

Übungsblatt 11

Stochastik 2

Abgabe von: **Linus Mußmächer**

2. Juli 2023

11.1 Zentralübung

(i) Die gemeinsame Dichte der n Zufallsvariablen X_1, \dots, X_n ist

$$f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\theta} x_i^{\frac{1-\theta}{\theta}} \mathbb{1}_{(0,1)}(x_i)$$

Für $(x_1, \dots, x_n) \in (0, 1)^n = \mathcal{X}$ ist dann die Likelihood-Funktion

$$f(\theta, x_1, \dots, x_n) = \frac{1}{\theta^n} \left(\prod_{i=1}^n x_i \right)^{\frac{1-\theta}{\theta}}$$

und ihre Ableitung

$$\frac{\partial}{\partial \theta} f(\theta, x_1, \dots, x_n) = - \left(\prod_{i=1}^n x_i \right)^{\frac{1-\theta}{\theta}} \frac{1}{\theta^{n+2}} \left(n\theta + \sum_{i=1}^n \log(x_i) \right)$$

Diese hat die Nullstelle $n\theta = -\sum_{i=1}^n \log(x_i)$, also ist der ML-Schätzer gleich

$$\hat{\theta}^{ML}(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \log(1/x_i)$$

(ii) Der Erwartungswert von X_1 beträgt

$$\int_0^1 x \cdot \frac{1}{\theta} x^{\frac{1-\theta}{\theta}} dx = \frac{1}{\theta} \int_0^1 x^{\frac{1}{\theta}} dx = \frac{1}{\theta} \left[\frac{1+\theta}{\theta} x^{\frac{1}{1+\theta}} \right]_0^1 = \frac{1}{1+\theta}$$

und die Gleichsetzung $\frac{1}{1+\theta} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ liefert den Schätzer

$$\hat{\theta}(x_1, \dots, x_n) = \frac{n}{\sum_{i=1}^n x_i} - 1$$

(iii) Der ML-Schätzer liefert hier

$$\hat{\theta}^{ML}(0.1, 0.22, 0.54, 0.36) = \frac{1}{4} \log \left(\frac{1}{0.1 \cdot 0.22 \cdot 0.54 \cdot 0.36} \right) \approx 1.36$$

und aus dem Momentenschätzer erhält man

$$\hat{\theta}(0.1, 0.22, 0.54, 0.36) = \frac{4}{0.1 + 0.22 + 0.54 + 0.36} - 1 = \frac{4}{1.22} - 1 \approx 2.27$$

11.2

Wir setzen das empirische erste und zweite Moment mit dem ersten und zweiten Moment der Log-Normalverteilung gleich (man beachte, dass n in der Aufgabenstellung doppelt verwendet wurde, als Anzahl der Y_i und zur Indizierung der Momente. Wir übernehmen nur die erste Verwendung):

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n x_i = \exp\left(\mu + \frac{\sigma^2}{2}\right) \Leftrightarrow \mu + \frac{1}{2}\sigma^2 = \log\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \quad (I)$$

$$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = \exp(2\mu + 2\sigma^2) \Leftrightarrow 2\mu + 2\sigma^2 = \log\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) \quad (II)$$

$(II) - 2(I)$ liefert

$$\begin{aligned} \sigma^2 &= \log\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) - \log\left(\left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right) = \log\left(\frac{n \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i\right)^2}\right) \\ &= \log(\hat{\mu}_2) - \log(\hat{\mu}_1^2) = \log\left(\frac{\hat{\mu}_2}{\hat{\mu}_1^2}\right) \end{aligned}$$

und damit einen Schätzer für σ^2 ; und aus $2(I) - 4(II)$ erhalten wir

$$\begin{aligned} \mu &= \log\left(\left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2\right) - \log\left(\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right)^4\right) = 2 \log\left(\frac{n \sum_{i=1}^n x_i}{\left(\sum_{i=1}^n x_i^2\right)^2}\right) \\ &= \log(\hat{\mu}_1^2) - \log(\hat{\mu}_2^4) = \log\left(\frac{\hat{\mu}_1^2}{\hat{\mu}_2^4}\right) = 2 \log\left(\frac{\hat{\mu}_1}{\hat{\mu}_2^2}\right) \end{aligned}$$

11.3

(i) Wir bestimmen die Ableitung durch die Produktregel:

$$\begin{aligned} \frac{\partial}{\partial \lambda} F_x(\lambda) &= e^{-\lambda} \sum_{k=0}^x \frac{k \cdot \lambda^{k-1}}{k!} + (-1) \exp(-\lambda) \sum_{k=0}^x \frac{\lambda^{k-1}}{k!} = e^{-\lambda} \sum_{k=0}^{x-1} \frac{\lambda^k}{k!} - \exp(-\lambda) \sum_{k=0}^x \frac{\lambda^{k-1}}{k!} \\ &= - \underbrace{e^{-\lambda}}_{>0} \underbrace{\frac{\lambda^x}{x!}}_{\geq 0} < 0 \end{aligned}$$

und somit ist $F_x(\lambda)$ für alle $x \in \mathbb{N}_0$ streng monoton steigend auf $(0, \infty)$. Weiterhin halten wir fest, dass $F_x(\lambda) = \mathbb{P}_\lambda\{X \leq x\}$ für eine zum Parameter λ Poisson-verteilte Zufallsvariable X gilt.

(ii) Wir wollen Satz 4.16 verwenden. Dann gilt:

$$\begin{aligned} n_r &= \min\left(n_0 \in \mathbb{N} \mid \sup_{\lambda \leq \lambda_0} \mathbb{P}_\lambda(\{n \in \mathbb{N}_0 \mid n \geq n_0\}) \leq \alpha\right) \\ &= \min\left(n_0 \in \mathbb{N} \mid \sup_{\lambda \leq \lambda_0} 1 - \mathbb{P}_\lambda(\{n \in \mathbb{N}_0 \mid n \leq n_0 + 1\}) \leq \alpha\right) \end{aligned}$$

$$\begin{aligned}
&= \min \left(n_0 \in \mathbb{N} \mid \sup_{\lambda \leq \lambda_0} 1 - F_{n_0+1}(\lambda) \leq \alpha \right) \quad | \text{ mit } F_{n_0+1} \text{ s.m.f.} \\
&= \min(n_0 \in \mathbb{N} \mid 1 - F_{n_0+1}(\lambda_0) \leq \alpha) = \min(n_0 \in \mathbb{N} \mid F_{n_0+1}(\lambda_0) \geq 0.95).
\end{aligned}$$

Da $F_{n_0+1}(\lambda_0)$ in n_0 streng monoton steigend ist, müssen wir lediglich eine Zahl $r \in \mathbb{R}$ mit $F_{r+1}(\lambda_0) = 0.95$ bestimmen und aufrunden, um $n_r = \lceil r \rceil$ zu erhalten. Dies ist allerdings nur numerisch möglich.

Für dieses n_r setzen wir dann $T = \mathbb{1}_{[n_r, \infty]}$ als unseren Test. Nach Satz 4.16 hat er Niveau α .

- (iii) Für $\lambda_0 = 10$ bestimmen wir numerisch $F_{r+1}(10) = 0.95 \Leftrightarrow r \approx 13.95$, also wählen wir $n_r = 14$ und erhalten einen Verwerfungsbereich von $[0, 13]$.

Für $\lambda_0 = 1$ erhalten wir analog $F_{r+1}(1) = 0.95 \Leftrightarrow x \approx 1.35$, also $n_r = 2$ und damit den Verwerfungsbereich $\{0, 1\}$.

11.4

- (i) Wir setzen $\mathcal{X} = \{0, 1\}$ (Antwort falsch, Antwort richtig) und betrachten \mathcal{X}^4 mit der Statistik $S : \mathcal{X}^4 \rightarrow \mathbb{R}, (x_1, x_2, x_3, x_4) \mapsto \sum_{k=1}^4 x_k$. Diese ist dann binomialverteilt zu einem Parameter $p \in [0, 1] = \Theta$, d.h. $S \sim \text{Bin}_{4,p}$.

Als Nullhypothese könnten wir nun $H_0 : p \in \Theta_0 = [0, 1/4]$, d.h. der Kandidat ist unqualifiziert und rät nur. Kurioserweise hat diese Versicherung die interessante Definition von "qualifiziert" als "besser als reines Raten". Dies ist nicht anspruchsvoll, und ich möchte mich bitte anderswo versichern lassen.

In diesem Beispiel wäre dann die Gegenhypothese $K : p \in (1/4, 1]$ und ein Test wäre eine Abbildung $T_4 : \mathcal{X}^4 \rightarrow \{0, 1\}$. Ein solcher Test ließe sich beispielsweise nach Wahl eines Signifikanzniveaus mittels Satz 4.16 konstruieren. Möchte man ein Signifikanzniveau $\alpha = 0$ sicherstellen, d.h. unter keinen Umständen einen unqualifizierten Bewerber einstellen, so bietet sich die Wahl von $T_4(x_1, x_2, x_3, x_4) = 1 \Leftrightarrow x_k = 1 \forall_{k=1, \dots, 4}$ an.

- (ii) Beim Fehler erster Art wird die Nullhypothese verworfen (d.h. der Kandidat als qualifiziert eingestuft und eingestellt), obwohl sie eigentlich erfüllt ist (der Kandidat rät nur). Im Fall, dass der Kandidat wirklich nur rät ist die Wahrscheinlichkeit hierfür laut Bernoulli

$$\mathbb{P}(S \geq 3) = p_{4,1/4}(3) + p_{4,1/4}(4) = \binom{4}{3} \frac{1}{4^3} \frac{3}{4} + \frac{1}{4^4} = \frac{13}{256} \approx 0.051.$$

Falls der Kandidat so wenig Ahnung hat, dass er des Öfteren falsche Antworten für richtig hält, dann ist die Wahrscheinlichkeit noch kleiner.

- (iii) Wie wir in (b) gesehen haben, dass beim Verlangen dreier richtiger Antworten immer noch eine Chance $> 5\%$ besteht, dass ein nur-ratender Kandidat akzeptiert wird, muss ein Test zum Signifikanzniveau 0.05 schärfer sein, d.h. wir müssen T wie in (a) wählen als $T(x_1, x_2, x_3, x_4) = 1 \Leftrightarrow x_k = 1 \forall_{k=1, \dots, 4}$.
- (iv) Dies entspricht der Wahrscheinlichkeit dafür, dass der Statistiker keine, eine oder zwei Antworten korrekt beantwortet bei $p = 0.75$. Bernoulli:

$$\mathbb{P}(S < 3) = p_{4,3/4}(0) + p_{4,3/4}(1) + p_{4,3/4}(2) = \frac{1}{4} + 4 \cdot \frac{3}{4} \cdot \frac{1}{4} + \binom{4}{2} \frac{3^2}{4} \frac{1}{4} = \frac{31}{256} \approx 0.121.$$

Etwas mehr Vorbereitung kann also nicht schaden.