Prof. Dr. E. Spodarev
D. Meschenmoser

Wintersemester 2009 / 2010
05. November 2009

Statistik II – Übungsblatt 4

Abgabe: 12. November 2009, vor den Übungen

Aufgabe 1 (4)

Es sei eine Stichprobe $(X_1, \ldots, X_n)$ von unabhängigen Zufallsvariablen gegeben mit $X_i \sim \mathrm{Bin}(k, p)$, wobei $k \in \mathbb{N}$ bekannt sei. Konstruiere einen besten Test für $H_0 : p \geq p_0$ gegen $H_1 : p < p_0$ zum Niveau $\alpha \in (0, 1)$.

Aufgabe 2

Bei vielen praktischen Problemen ist die Verteilung von auftretenden Zufallsgrößen unbekannt oder analytisch schwer in den Griff zu kriegen. Um hier trotzdem Tests durchführen zu können, werden oft so genannte Monte-Carlo-Methoden angewendet. Hier wird per Simulation die Verteilung approximiert, ohne Quantile o.ä. analytisch berechnen zu müssen.

a) Zeige: Ist $(X_1, \ldots, X_n)$ eine Zufallsstichprobe von unabhängigen und identisch verteilten absolutstetigen Zufallsvariablen, so ist $\mathbb{P}(\mathrm{rg}(X_1) \leq m) = \frac{m}{n}$. (3)
Hinweis: $\mathrm{rg}(X_1)$ ist der Rang von X_1, also die Anzahl der Zufallsvariablen in der Stichprobe, die kleiner oder gleich X_1 sind. Außerdem gilt: $(X_1, \ldots, X_n) \overset{d}{=} \pi(X_1, \ldots, X_n)$ für alle Permutationen π von $(X_1, \ldots, X_n)$.

b) Nun soll mit Hilfe des Rangs ein Test auf Gleichheit der Verteilung konstruiert werden. Es sei eine Zufallsstichprobe $X_1, \ldots, X_n$ von absolutstetigen unabhängigen Zufallsvariablen mit einer Verteilung F gegeben, und noch eine weitere davon unabhängige Zufallsvariable Y, für die überprüft werden soll, ob sie auch die Verteilung F hat. Dabei soll die Nullhypothese abgelehnt werden, wenn Y signifikant größer ist als die Stichprobenwerte $X_1, \ldots, X_n$. Wie sieht der Test zum Niveau α aus? (3)

c) Nun wende den Test aus Teil b) auf folgendes Problem an: Bei einem Geysir wird davon ausgegangen, dass die Höhe der Wasserfontänen ungefähr exponentialverteilt ist mit einem Erwartungswert von 10 m. Da die Messung der Fontänenhöhen allerdings sehr schwierig ist, kommt ein Messfehler dazu. Von diesem wird angenommen, dass er ungefähr normalverteilt ist mit $\mu = 0\,\mathrm{m}$ und $\sigma^2 = 0.25\,\mathrm{m}^2$. Er ist von der Fontänenhöhe unabhängig. Eine Messung hat eine Höhe von 11 m ergeben. Stammt diese Messung vermutlich aus der beschriebenen Messreihe, oder ist der Wert zu groß? Betrachte das Niveau $\alpha = 0.05$ und $n = 99$. (2)
Hinweis: Siehe Abschnitt „Auswählen von Teilvektoren" im R-Skript, um den Rang mit R zu bestimmen.

Aufgabe 3

Betrachte eine Exponentialfamilie Q_θ und einen gemäß Q_θ verteilten Zufallsvektor $X = (X_1, \ldots, X_n)$. Also hat die Dichte die Form

$$f_\theta(x) = \exp\{c(\theta) T(x) + a(\theta)\} l(x)$$

für $x = (x_1, \ldots, x_n) \in B$. Zusätzlich gelte: $c(\theta) = \theta$, $a(\theta)$ ist stetig differenzierbar für alle $\theta \in \Theta$ und $0 < \mathrm{Var}_\theta T(X_1, \ldots, X_n) < \infty$.

a) Beweise folgende Aussage, die für Teil b) und c) benötigt wird: Besitzt X die μ-Dichte einer einparametrigen Exponentialfamilie mit den oben genannten Eigenschaften, und ist φ eine reellwertige (messbare) Funktion auf B mit $\mathbb{E}_\theta(|\varphi(X)|) < \infty$ für alle $\theta \in \Theta$, so ist die Funktion (5)

$$A_\varphi(\theta) = \int_B \varphi(x) \exp(\theta T(x)) l(x)\, dx, \quad \theta \in \Theta$$

beliebig oft in θ differenzierbar, und die Differentiation kann unter dem Integralzeichen vorgenommen werden.

b) Zeige, dass $a(\theta)$ beliebig oft differenzierbar ist. (2)

c) Berechne den Erwartungswert von $T(X)$. (2)

Aufgabe 4

Sei $Z = (Z_1, \ldots, Z_r) \sim M_{r-1}(n, p)$, also multinomialverteilt mit Parametern $n \geq 1$ und $p = (p_1, \ldots, p_{r-1})$, mit $p_i > 0$, d.h. für alle $k_1, \ldots, k_r \in \mathbb{N}_0$ mit $k_1 + \ldots + k_r = n$ gilt

$$\mathbb{P}(Z_1 = k_1, \ldots, Z_r = k_r) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r}$$

a) Zeige: Die charakteristische Funktion φ_Z von Z ist gegeben durch (2)

$$\varphi_Z(t) = \left(\sum_{j=1}^{r} p_j e^{it_j} \right)^n, \quad t = (t_1, \ldots, t_r),$$

wobei $p_r = 1 - \sum_{i=1}^{r-1} p_i$.
Hinweis: Der polynomiale Satz lautet: Für $a_1, \ldots, a_k \in \mathbb{C}$ und $l \in \mathbb{N}$ gilt

$$(a_1 + \ldots + a_k)^l = \sum_{\substack{i_1 + \ldots + i_k = l \\ i_j \geq 0}} \frac{l!}{i_1! \cdots i_k!} a_1^{i_1} \cdots a_k^{i_k}.$$

b) Es seien nun $Y_i = (Y_{i1}, \ldots, Y_{ir}) \sim M_{r-1}(1, p)$ mit $p = (p_1, \ldots, p_{r-1})$, $p_i > 0$ und $i = 1, \ldots, n$ unabhängige Zufallsvektoren. Zeige, dass gilt (2)

$$\sum_{j=1}^{n} Y_j \sim \mathrm{M}_{r-1}(n, p).$$

c) Zeige, dass $\mathbb{E}Z = n(p_1, \ldots, p_r)$. (2)