Institut für Angewandte Mathematik
SS 2017
Patrik Ferrari, Eva Kopfer & Angelo Profeta

10. Übungsblatt „Einführung in die Statistik"

Abgabe: Mittwoch 12.7.2017 in der Vorlesung

---

1. (Fischers exakter Test auf Unabhängigkeit) *[8 Pkt]*

Betrachten Sie das unendliche Produktmodell für den Chiquadrat-Unabhängigkeitstest, d.h. $(E^{\mathbb{N}}, \mathcal{P}(E)^{\otimes \mathbb{N}}, \vartheta^{\otimes \mathbb{N}} : \vartheta \in \Theta)$ mit $E = A \times B$ für $A = B = \{1, 2\}$, wobei Θ eine Teilmenge der Wahrscheinlichkeitsmaße auf E ist

$$\Theta \;=\; \left\{ \vartheta \in (0,1)^E \mid \textstyle\sum_{(i,j)\in E} \vartheta(i,j) = 1 \right\} \;\subset\; \mathcal{M}_1(E).$$

a) Zeigen Sie, dass genau dann $\vartheta = \vartheta^A \otimes \vartheta^B$ gilt, wenn $\vartheta(1,1) = \vartheta^A(1)\vartheta^B(1)$ ist.

b) Zeigen Sie, dass für alle $n \in \mathbb{N}, k, n_A, n_B \in \mathbb{Z}_+$ und $\vartheta \in \Theta_0$ gilt

$$\begin{aligned} \mathrm{P}_\vartheta\left[h_n(1,1) = k \,\middle|\, h_n^A(1) = n_A, h_n^B(1) = n_B \right] \;&=\; \mathrm{Hyp}(n_B, n_A, n - n_A)\big[\{k\}\big] \\ &=\; \mathrm{Hyp}(n_A, n_B, n - n_B)\big[\{k\}\big]. \end{aligned}$$

c) Entwickeln Sie zu einem vorgegebenen Irrtumsniveau α einen Test der Nullhypothese $H_0 : \vartheta = \vartheta^A \otimes \vartheta^B$ gegen die Alternative $H_1 : \vartheta \neq \vartheta^A \otimes \vartheta^B$.

d) In dem Artikel von Kuss et. al.

> The fouled player should not take the penalty himself: an empirical investigation of an old German football myth, *J. Sports Science*, **25**, 963-967, 2007

wird über die Ausgang von Strafelfmetern in der 1.ten Fußballbundesliga im Zeitraum von August 1993 bis Februar 1995 berichtet:

	verwandelt	gehalten	Total
Gefoulte schießt selbst	74	28	102
anderer Spieler schießt	547	186	733
Total	621	214	835

Stützen die Daten die These, dass der Gefoulte den Elfmeter nicht selbst schießen soll? Führen Sie dazu sowohl den Fischer-Test als auch den Chiquadrat-Test durch.

2. (Simpson Paradoxon) [*4 Pkt*]

Die folgende Tabelle zeigt die Studiendauer (in Semestern) sowie das Einstiegsgehalt (in 1000 Euro) der Absolventen eines Jahres am Fachbereich Mathematik und Informatik der (hypothetischen) Yule-Universität:

Semester	12	14	16	12	15	14	13	14	11	13	10	12	14
Gehalt	44.4	43.2	42.4	44.5	36.8	40.3	44.1	40.2	42.9	40.7	46	44.9	38.2
Semester	9	11	9	9	12	13	11	10	10	10	9	10	12
Gehalt	33.7	35.9	36.1	34.2	29.9	31.9	33.3	36.2	33.8	32.9	33.3	35.1	34.2

a) Schlägt sich für diese Absolventen ein längeres Studium in einem höheren Anfangsgehalt nieder? Bestimmen Sie die Regressionsgerade für Studiendauer gegen Anfangsgehalt.

b) Wie ändert sich Ihr Befund, wenn Sie zusätzlich erfahren, dass die oberen beiden Zeilen der Tabelle sich auf die Absolventen des Fachs Informatik, die unteren beiden sich auf die Absolventen des Fachs Mathematik beziehen? Führen Sie jeweils eine Regression innerhalb der beiden Gruppen durchführen.

c) Stellen Sie sowohl die Punktwolke der gegebenen Daten wie auch die Regressionsgraden aus den *Aufgabenteilen a)* und *b)* graphisch dar.

3. (Autoregressives Modell) [*6 Pkt*]

Zur Beschreibung zeitlicher Entwicklungen mit deterministischer Wachstumstendenz und zufälligen Störungen verwendet man oft das folgende autoregressive Model:

$$X_k \;=\; \gamma\, X_{k-1} + \sqrt{v}\,\xi_k, \qquad 1 \le k \le n.$$

Dabei sind $\gamma \in \mathbb{R}$ und $v > 0$ unbekannte Parameter, $X_0, \dots, X_n$ die Beobachtungen und $\xi_1, \dots, \xi_n$ unabhängige zufällige Störungen mit $\mathrm{E}[\xi_k] = 0, \mathrm{Var}[\xi_k] = 1$.

a) Machen Sie einen Ansatz für den quadratischen Fehler, und bestimmen Sie den Kleinste-Quadrate-Schätzer $\hat{\gamma}$ für γ. Ist dieser Schätzer erwartungstreu?

b) Sei nun $X_0 = 0$, und die Fehlervariablen ξ_k seien standardnormalverteilt. Zeigen Sie, dass die Likelihood-Funktion des Modells gegeben ist durch

$$\varrho\big(X, (\gamma, v)\big) \;=\; (2\pi v)^{-n/2} \exp\Big(-\frac{1}{2v}\sum_{k=1}^{n}(X_k - \gamma\, X_{k-1})^2\Big).$$

c) Betrachten Sie das Testproblem $H_0 : \gamma = 0$ (*keine Abhängigkeit*) gegen $H_1 : \gamma \neq 0$. Zeigen Sie unter den Vorraussetzungen des *Aufgabenteils b)*, dass der Likelihood-Quotient eine monotone Funktion des Betrags der Stichprobenkorrelation

$$\hat{r}(X) \;=\; \frac{\sum_{k=1}^{n} X_k\, X_{k-1}}{\sqrt{\big(\sum_{k=1}^{n} X_k^2\big)\,\big(\sum_{k=1}^{n} X_{k-1}^2\big)}}$$

ist.