Übungen zur Vorlesung

Mathematische Statistik

Sommersemester 2012

Institut für Mathematik

Jun.-Prof. Dr. Thorsten Dickhaus
RUD25, Raum 1.203
E-Mail: dickhaus@math.hu-berlin.de

Semesterwoche 11

Abgabe bis Montag, 09. Juli 2012, 13:15 Uhr
Jede komplett richtig gelöste Aufgabe ergibt 4 Punkte.

Aufgaben

1. (a) Komplettieren Sie den Beweis von Resultat [R1] im Beweis von Theorem 12.2.3 im Buch von Lehmann und Romano (2005). D. h., zeigen Sie, dass unter den dort gemachten Bezeichnungen und Annahmen gilt:

$$4\int_\Omega \left| \left\langle \frac{\partial}{\partial \vartheta} \sqrt{\ell(\vartheta, x)}\Big|_{\vartheta=\vartheta_0}, h \right\rangle \right|^2 d\mu(x) = \langle h, I(\vartheta_0)h \rangle.$$

   (b) Die Hellinger-Distanz $H(P_0, P_1)$ von zwei Wahrscheinlichkeitsmaßen P_0 und P_1 auf $(\Omega, \mathcal{F})$, welche beide von dem Maß μ dominiert sind, ist gegeben durch

$$H^2(P_0, P_1) = \frac{1}{2}\int_\Omega \left[\sqrt{p_1(x)} - \sqrt{p_0(x)}\right]^2 d\mu(x),$$

   wobei p_i eine Version der μ-Dichte von P_i bezeichnet, $i = 0, 1$. Zeigen Sie: Für eine L_2-differenzierbare Familie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ gilt, dass $nH^2(\mathbb{P}_{\vartheta_0+hn^{-1/2}}, \mathbb{P}_{\vartheta_0+h_n n^{-1/2}}) \to 0$, falls die Folge $\{h_n\}_{n\geq 1}$ gegen h konvergent ist.

2. Wir betrachten noch einmal das Gaußmodell $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), ((\mathcal{N}(\mu, \sigma^2))^n)_{\vartheta=(\mu,\sigma^2)\in\Theta=\mathbb{R}\times(0,\infty)})$. Zeigen Sie:

   (a) Der Likelihood Ratio-Test für das einseitige Testproblem $H_0 = \{\mu \leq 0\}$ gegen $H_1 = \{\mu > 0\}$ lehnt ab, falls

$$\frac{\sqrt{n}\bar{x}_n}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2}} > c(\alpha).$$

   Der kritische Wert $c(\alpha)$ wird dabei wie folgt berechnet: Es sei $k > 1$ der kritische Wert für $\Lambda(x)$. Setzen wir $K := k^{-2/n}$, so ist $c(\alpha) = \sqrt{K-1}$.

   (b) Der Likelihood Ratio-Test für das einseitige Testproblem $H_0 = \{\sigma \leq \sigma_0\}$ gegen $H_1 = \{\sigma > \sigma_0\}$ lehnt ab, falls

$$\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{n\sigma_0^2} > c(\alpha).$$

   Der kritische Wert $c(\alpha)$ ist dabei die kleinere Lösung der Gleichung $Kx = \exp(x-1)$, wobei K wie unter Teil (a) definiert ist.

3. In einer genetischen Assoziationsstudie hat das Wellcome Trust Case Control Consortium (WTCCC, Nature, Vol. 447, 2007) unter anderem versucht, Positionen auf dem menschlichen Genom zu finden, die mit dem Krankheitsbild „Morbus Crohn“ assoziiert sind. Dazu wurde Blut von ca. 1800 Morbus Crohn-PatientInnen und von ca. 3000 gesunden Kontrollpersonen abgenommen und es wurde der Genotyp aller dieser Versuchspersonen an bestimmten Genorten (Loci) festgestellt. Der Genotyp an einem Locus setzt sich zusammnen aus zwei Allelen A_1, A_2 mit Werten in $\{A, C, G, T\}$, nämlich aus der maternalen und der paternalen DNA-Base an diesem Genort. Für Genort `rs5987140` zum Beispiel ergab sich das folgende Datenmaterial.

Genotyp	CC	CT	TT	$\sum$
erkrankt	101	234	1413	1748
gesund	196	322	2413	2931
$\sum$	297	556	3826	4679

(a) Berechnen Sie das minimale Signifikanzniveau $\alpha_{\text{min.}}$, für welches ein asymptotischer χ^2-Tests die Nullhypothese stochastischer Unabhängigkeit von Krankheitsstatus und Genotyp am Genort `rs5987140` anhand des obigen Datenmaterials ablehnt. Interpretieren Sie Ihr Ergebnis.

(b) Insgesamt sind in der erwähnten Studie $m = 455.086$ Genorte simultan daraufhin überprüft worden, ob Sie mit dem Erkrankungsrisiko für Morbus Crohn assoziiert sind. Wie könnte man die m „lokalen“ Signifikanzniveaus, die für die entsprechenden Assoziationstests an den jeweiligen Genorten verwendet werden, so „adjustieren“, dass gewährleistet ist, dass die Wahrscheinlichkeit für <u>irgendeinen</u> Typ I-Fehler unter den m Einzeltests durch eine fest vorgegebene obere Schranke $\alpha \in (0, 1)$ beschränkt ist? Interpretieren Sie Ihr Ergebnis unter (a) in diesem Lichte erneut.

4. Angenommen, man beobachtet Realisierungen von stochastisch unabhängigen und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$, wobei X_1 Gamma-verteilt ist mit unbekannten Parametern $\alpha > 0$ und $r > 0$. D. h., die Lebesguedichte von X_1 auf $\mathbb{R}_{\geq 0}$ unter (α, r) ist gegeben durch

$$f_{\alpha,r}(x) = \frac{\alpha^r}{\Gamma(r)} x^{r-1} \exp(-\alpha x), \; x \geq 0.$$

Betrachten Sie das Testproblem $H_0 = \{r = 1\}$ versus $H_1 = \Theta \setminus H_0$, wobei $\Theta = (0, \infty)^2$ den Parameterraum des Modells bezeichnet. Leiten Sie die Gestalt der Likelihood-Quotienten Statistik Λ_n für dieses Problem sowie die Grenzverteilung von $\Lambda_n(X)$ für $n \to \infty$ her, wobei $X = (X_1, \ldots, X_n)$ gilt.