UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Einführung in die Wahrscheinlichkeitstheorie und Statistik
Prof. Dr. Rainer Dahlhaus
Nathawut Phandoidaen
Wintersemester 2019/2020

13. Übungsblatt – Lösungen

Aufgabe 49 (Maximum-Likelihood-Schätzer 1).

Seien $X_1, \dots, X_n \overset{\text{iid}}{\sim}$ Rayleigh(σ)-verteilt mit Parameter $\sigma > 0$ und Wahrscheinlichkeitsdichten

$$f_\sigma(x) = \frac{x}{\sigma^2} \cdot e^{-\frac{x^2}{2\sigma^2}} \mathbb{1}_{[0,\infty)}(x).$$

(a) Berechnen Sie den Maximum-Likelihood-Schätzer $\hat{\sigma}_n(X_1, \dots, X_n)$ von σ.

(b) Zeigen Sie, dass $\hat{\sigma}_n$ ein konsistenter Schätzer für σ ist.
Es gilt $\mathbb{E}[X_1] = \sigma\sqrt{\frac{\pi}{2}}$, $\operatorname{Var}(X_1) = \frac{4-\pi}{2}\sigma^2$ *und* $\mathbb{E}[X_1^4] = 8\sigma^4$.

(c) Berechnen Sie die Fisher-Informationsmatrix $I(\sigma)$ von σ auf Basis einer Beobachtung X_1.

(d) Zeigen Sie, dass $\hat{\sigma}_n$ ein asymptotisch Fisher-effizienter Schätzer ist.

Lösung:

(a) Wir betrachten die negative log-Likelihood

$$\begin{aligned} L_n(\sigma) &= -\frac{1}{n}\sum_{i=1}^n \log f_\sigma(x_i)) = -\frac{1}{n}\sum_{i=1}^n \Big[\log(x_i) - 2\log(\sigma) - \frac{x_i^2}{2\sigma^2}\Big] \\ &= -\frac{1}{n}\sum_{i=1}^n \log(x_i) + 2\log(\sigma) + \frac{1}{2\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^n x_i^2. \end{aligned}$$

Für den ML-Schätzer suchen wir die Extremstellen von L_n. Damit ergibt sich

$$0 \overset{!}{=} L_n'(\sigma) = \frac{2}{\sigma} - \frac{1}{\sigma^3}\frac{1}{n}\sum_{i=1}^n x_i^2 \quad\iff\quad \sigma_n = \sqrt{\frac{1}{2n}\sum_{i=1}^n x_i^2}.$$

Da $\mathbb{P}(X_1 = 0) = 0$, ist $\sigma_n > 0$ und wir erhalten

$$L_n''(\sigma) = -\frac{2}{\sigma^2} + \frac{3}{\sigma^4}\cdot\frac{1}{n}\sum_{i=1}^n x_i^2 \quad\Rightarrow\quad L_n''(\sigma_n) = -\frac{2n}{\sigma_n^2} + \frac{6n}{\sigma_n^2} = \frac{4n}{\sigma_n^2} > 0.$$

Daher minimiert σ_n die Likelihood-Funktion L_n. Da L_n stetig differenzierbar ist und keine weiteren Extremstellen besitzt, ist σ_n schon das globale Minimum. Es ist also

$$\hat{\sigma}_n = \sqrt{\frac{1}{2n}\sum_{i=1}^n X_i^2}.$$

(b) Nach dem schwachen Gesetz der großen Zahlen ($\mathbb{E}[X_1^4] < \infty$) gilt

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} \mathbb{E}[X_1^2] = \mathrm{Var}(X_1) + (\mathbb{E}[X_1])^2 = \frac{4-\pi}{2}\cdot\sigma^2 + \left(\sigma\cdot\sqrt{\frac{\pi}{2}}\right)^2 = 2\sigma^2.$$

Wir wollen nun das Continuous Mapping Theorem (CMT) bzw. die Proposition 14.9(iv) mit der in $2\sigma^2 > 0$ stetigen Funktion $g : \mathbb{R} \to \mathbb{R}, g(x) = \sqrt{\frac{1}{2}x}$ anwenden. Es folgt

$$\hat{\sigma}_n = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2} = g\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) \xrightarrow{p} g(2\sigma^2) = \sigma,$$

wobei wir g für $x < 0$ beliebig ergänzen. Also ist $\hat{\sigma}_n$ konsistent.

(c) Die Formel für die Fisher-Information lautet

$$\begin{aligned} I(\sigma) &= \mathbb{E}\left[\Big(\frac{\partial}{\partial\sigma}(\log f_\sigma(X_1))\Big)^2\right] \\ &= \mathbb{E}\left[\Big(-\frac{2}{\sigma} + \frac{X_1^2}{\sigma^3}\Big)^2\right] \overset{\mathbb{E}[X_1^2]=2\sigma^2}{=} \frac{1}{\sigma^6}\mathrm{Var}(X_1^2) = \frac{1}{\sigma^6}\Big(\mathbb{E}[X_1^4] - (\mathbb{E}[X_1^2])^2\Big) = \frac{4}{\sigma^2}. \end{aligned}$$

Mit $X_1 \sim \mathrm{Rayleigh}(\sigma)$ kann die Indikatorfunktion wegen $X_i \geq 0$ ignoriert werden.

(d) Wir zeigen zunächst, dass der ML-Schätzer σ_n asymptotisch normal ist. Daher nutzen wir den zentralen Grenzwertsatz

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mathbb{E}[X_1^2]}{\sqrt{\mathrm{Var}(X_1^2)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

bzw.

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}(X_1^2)) = \mathcal{N}(0, 4\sigma^4).$$

Der ZGWS (CLT) ist anwendbar, da wegen $X_1, ..., X_n$ iid auch $X_1^2, ..., X_n^2$ iid ist, und $\mathrm{Var}(X_1^2) < \infty$.
Nach der Delta-Methode (12. Übungsblatt, Aufgabe 46(c)), angewendet auf die in $c = 2\sigma^2$ stetig differenzierbaren Funktion $g : \mathbb{R} \to \mathbb{R}, g(x) = \sqrt{\frac{1}{2}x}$ mit $g'(c) = 2^{-3/2}c^{-1/2} = \frac{1}{4\sigma}$, gilt gerade

$$\sqrt{n}(\sigma_n - \sigma) = \sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - g(2\sigma^2)\right) \xrightarrow{\mathcal{D}} g'(c)\mathcal{N}(0, 4\sigma^4) = \mathcal{N}\Big(0, \underbrace{\frac{\sigma^2}{4}}_{=:v(\sigma)}\Big).$$

Damit entspricht die Varianz $v(\sigma)$ der asymptotischen Verteilung genau dem Inversen $I(\sigma)^{-1} = \frac{\sigma^2}{4}$ der Fisher-Informations-Matrix. Das bedeutet, dass σ_n asymptotisch Fisher-effizient ist.

Aufgabe 50 (Maximum-Likelihood-Schätzer 2).

Seien $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Bin}(m, p)$ unabhängige und identisch Binomial-verteilte Zufallsvariablen mit Parametern $m \geq 1$ und $p \in (0, 1)$. Die Zähldichte der Binomial-Verteilung mit Parametern (m, p) ist gegeben durch

$$p(k) = \binom{m}{k} p^k (1-p)^{m-k} \qquad \text{für } k \in \{0, \ldots, m\}.$$

Außerdem gilt $\mathbb{E}[X_1] = m \cdot p$ und $\mathrm{Var}(X_1) = m \cdot p \cdot (1 - p)$.

(a) Zeigen Sie, dass der Maximum-Likelihood-Schätzer $\hat{p}_n$ von p basierend auf $X_1, \ldots, X_n$ für bekanntes m durch $\hat{p}_n = \frac{1}{nm}\sum_{i=1}^n X_i$ gegeben ist.

(b) Zeigen Sie, dass $\hat{p}_n$ ein erwartungstreuer und konsistenter Schätzer für p ist.

(c) Weisen Sie nach, dass zudem $\sqrt{n}(\hat{p}_n - p) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \frac{p(1-p)}{m}\big)$ für $n \to \infty$ gilt.

Lösung:

(a) Wir betrachten die negative log-Likelihood

$$\begin{aligned} L_n(m,p) &= -\frac{1}{n}\sum_{i=1}^n \log p(x_i) = -\frac{1}{n}\sum_{i=1}^n \Big[\log(\binom{m}{x_i}) + \log(p)x_i + \log(1-p)(m-x_i)\Big] \\ &= -\frac{1}{n}\sum_{i=1}^n \log(\binom{m}{x_i}) - \log(p)\frac{1}{n}\sum_{i=1}^n x_i - \log(1-p)\cdot(m - \frac{1}{n}\sum_{i=1}^n x_i). \end{aligned}$$

Für den ML-Schätzer suchen wir die Extremstellen von L_n bei festem m und variablen p. Damit ergibt sich

$$0 \stackrel{!}{=} L_n'(m,p) = -\frac{1}{p}\frac{1}{n}\sum_{i=1}^n x_i + \frac{1}{1-p}(m - \frac{1}{n}\sum_{i=1}^n x_i) \qquad \Longleftrightarrow \qquad p_n = \frac{1}{mn}\sum_{i=1}^n x_i.$$

Ferner gilt mit $x := \sum_{i=1}^n x_i$,

$$L_n''(m,p) = \frac{x}{p^2} + \frac{1}{(1-p)^2}(m-x) > 0.$$

Alternativ kann argumentiert werden, dass L_n stetig ist und $\lim_{p\to 0} L_n(p) = \infty$ sowie $\lim_{p\to 1} L_n(p) = \infty$.
Daher minimiert σ_n die Likelihood-Funktion L_n. Da L_n stetig differenzierbar ist und keine weiteren Extremstellen besitzt, ist σ_n schon das globale Minimum.

(b) Da $\mathbb{E}[X_1] = mp$ und $\mathrm{Var}(X_1) = mp(1-p)$ folgt mit der Linearität des Erwartungswertes

$$\mathbb{E}[\hat{p}_n] = \frac{1}{nm}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{nm}nmp = p.$$

Damit ist $\hat{p}_n$ erwartungstreu für p.

Mit (X_i) iid, ist auch $\left(\frac{1}{m}X_i\right)$ iid. Weiter gilt $\mathbb{E}\left[\frac{1}{m^2}X_1^2\right] = \frac{1}{m^2}(n^2p^2 - np^2 + np) < \infty$ und $\mathrm{Var}\left(\frac{1}{m}X_i\right) = \frac{1}{m}(p(1-p)) < \infty$. Nach den schwachen Gesetz der großen Zahlen erhalten wir

$$\hat{p}_n = \frac{1}{nm}\sum_{i=1}^n X_i = \frac{1}{n}\sum_{i=1}^n \frac{1}{m}X_i \xrightarrow{p} \mathbb{E}\Big[\frac{1}{m}X_i\Big] = p,$$

und also die Konsistenz des Schätzers.

(c) Nach dem zentralen Grenzwertsatz mit $\mathbb{E}\left[\frac{1}{m}X_i\right] = p$ und $\mathrm{Var}\left(\frac{1}{m}X_i\right) = \frac{1}{m}(p(1-p))$ gilt

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{m}X_i - p}{\sqrt{p(1-p)/m}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Das CMT liefert

$$\sqrt{n}(\hat{p}_n - p) = \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n \frac{1}{m}X_i - p\Big) \xrightarrow{\mathcal{D}} \sqrt{p(1-p)/m}\,\mathcal{N}(0,1) = \mathcal{N}(0, p(1-p)/m).$$

Aufgabe 51 (Bed. Varianz und Projektionseigenschaft des bed. Erw.).
Seien X, Y reelle Zufallsvariablen und sei $\mathrm{Var}(Y) < \infty$. Die sogenannte bedingte Varianz von Y gegeben X ist definiert durch

$$\mathrm{Var}(Y|X) := \mathbb{E}[(Y - \mathbb{E}(Y|X))^2|X] = \mathbb{E}[Y^2|X] - \mathbb{E}[Y|X]^2.$$

(a) Zeigen Sie, dass
$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]).$$

(b) Zeigen Sie, dass der bedingte Erwartungswert $h^*(X) = \mathbb{E}[Y|X]$ den Abstand
$$\mathbb{E}[(Y - h(X))^2]$$
unter allen (messbaren) Funktionen $h : \mathbb{R} \to \mathbb{R}$ minimiert.

Lösung:

(a) Es ist

$$\begin{aligned}
& \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]) \\
= \quad & \mathbb{E}\big[\mathbb{E}[Y^2|X] - \mathbb{E}[Y|X]^2\big] + \mathbb{E}\big[\mathbb{E}[Y|X]^2\big] - \mathbb{E}\big[\mathbb{E}[Y|X]\big]^2 \\
= \quad & \mathbb{E}\big[\mathbb{E}[Y^2|X]\big] - \mathbb{E}\big[\mathbb{E}[Y|X]^2\big] + \mathbb{E}\big[\mathbb{E}[Y|X]^2\big] - \mathbb{E}\big[\mathbb{E}[Y|X]\big]^2 \\
\overset{\text{Satz 16.7 (i)}}{=} \quad & \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\
= \quad & \mathrm{Var}(Y)
\end{aligned}$$

(b) Es gilt

$$\begin{aligned}
& \mathbb{E}[(Y - h(X))^2] \\
= \; & \mathbb{E}[((Y - \mathbb{E}[Y|X]) + (\mathbb{E}[Y|X] - h(X)))^2] \\
= \; & \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + 2\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - h(X))] + \mathbb{E}[(\mathbb{E}[Y|X] - h(X))^2]
\end{aligned}$$

Für den mittleren Summanden gilt weiterhin

$$\begin{aligned}
& \mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - h(X))] \\
= \quad & \mathbb{E}\Big[\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - h(X))|X]\Big] \\
\overset{\text{Satz 16.7 (iii)}}{=} \quad & \mathbb{E}\Big[(\mathbb{E}[Y|X] - h(X))\mathbb{E}[(Y - \mathbb{E}[Y|X])|X]\Big] \\
\overset{\text{Satz 16.7 (iv)}}{=} \quad & \mathbb{E}\Big[(\mathbb{E}[Y|X] - h(X))(\mathbb{E}[Y|X] - \mathbb{E}[Y|X])\Big] \\
= \quad & 0.
\end{aligned}$$

Damit bleibt $\mathbb{E}[(Y - h(X))^2] = \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] + \mathbb{E}[(\mathbb{E}[Y|X] - h(X))^2]$. Allerdings wird der zweite Summand $\mathbb{E}[(\mathbb{E}[Y|X] - h(X))^2]$ durch $h(X) = \mathbb{E}[Y|X]$ minimiert. Der erste Summand ist unabhängig von $h(X)$. Damit ist $h^*(X) = \mathbb{E}[Y|X]$ der Minimierer für alle Funktionen h.

Aufgabe 52 (Bedingte Erwartungswerte für stetige und diskrete Verteilungen).
Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X, Y zwei Zufallsvariablen.

(a) Es habe (X, Y) im $\mathbb{R}^2$ die gemeinsame Dichte
$$f_{X,Y}(x, y) = e^{-y}\mathbb{1}_{\{0 \le x \le y\}}.$$
Berechnen Sie $\mathbb{E}[X|Y]$ und $\mathbb{E}[Y|X]$.

(b) Sei nun $r, p \in \mathbb{R}$ mit $0 \leq r \leq p \leq 1$ und $1 - 2p + r \geq 0$. Die Verteilungen von X und Y seien gegeben durch

$$\begin{aligned} \mathbb{P}(X = 1, Y = 1) &= r, \qquad \mathbb{P}(X = 0, Y = 1) = p - r, \\ \mathbb{P}(X = 1, Y = 0) &= p - r, \qquad \mathbb{P}(X = 0, Y = 0) = 1 - 2p + r. \end{aligned}$$

Berechnen Sie die Verteilung von Y und den bedingten Erwartungswert $\mathbb{E}[X|Y]$.

Lösung:

(a) Wir bestimmen die Randdichten f_X und f_Y, d.h.

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}y = \int_x^\infty e^{-y}\mathrm{d}y\, \mathbb{1}_{\{x \geq 0\}} = e^{-x}\mathbb{1}_{\{x \geq 0\}}$$

und

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}x = e^{-y}\int_0^y 1\,\mathrm{d}x\, \mathbb{1}_{\{y \geq 0\}} = ye^{-y}\mathbb{1}_{\{y \geq 0\}}.$$

Damit erhalten wir für die bedingten Dichten

$$f_{X|Y=y}(x) = \begin{cases} f_{X,Y}(x,y)/f_Y(y), & f_Y(y) > 0 \\ 0, & f_Y(y) = 0 \end{cases} = \begin{cases} \frac{e^{-y}\mathbb{1}_{\{0 \leq x \leq y\}}}{ye^{-y}}, & y \geq 0 \\ 0, & y < 0 \end{cases} = \frac{1}{y}\mathbb{1}_{\{0 \leq x \leq y\}}$$

und

$$f_{Y|X=x}(y) = \begin{cases} f_{X,Y}(x,y)/f_X(x), & f_X(x) > 0 \\ 0, & f_X(x) = 0 \end{cases} = \begin{cases} \frac{e^{-y}\mathbb{1}_{\{0 \leq x \leq y\}}}{e^{-x}}, & x \geq 0 \\ 0, & x < 0 \end{cases} = e^x e^{-y}\mathbb{1}_{\{0 \leq x \leq y\}}.$$

Für die bedingten Erwartungswerte berechnen wir

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} x f_{X|Y=y}(x)\mathrm{d}x = \int_{\mathbb{R}} x\frac{1}{y}\mathbb{1}_{\{0 \leq x \leq y\}}\mathrm{d}x = \frac{1}{y}\int_0^y x\mathrm{d}x\, \mathbb{1}_{\{y \geq 0\}} = \frac{y}{2}\mathbb{1}_{\{y \geq 0\}}$$

und

$$\begin{aligned} \mathbb{E}[Y|X = x] &= \int_{\mathbb{R}} y f_{Y|X=x}(y)\mathrm{d}y = \int_{\mathbb{R}} ye^x e^{-y}\mathbb{1}_{\{0 \leq x \leq y\}}\mathrm{d}y = e^x\int_x^\infty ye^{-y}\mathrm{d}y\, \mathbb{1}_{\{x \geq 0\}} \\ &= (x+1)\mathbb{1}_{\{x \geq 0\}}. \end{aligned}$$

Damit erhalten wir insgesamt

$$\mathbb{E}[X|Y] = \frac{Y}{2}\mathbb{1}_{\{Y \geq 0\}} \qquad \text{und} \qquad \mathbb{E}[Y|X] = (X+1)\mathbb{1}_{\{X \geq 0\}},$$

da $\mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y = Y(\omega)] = \frac{Y(\omega)}{2}\mathbb{1}_{\{Y(\omega) \geq 0\}}$ bzw. $\mathbb{E}[Y|X](\omega) = \mathbb{E}[Y|X = X(\omega)] = (X(\omega)+1)\mathbb{1}_{\{X(\omega) \geq 0\}}$.

(b) Wir haben

$$\mathbb{P}(Y = 0) = \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 1, Y = 0) = 1 - p,$$

weswegen $\mathbb{P}(Y = 1) = p$. Weiterhin folgt

$$\mathbb{E}[X|Y = 0] = 0 \cdot \mathbb{P}(X = 0|Y = 0) + 1 \cdot \mathbb{P}(X = 1|Y = 0) = \frac{\mathbb{P}(X = 1, Y = 0)}{\mathbb{P}(Y = 0)} = \frac{p - r}{1 - p}$$

und

$$\mathbb{E}[X|Y=1] = 0 \cdot \mathbb{P}(X=0|Y=1) + 1 \cdot \mathbb{P}(X=1|Y=1) = \frac{\mathbb{P}(X=1, Y=1)}{\mathbb{P}(Y=1)} = \frac{r}{p}.$$

Insgsamt gilt wegen $\mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y=Y(\omega)]$,

$$\mathbb{E}[X|Y] = \begin{cases} \frac{p-r}{1-p}, & Y=0 \\ \frac{r}{p}, & Y=1 \end{cases}.$$

Abgabe:
Dieser Zettel wird **nicht** abgegeben.

Homepage der Vorlesung:
`https://ssp.math.uni-heidelberg.de/ews-WS1920/index.html`