Lösungsvorschlag zur Nachklausur in
Stochastik und Statistik
Sommersemester 2017

long.xingyu@campus.lmu.de

Letzte Änderung: 16. Juli 2019

Alle Angaben ohne Gewähr

Aufgabe 1

(a)

Zeitschrift	Nord (N)	Süd (S)	Summe
A	180.625	159.375	340
B	180.625	159.375	340
C	148.75	131.25	280
Summe	510	450	960

(b) $k = (r-1) \cdot (c-1) = (2-1) \cdot (3-1) = 2$
Um die Unabhängigkeit zu testen werden insgesamt $r \cdot c$ Kategorien berücksichtigt. Durch Bildung der Abweichungen von Erwartungswerten verlieren beide Variablen jeweils einen Freiheitsgrad (analog zur Bessel-korrigierten Stichprobenvarianz): $\implies k = (r-1) \cdot (c-1)$

(c) $\chi^2_{0.99,2} \approx 9.21 < 31.8 = \chi^2 \implies$ Ja, die Nullhypothese wird abgelehnt: Bevorzugte Zeitschrift ist *nicht* unabhängig von der Region des Käufers.

(d) Devianz: D folgt der χ^2-Verteilung.

Aufgabe 2

Seien Zufallsvariablen X für Testergebnis, Y für Genotyp, Z für Gesundheitsstand.

(a)
$$\begin{aligned} P(X = \text{pos}) &\overset{\text{(Totale Wkt)}}{=} P(Y = \text{aa}) \cdot P(X = \text{pos}|Y = \text{aa}) + \\ &\quad P(Y = \text{ab}) \cdot P(X = \text{pos}|Y = \text{ab}) + \\ &\quad P(Y = \text{bb}) \cdot P(X = \text{pos}|Y = \text{bb}) \\ &= 0.5 \cdot 0.05 + 0.45 \cdot 0.1 + 0.05 \cdot 0.95 \\ &= 0.1175 \end{aligned}$$

(b)
$$\begin{aligned} P(Y = \text{bb}|X = \text{pos}) &\overset{\text{(Bayes)}}{=} \frac{P(Y = \text{bb}, X = \text{pos})}{P(X = \text{pos})} \overset{\text{(Bayes)}}{=} \frac{P(Y = \text{bb}) \cdot P(X = \text{pos}|Y = \text{bb})}{P(X = \text{pos})} \\ &= \frac{0.05 \cdot 0.95}{0.1175} \\ &\approx 0.404 \end{aligned}$$

(c) $$P(Z=\text{krank}|X=\text{pos}) \overset{\text{(Bayes)}}{=} \frac{P(Z=\text{krank}, X=\text{pos})}{P(X=\text{pos})} \overset{\text{(Unabh.)}}{=} \frac{P(Z=\text{krank})\cdot P(X=\text{pos})}{P(X=\text{pos})} = P(Z=\text{krank}) = 0.3$$

Aufgabe 3

(a) Likelihoodfunktion: $$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} (\theta-1)x_i^{-\theta} I_{[1,\infty)}(x_i) = \prod_{i=1}^{n}(\theta-1)x_i^{-\theta} \text{ mit } x_i \in [1,\infty)$$

Loglikelihoodfunktion: $$l(\theta) = \ln L(\theta) = \ln \prod_{i=1}^{n}(\theta-1)x_i^{-\theta} = \sum_{i=1}^{n} \ln(\theta-1)x_i^{-\theta} = n\ln(\theta-1) + \sum_{i=1}^{n}\ln x_i^{-\theta}$$
$$= n\ln(\theta-1) - \theta\sum_{i=1}^{n}\ln x_i$$

Erste Ableitung $\overset{!}{=} 0$: $$\frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta) = \frac{n}{\theta-1} - \sum_{i=1}^{n}\ln x_i \overset{!}{=} 0 \implies \theta = \frac{n}{\sum_{i=1}^{n}\ln x_i} + 1$$

Zweite Ableitung < 0: $$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\theta) = -\frac{n}{(\theta-1)^2} < 0 \text{ gilt stets wegen } n > 0$$

$\implies$ ML-Schätzer: $$\hat{\theta}_{\text{ML}} = \frac{n}{\sum_{i=1}^{n}\ln x_i} + 1$$

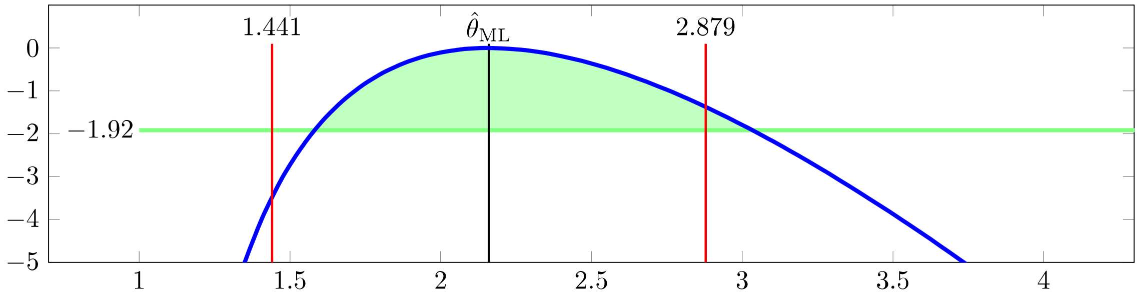
(b) Zweite Ableitung: $$l''(\hat{\theta}_{\text{ML}}) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\theta) \,\Big|\, \theta = \hat{\theta}_{\text{ML}}$$
$$= -\frac{n}{(\theta-1)^2} \,\Big|\, \theta = \frac{n}{\sum_{i=1}^{n}\ln x_i} + 1$$
$$= -\frac{\left(\sum_{i=1}^{n}\ln x_i\right)^2}{n}$$

$\implies$ Standardfehler: $$\text{SE}(\hat{\theta}_{\text{ML}}) = \sqrt{\left[-l''(\hat{\theta}_{\text{ML}})\right]^{-1}} = \sqrt{\frac{n}{\left(\sum_{i=1}^{n}\ln x_i\right)^2}} = \frac{\sqrt{n}}{\sum_{i=1}^{n}\ln x_i}$$

(c) $(1-a) = 95\% \implies a = 5\% \implies d = (1-\frac{5\%}{2})$-Quantil der Standardnormalverteilung
$= 0.975$-Quantil der Standardnormalverteilung
≈ 1.96

$$\text{95\%-Wald-Konfidenzintervall} = \left[\hat{\theta}_{\text{ML}} - d\cdot\text{SE}(\hat{\theta}_{\text{ML}}),\ \hat{\theta}_{\text{ML}} + d\cdot\text{SE}(\hat{\theta}_{\text{ML}})\right] \Big|\ n=10, \sum_{i=1}^{10}\ln x_i = 8.62$$
$$\approx \left[\frac{n-1.96\cdot\sqrt{n}}{\sum_{i=1}^{n}\ln x_i} + 1,\ \frac{n+1.96\cdot\sqrt{n}}{\sum_{i=1}^{n}\ln x_i} + 1\right] \Big|\ n=10, \sum_{i=1}^{10}\ln x_i = 8.62$$
$$\approx [1.441,\ 2.879]$$

(d) Die Loglikelihoodfunktion ist stark asymmetrisch, während das approximative Wald-Intervall auf Symmetrie um den ML-Schätzer setzt.



In der konkreten Datensituation ragt die untere Grenze des 95%-Wald-Intervalls über die untere Grenze des 95%-Likelihood-Intervalls ($\tilde{l}(\theta) \geq -1.92$) hinaus, was die Vertrauenswürdigkeit verletzt.

Aufgabe 4

(a) $\displaystyle\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_0^1 [\lambda + c\cdot(1-\lambda)\cdot x]\,\mathrm{d}x = \left[\lambda\cdot x + \frac{c}{2}\cdot(1-\lambda)\cdot x^2\right]_{x=0}^{x=1} = \lambda + \frac{c}{2}\cdot(1-\lambda) \overset{!}{=} 1$

$$\text{Falls } \lambda = 1: \; 1 + \frac{c}{2}\cdot 0 = 1 \implies c \in \mathbb{R}$$

$$\text{Falls } \lambda \neq 1: \; \frac{c}{2} = \frac{1-\lambda}{1-\lambda} \implies c = 2$$

$$\implies c = 2 \text{ erfüllt } \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \text{ für alle } \lambda \in \mathbb{R}.$$

(b) $\displaystyle\mathbb{E}(X) = \int_{-\infty}^{+\infty} x\cdot f(x)\mathrm{d}x = \int_0^1 x\cdot[\lambda + 2\cdot(1-\lambda)\cdot x]\,\mathrm{d}x = \left[\frac{\lambda}{2}\cdot x^2 + \frac{2\cdot(1-\lambda)}{3}\cdot x^3\right]_{x=0}^{x=1} = \frac{2}{3} - \frac{\lambda}{6}$

(c)
$$\begin{aligned}\mathbb{E}(\hat{\lambda}) = \mathbb{E}\left(4 - \frac{6}{n}\sum_{i=1}^n X_i\right) &= 4 - \frac{6}{n}\sum_{i=1}^n \mathbb{E}(X_i)\\ &\overset{(X_i \text{ i.i.d.})}{=} 4 - \frac{6}{n}\cdot\left[n\cdot\left(\frac{2}{3} - \frac{\lambda}{6}\right)\right]\\ &= \lambda\end{aligned}$$

$$\implies \mathbb{E}(\hat{\lambda}) = \lambda \implies \hat{\lambda} \text{ ist erwartungstreu.}$$

(d) X ist eine stetige Zufallsvariable $\implies \forall x \in \mathbb{R} : P(X = x) = 0$.

Aufgabe 5

(a) Es gibt zwei verschiedene Interpretationen der Aussage „*In diesem Fall bleibt das Guthaben bei 0 Euro und das Spiel ist beendet*“. Die daraus resultierende Matrix kann entsprechend entweder stochastisch oder substochastisch sein. In Anbetracht der Teilaufgabe (c) nehme ich hier an, dass es sich dabei um den ersteren Fall handelt.

$$\text{Übergangsmatrix: } P = \begin{array}{c} \\ 0\\1\\2\\3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 2/3 & 1/3 & 0 & 0\\ 1/3 & 1/3 & 1/3 & 0\\ 0 & 1/3 & 1/3 & 1/3\\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}\end{array}$$

(b) $\mu^0 = (0,1,0,0)$

$$\mu^1 = \mu^0 \cdot P = (0,1,0,0) \cdot \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0\right)$$

$$\mu^2 = \mu^1 \cdot P = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0\right) \cdot \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix} = \left(\frac{1}{3}, \frac{1}{3}, \frac{2}{9}, \frac{1}{9}\right)$$

(c) $$\pi_1 \cdot P = \left(\frac{1}{2}, 0, 0, \frac{1}{2}\right) \cdot \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix} = \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}\right) \neq \pi_1 \implies \pi_1 \text{ ist nicht stationär}$$

$$\pi_2 \cdot P = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \cdot \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) = \pi_2 \implies \pi_2 \text{ ist stationär}$$

Bemerkung: Mit der substochastischen Übergangsmatrix ist weder π_1 noch π_2 stationär.

Aufgabe 6

Lineare Regression wurde im SoSe 2019 nicht behandelt sry