \quad & \int_A \mathbb{f}_1(x)dx - \int_{\tilde{A}} \mathbb{f}_1(x)dx && \geq k\left(\int_A \mathbb{f}_0(x)dx - \int_{\tilde{A}} \mathbb{f}_0(x)dx\right) \\
\implies \quad & \mathbb{P}_1(A) - \mathbb{P}_1(\tilde{A}) && \geq \underbrace{k\left(\underbrace{\mathbb{P}_0(A)}_{=\alpha} - \underbrace{\mathbb{P}_0(\tilde{A})}_{\leq\alpha}\right)}_{\geq 0} \\
\implies \quad & \mathbb{P}_1(A) && \geq \mathbb{P}_1(\tilde{A}) \\
\implies \quad & 1 - \mathbb{P}_1(A) = \mathbb{P}_1(A^c) = \mathbb{P}_1(\varphi = 0) && \leq \mathbb{P}_1(\tilde{A}^c) = 1 - \mathbb{P}_1(\tilde{A}) = \mathbb{P}_1(\tilde{\varphi} = 0).
\end{aligned}$$

Aufgabe 18 (Neyman-Pearson-Lemma, Konfidenzbereiche, 4 = 4 × 1 Punkte).
Eine Forschungsgruppe untersucht die Halbwertszeit von ^{60}Co. Die bisher vermutete Halbwertszeit beträgt $\lambda_0 = 5.2714$ Jahre. Die Forschenden beobachten nun in einem Experiment

eine Halbwertszeit von $X = 13.0$ Jahren.
In ihrem Modell nehmen die Forscher an, dass die Halbwertszeit exponentialverteilt ist mit Parameter $\frac{1}{\lambda}$ (wobei $\lambda > 0$), d.h. $X \sim \mathrm{Exp}_{\frac{1}{\lambda}} =: \mathbb{P}_\lambda$. Die Forschenden wollen basierend auf ihrer Messung nachweisen, dass die wahre Halbwertszeit größer als die bisher vermutete ist.

(a) Sei zunächst $\lambda_1 > \lambda_0 > 0$ fest gewählt. Formulieren Sie den Neyman-Pearson-Test $\varphi : \mathbb{R} \to \{0,1\}$ für die $\mathbb{P}_\lambda$-Verteilung zu den Hypothesen

$$H_0 : \lambda = \lambda_0 \quad \text{gegen} \quad H_1 : \lambda = \lambda_1$$

zum Niveau $\alpha \in (0,1)$. Vereinfachen Sie dann diesen Test mittels der Technik des monotonen Likelihood-Quotienten.

(b) Begründen Sie, warum φ aus (a) auch ein gleichmäßig bester Test zum Niveau α für die Hypothesen

$$H_0 : \lambda \leq \lambda_0 \quad \text{gegen} \quad H_1' : \lambda > \lambda_0$$

ist.

(c) Die Forschende wollen den peinlichen Fehler, falsch zu liegen, nur mit einer Wahrscheinlichkeit von 5% begehen. Werden sie sich auf Basis ihrer Beobachtung X und dem Test φ dazu entscheiden, ihre Ergebnisse zu publizieren?

(d) Geben Sie einen gleichmäßig besten Konfidenzbereich zum Niveau $1-\alpha$ für die falschen Parameter $\mathcal{F}_\lambda = (0, \lambda)$ an.

Lösung 18. (a) Wir untersuchen Verteilungen $\mathbb{P}_\lambda = \mathrm{Exp}(\frac{1}{\lambda})$ mit Dichten $\mathbb{f}_\lambda(x) = \frac{1}{\lambda}\exp(-\frac{1}{\lambda}x)\mathbb{1}_{\{x \geq 0\}}$.
Laut dem Neyman-Pearson-Lemma lautet der Neyman-Pearson-Test zum Niveau $\alpha \in (0,1)$ für die Hypothesen $H_0 : \lambda = \lambda_0$ gegen $H_1 : \lambda = \lambda_1$:

$$\varphi : \mathbb{R} \to \{0,1\}, \quad \varphi(x) = \mathbb{1}_{\{f_{\lambda_1} \geq k f_{\lambda_0}\}} = \begin{cases} 1, & L(x) \geq k, \\ 0, & L(x) < k \end{cases},$$

wobei $L(x) := \frac{\mathbb{f}_{\lambda_1}(x)}{\mathbb{f}_{\lambda_0}(x)}$ (für $x \geq 0$) der Likelihood-Quotient ist und $\mathbb{P}_{\lambda_0}(\varphi = 1) = \alpha$ gelten muss. Hier haben wir für $x \geq 0$:

$$L(x) = \frac{\mathbb{f}_{\lambda_1}(x)}{\mathbb{f}_{\lambda_0}(x)} = \frac{\lambda_0}{\lambda_1} \cdot \exp\left(x \cdot \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)\right).$$

Da $\lambda_1 > \lambda_0 > 0$ und daher $\frac{1}{\lambda_0} - \frac{1}{\lambda_1} > 0$, ist der Likelihoodquotient $L(x)$ monoton wachsend in x und der Test vereinfacht sich zu

$$\varphi : \mathbb{R} \to \{0,1\}, \quad \varphi(x) = \begin{cases} 1, & x \geq c^*, \\ 0, & x < c^* \end{cases}, \tag{1}$$

mit

$$\alpha = \mathbb{P}_{\lambda_0}(\varphi = 1) = \mathbb{P}_{\lambda_0}(\{x \geq c^*\}) = 1 - \mathbb{P}_{\lambda_0}((-\infty, c^*]) = 1 - \mathbb{F}_{\lambda_0}(c^*),$$

wobei $\mathbb{F}_\lambda(x) = \left(1 - e^{-\frac{1}{\lambda}x}\right)\mathbb{1}_{\{x \geq 0\}}$ die Verteilungsfunktion der Exponentialverteilung mit Parameter $\frac{1}{\lambda}$ ist. Da für $c^* \leq 0$ offensichtlich keine Lösung existiert, nehmen wir $c^* > 0$ an und finden damit die Lösung

$$\alpha = e^{-\frac{1}{\lambda_0}c^*} \Leftrightarrow \underline{c^* = -\lambda_0 \cdot \ln(\alpha)} \tag{2}$$

(b) Wie man an (1) und (2) sieht, hängt der gesamte Test φ nicht von der konkreten Wahl von λ_1 mit $\lambda_1 > \lambda_0$ ab. Für jedes $\lambda_1 > \lambda_0$ und Hypothesen

$$H_0 : \lambda = \lambda_0 \quad \text{gegen} \quad H_1 : \lambda = \lambda_1$$

erhalten wir also denselben Test φ mit der Optimalitätsaussage: φ minimiert den Fehler 2. Art $\mathbb{P}_{\lambda_1}(\phi = 0)$ unter allen Tests ϕ mit Fehler 1. Art $\mathbb{P}_{\lambda_0}(\varphi = 1)$.
Damit ist φ gleichmäßig bester Test (zum Niveau α)für die Hypothesen

$$H_0 : \lambda = \lambda_0 \quad \text{gegen} \quad H_1'' : \lambda > \lambda_0.$$

Es gilt außerdem für alle $\lambda \leq \lambda_0$:

$$\mathbb{P}_\lambda(\varphi = 1) = 1 - F_\lambda(c^*) = 1 - (1 - e^{-\frac{1}{\lambda}c^*}) = e^{-\frac{1}{\lambda}c^*} \leq e^{-\frac{1}{\lambda_0}c^*} = \alpha,$$

der Test hält also auch für alle $\lambda \leq \lambda_0$ das Niveau α ein. D.h. φ ist sogar ein gleichmäßig bester Test für die Hypothesen

$$H_0 : \lambda \leq \lambda_0 \quad \text{gegen} \quad H_1' : \lambda > \lambda_0.$$

(c) Mit der Bemerkung über den peinlichen Fehler wird ein Niveau $\alpha = 0.05$ für den Fehler 1. Art vorgegeben (die Hypothesen wurden in der Aufgabenstellung (a) schon richtig herum definiert, damit der peinliche Fehler wirklich dem Fehler 1. Art entspricht).
Wir erhalten damit

$$c^* = -\lambda_0 \cdot \ln(\alpha) \approx 15.8.$$

Das bedeutet, der Test φ liefert für die Beobachtung $X = 13.0$ trotzdem $\varphi(X) = 0$. Das bedeutet, der Test verwirft die Nullhypothese $\lambda \leq \lambda_0$ nicht. Die Forschende können also auf Basis des Tests φ und ihrer Beobachtung X zum Niveau $\alpha = 0.05$ nicht nachweisen, dass die Halbwertszeit tatsächlich größer ist. Sie sollten ihre Ergebnisse besser nicht publizieren.

(d) Wir haben $\mathbb{R}^+ = \mathcal{R}_\lambda \dot\cup \mathcal{F}_\lambda$ mit $\mathcal{F}_\lambda = (0, \lambda)$ und $\mathcal{R}_\lambda = [\lambda, \infty)$. Analog zu Beispiel 12.31. ist die assoziierte Familie von Partitionen gegeben durch
$\mathscr{H}^0_{\lambda_0} = \{\lambda \in \mathbb{R}^+ \mid \lambda_0 \in R_\lambda\} = \{\lambda \in \mathbb{R}^+ \mid \lambda_0 \in [\lambda, \infty)\} = \{\lambda \in \mathbb{R}^+ \mid \lambda_0 \geq \lambda\} = (0, \lambda_0]$ und $\mathscr{H}^1_{\lambda_0} = (\lambda_0, \infty)$.
In Aufgabenteil (b) haben wir einen gleichmäßig besten Test $\varphi_{\lambda_0} = \mathbb{1}_{A_{\lambda_0}}$ zum Niveau α für das Testproblem

$$H_0 : \mathscr{H}^0_{\lambda_0} = (0, \lambda_0] \quad \text{gegen} \quad H_1 : \mathscr{H}^1_{\lambda_0} = (\lambda_0, \infty)$$

gefunden, nach Satz 12.33 ist die assoziierte Bereichsschätzfunktion B ein gleichmäßig bester $(1 - \alpha)$ Konfidenzbereich, wobei B gegeben ist durch

$$\begin{aligned} B(x) &= \big\{\lambda \in \mathbb{R}^+ \mid x \notin \mathcal{A}_\lambda\big\} \\ &= \big\{\lambda \in \mathbb{R}^+ \mid x \notin [c^*, \infty)\big\} \\ &= \big\{\lambda \in \mathbb{R}^+ \mid x < -\lambda \ln(\alpha)\big\} \\ &= \big\{\lambda \in \mathbb{R}^+ \mid \ln(1/\alpha)^{-1} x < \lambda\big\} \\ &= (x \ln(1/\alpha)^{-1}, \infty). \end{aligned}$$

Aufgabe 19 (Berechnung MLS, 4 = 2 + 2 Punkte).

Seien $X_1, ..., X_n$ identisch verteilte Zufallsvariablen mit gemeinsamer Produktdichte $\mathbb{f}(x_1, \ldots, x_n) = \prod_{i=1}^n \mathbb{f}_i(x_i)$, wobei $\mathbb{f}_i$ die Wahrscheinlichkeitsdichte von X_i sei. Berechnen Sie den zugehörigen Maximum-Likelihood-Schätzer $\hat{\theta}_n$ für θ, falls

(a) $X_1 \sim \mathrm{N}_{(\mu,\sigma^2)}$ normalverteilt mit Parametervektor $\theta = (\mu, \sigma^2)$,
Hinweis: Sie müssen nicht nachrechnen, dass $\hat{\theta}_n$ ein globales Minimum/Maximum ist.

(b) $X_1 \sim \mathrm{Bin}_{(m,p)}$ binomialverteilt mit Parameter $\theta = p \in (0, 1)$, $m \in \mathbb{N}$ sei hier bekannt.

Lösung 19. (a) Erinnerung: Die Dichte der Normalverteilung N_{μ,σ^2} ist gegeben durch:

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Ein Maximum-Likelihood-Schätzer ist eine Funktion $\hat{\theta}_n = \hat{\theta}_n(X_1, ..., X_n)$ mit der Eigenschaft

$$\hat{\theta}_n \in \underset{\theta\in\Theta}{\arg\max}\ L_n(\theta) \qquad \text{mit} \qquad L_n(\theta) = \prod_{i=1}^n \mathbb{f}_\theta(X_i).$$

Anstatt $L_n(\theta)$ zu maximieren, können wir auch $l_n(\theta) := -\frac{1}{n}\log(L_n(\theta))$ bzgl θ minimieren.
Gleich zu Beginn stellen wir folgendes fest: Wir dürfen annehmen, dass nicht alle X_i denselben Wert haben, denn

$$\mathbb{P}(X_1 = ... = X_n) = \int_{\{x_1=...=x_n\}} f_{X_1,...,X_n}(x_1, ..., x_n)\ \mathrm{d}(x_1, ..., x_n) = 0,$$

($\{x_1, ..., x_n\} \in \mathbb{R}^n : x_1 = ... = x_n\}$ ist eine Lebesgue-Nullmenge in $\mathbb{R}^n$) d.h. der Fall $X_1 = ... = X_n$ tritt mit Wahrscheinlichkeit Null auf.
In unserem Fall ist

$$\begin{aligned} l_n(\mu, \underbrace{\sigma^2}_{=:s}) &= -\frac{1}{n}\sum_{i=1}^n \log(f_\theta(X_i)) = -\frac{1}{n}\sum_{i=1}^n \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)\right) \\ &= -\frac{1}{n}\sum_{i=1}^n \Big(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(X_i-\mu)^2}{2\sigma^2}\Big) \\ &= \frac{1}{2}\log(2\pi) + \frac{1}{2}\log(s) + \frac{1}{2\cdot n\cdot s}\sum_{i=1}^n (X_i-\mu)^2 \end{aligned}$$

Zur Bestimmung des Minimums bilden wir die erste Ableitung:

$$0 \overset{!}{=} \nabla l_n(\mu, s) = \begin{pmatrix} \frac{\partial l_n}{\partial \mu} \\ \frac{\partial l_n}{\partial s} \end{pmatrix} = \begin{pmatrix} -\frac{1}{n\cdot s}\cdot\sum_{i=1}^n (X_i-\mu) \\ \frac{1}{2s} - \frac{1}{2\cdot n\cdot s^2}\sum_{i=1}^n (X_i-\mu)^2 \end{pmatrix}$$

Aus der Gleichung in der ersten Komponente folgt:

$$0 = -\frac{1}{n\cdot s}\cdot\sum_{i=1}^n (X_i-\mu) \qquad \Rightarrow \qquad \underline{\underline{\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X_n},}}$$

aus der Gleichung in der zweiten Komponente folgt:

$$0 = \frac{1}{2s} - \frac{1}{2 \cdot n \cdot s^2} \sum_{i=1}^{n} (X_i - \mu)^2 \qquad \Rightarrow \qquad \underline{\underline{\hat{s}_n}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2 = \underline{\underline{\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X_n})^2}}$$

Wir müssen nun überprüfen, ob an der Stelle $(\hat{\mu}_n, \hat{s}_n)$ wirklich ein (lokales) Minimum der Funktion l_n vorliegt. Überprüfe dazu die Hesse-Matrix:

$$(\mathrm{Hess} l_n)(\hat{\mu}_n, \hat{s}_n) = \begin{pmatrix} \frac{\partial^2 l_n}{\partial \mu^2} & \frac{\partial^2 l_n}{\partial \mu \partial s} \\ \frac{\partial^2 l_n}{\partial s \partial \mu} & \frac{\partial^2 l_n}{\partial s^2} \end{pmatrix} (\hat{\mu}_n, \hat{s}_n) = \begin{pmatrix} \frac{1}{\hat{s}_n} & \frac{1}{n \cdot \hat{s}_n^2} \sum_{i=1}^{n} (X_i - \hat{\mu}_n) \\ \frac{1}{n \cdot \hat{s}_n^2} \sum_{i=1}^{n} (X_i - \hat{\mu}_n) & -\frac{1}{2\hat{s}_n^2} + \frac{1}{\hat{s}_n^3} \cdot \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2 \end{pmatrix}$$

Mit $\sum_{i=1}^{n} (X_i - \hat{\mu}_n) = (\sum_{i=1}^{n} X_i) - n \cdot \hat{\mu}_n = n \cdot \overline{X_n} - n \cdot \overline{X_n} = 0$ und $\hat{s}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2$ ist:

$$= \begin{pmatrix} \frac{1}{\hat{s}_n} & 0 \\ 0 & \frac{1}{2 \cdot \hat{s}_n^2} \end{pmatrix}.$$

Die Eigenwerte der Hesse-Matrix sind also $\frac{1}{\hat{s}_n}$ und $\frac{1}{2 \cdot \hat{s}_n}$. Es ist $\hat{s}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2 > 0$ (denn $\hat{s}_n = 0 \Leftrightarrow X_1 = ... = X_n = \hat{\mu}_n$, und das haben wir oben ausgeschlossen) und insbesondere hat die Hesse-Matrix nur positive Eigenwerte, ist also positiv definit. Damit besitzt l_n bei $(\hat{\mu}_n, \hat{\sigma}_n^2)$ ein striktes lokales Minimum.
Man kann zeigen, dass $(\hat{\mu}_n, \hat{\sigma}_n^2)$ sogar ein globales Minimum von l_n ist (Analysis). Damit ist

$$\hat{\theta}_n = (\hat{\mu}_n, \hat{s}_n) = \left(\overline{X_n}, \quad \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X_n})^2 \right)$$

der Maximum-Likelihood-Schätzer für den Parametervektor $\theta = (\mu, \sigma^2)$.

(b) Die Dichte der Binomialverteilung ist gegeben durch

$$\mathbb{f}_\theta(x) = \binom{m}{x} \theta^x (1 - \theta)^{m-x}.$$

Ein Maximum-Likelihood-Schätzer ist eine Funktion $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ mit der Eigenschaft

$$\hat{\theta}_n \in \arg\max_{\theta \in \Theta} L_n(\theta) \quad \text{mit} \quad L_n(\theta) = \prod_{i=1}^{n} \mathbb{f}_\theta(X_i).$$

Anstatt $L_n(\theta)$ zu maximieren, können wir auch die log-Liklihood $l_n(\theta) = \log(L_n(\theta))$ maximieren. Wir erhalten für $\theta = p$

$$\begin{aligned} l_n(p) = \log(L_n(p)) &= \log \left(\prod_{i=1}^{n} \binom{m}{X_i} p^{X_i} (1-p)^{m - X_i} \right) \\ &= \log \left(\left(\prod_{i=1}^{n} \binom{m}{X_i} \right) p^{\sum_{i=1}^{n} X_i} (1-p)^{m \cdot n - \sum_{i=1}^{n} X_i} \right) \\ &= \log \left(\prod_{i=1}^{n} \binom{m}{X_i} \right) + \log \left(p^{\sum_{i=1}^{n} X_i} \right) + \log \left((1-p)^{m \cdot n - \sum_{i=1}^{n} X_i} \right) \\ &= \log \left(\prod_{i=1}^{n} \binom{m}{X_i} \right) + \sum_{i=1}^{n} X_i \log(p) + \left(m \cdot n - \sum_{i=1}^{n} X_i \right) \log(1-p). \end{aligned}$$

Zur Bestimmung des Maximums bilden wir die erste Ableitung:

$$\frac{\mathrm{d}}{\mathrm{d}p} l_n(p) = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(m \cdot n - \sum_{i=1}^{n} X_i\right)$$

Es folgt:

$$0 \stackrel{!}{=} \frac{\mathrm{d}}{\mathrm{d}p} l_n(\hat{p}_n) \qquad \Longrightarrow \qquad \underline{\underline{\hat{p}_n = \frac{1}{n \cdot m}\sum_{i=1}^{n} X_i = \frac{1}{m}\overline{X_n}}},$$

Wir müssen noch überprüfen, ob an der Stelle $\hat{p}_n$ tatsächlich ein Maximum vorliegt, dazu bilden wir die zweite Ableitung:

$$\frac{\mathrm{d}^2}{\mathrm{d}p^2} l_n(p) = -\frac{1}{p^2}\sum_{i=1}^{n} X_i - \frac{1}{(1-p)^2}\left(m \cdot n - \sum_{i=1}^{n} X_i\right)$$

Da $\sum_{i=1}^{n} X_i \leq n \cdot m$ ist, handelt es sich bei $\hat{p}_n$ tatsächlich um ein Maximum. Damit ist

$$\hat{p}_n = \frac{1}{m}\overline{X_n}$$

der Maximum-Likelihood-Schätzer für den Parameter p.

Aufgabe 20 (Beispiel: Nichtexistenz MLS, 4 = 3 + 1 Punkte).
Wir wollen die durchschnittliche Anzahl der Zerfälle $\lambda > 0$ pro Stunde von ^{60}Co bestimmen. Wir wissen, dass diese Anzahl Poi(λ)-verteilt ist. Wir haben n-mal das exakt gleiche ^{60}Co-Präparat beschafft (d.h. wir können annehmen, dass die gemeinsame Wahrscheinlichkeitsdichte durch die Produktdichte gegeben ist) und führen nun mit jedem Präparat den folgenden Versuch durch ($i = 1, ..., n$):
Wir halten das Präparat eine Stunde an den Geigerzähler und notieren die Anzahl der Zerfälle X_i, die er uns anzeigt. Leider entdecken wir erst am Ende unserer Versuchsreihe, dass der Geigerzähler defekt ist und nur noch erkennt, ob in der Stunde überhaupt ein Zerfall geschehen ist oder nicht (d.h. wir beobachten nur $X_i = 0$ oder $X_i = 1$).
Wir müssen also gezwungenermaßen mit unseren Beobachtungen $X_1, ..., X_n$ auskommen. Auch hier können wir annehmen, dass die gemeinsame Dichte der Produktdichte entspricht.

(a) Zeigen Sie, dass der Maximum-Likelihood-Schätzer $\hat{\lambda}_n$ für λ basierend auf $X_1, ..., X_n$ durch

$$\hat{\lambda}_n = -\log(1 - \overline{X_n})$$

gegeben ist, sofern er existiert. Hierbei bezeichnet $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Hinweis: Ist $X \sim Bin_{(1,p)}$, so ist die Zähldichte gegeben durch $\mathbb{p}_X(k) = p^k \cdot (1-p)^{1-k}$ für $k \in \{0, 1\}$.

(b) Zeigen Sie, dass der Maximum-Likelihood-Schätzer mit positiver Wahrscheinlichkeit nicht existiert.

Lösung 20. (a) Sei $Y_1, ..., Y_n \sim \mathrm{Poi}_\lambda$ die Anzahl der tatsächlich geschehenen Zerfälle in der Stunde (da die Präparate unabhängig voneinander untersucht wurden und alle gleich beschaffen sind, nehmen wir an, dass die gemeinsame Verteilung durch die Produktdichte beschrieben wird). Die X_i lassen sich durch die Y_i ausdrücken:

$$X_i = \begin{cases} 0, & Y_i = 0, \\ 1, & Y_i \geq 1. \end{cases}$$

Daher sind die $X_1, ..., X_n$ verteilt mit

$$\mathbb{P}(X_i = 0) = \mathbb{P}(Y_i = 0) = e^{-\lambda}, \quad \mathbb{P}(X_i = 1) = \mathbb{P}(Y_i \geq 1) = 1 - e^{-\lambda},$$

d.h. $X_1, ..., X_n \sim \text{Bin}_{1,1-e^{-\lambda}}$ folgen einer Bernoulli-Verteilung. Die Zähldichte ist gegeben durch

$$\mathbb{p}_\lambda(k) = (1 - e^{-\lambda})^k \cdot (e^{-\lambda})^{1-k}.$$

Wir erhalten für die gemeinsame negative log-Likelihood (wie üblich $\overline{X_n} := \frac{1}{n} \sum_{i=1}^n X_i$):

$$\begin{aligned} l_n(\lambda) &= -\frac{1}{n} \sum_{i=1}^n \log \mathbb{p}_\lambda(X_i) = -\frac{1}{n} \sum_{i=1}^n \Big(X_i \cdot \log(1 - e^{-\lambda}) + (1 - X_i) \cdot (-\lambda) \Big) \\ &= -\overline{X_n} \cdot \Big(\lambda + \log(1 - e^{-\lambda}) \Big) + \lambda \end{aligned}$$

Zur Bestimmung des Minimums berechnen wir die ersten zwei Ableitungen von $l_n(\lambda)$:

$$\begin{aligned} l_n'(\lambda) &= -\overline{X_n}\Big(1 + \frac{e^{-\lambda}}{1 - e^{-\lambda}}\Big) + 1 = 1 - \overline{X_n} \cdot (1 - e^{-\lambda})^{-1}, \\ l_n''(\lambda) &= \overline{X_n} \cdot (1 - e^{-\lambda})^{-2} e^{-\lambda}, \end{aligned}$$

und lösen nun

$$0 = l_n'(\lambda) \quad \Leftrightarrow \quad 1 - e^{-\lambda} = \overline{X_n} \quad \Leftrightarrow \quad \lambda = -\log(1 - \overline{X_n}) =: \hat{\lambda}_n.$$

Die letzte Umformung ist aber nur möglich, falls $\overline{X_n} \neq 1$. Im Falle $\overline{X_n} = 1$ existiert der Maximum-Likelihood-Schätzer nicht. Formal besitzt die Funktion $l_n(\lambda) = -\log(1 - e^{-\lambda})$ in diesem Falle kein Minimum.

Im Falle $\overline{X_n} = 0$ ist die Lösung $\lambda = 0$ strenggenommen ebenfalls nicht im zulässigen Bereich $\lambda > 0$ enthalten und der Maximum-Likelihood-Schätzer existiert auch dann nicht.

Ist $\overline{X_n} \notin \{0, 1\}$, so gilt

$$l_n''(\hat{\lambda}_n) = \overline{X_n} \cdot (\overline{X_n})^{-2} \cdot (1 - \overline{X_n}) > 0,$$

d.h. dann besitzt l_n in $\hat{\lambda}_n$ wirklich ein lokales Minimum. Da $\hat{\lambda}_n$ die einzige Extremstelle war und die stetige Funktion l_n über ein offenes Intervall $\lambda \in (0, \infty)$ minimiert wird, ist $\hat{\lambda}_n$ auch das globale Minimum.

(b) Wir haben schon in (a) gesehen, dass der Maximum-Likelihood-Schätzer nicht existiert, wenn $\overline{X_n} = 1 \Leftrightarrow X_1 = ... = X_n = 1$. Die Wahrscheinlichkeit dafür ist

$$\mathbb{P}(\overline{X_n} = 1) = \mathbb{P}(X_1 = 1)^n = (1 - e^{-\lambda})^n > 0,$$

also positiv.
Auch für $\overline{X_n} = 0 \Leftrightarrow X_1 = \cdots = X_n = 0$ existiert der Maximum-Likelihood-Schätzer nicht (siehe (a)), und auch hierfür ist die Wahrscheinlichkeit positiv:

$$\mathbb{P}(\overline{X_n} = 0) = \mathbb{P}(X_1 = 0)^n = e^{-n\lambda} > 0$$

Abgabe:
In Zweiergruppen, bis spätestens Montag, den **14. Dezember 2020, 09:00 Uhr**.

Homepage der Vorlesung:
`https://sip.math.uni-heidelberg.de/vl/ews-ws20/`