Übung 4

1. Verbesserung einer Schätzung durch Suffizienz: Ein Beispiel aus der Qualitätskontrolle. Aus r Sendungen wird je eine Stichprobe des Umfangs n entnommen. In der i-ten Stichprobe befinden sich X_i defekte Stücke. Der erwartete Anteil defekter Stücke p $(0 \le p \le 1)$ sei in jeder Sendung gleich, aber unbekannt. Bei der Stichprobenerhebung wird ein ausgewähltes Stück jeweils zurückgelegt, bevor man das nächste herausgreift.

a) Finde eine suffiziente Statistik S für die Verteilung von $(X_1, \ldots, X_r)$. Bestimme die Verteilung von S.

b) Man möchte $P_p[X_1 \le 1]$ schätzen. Betrachte die Schätzung $T(X_1, \ldots, X_r) = I_{\{X_1 \le 1\}}$ und bestimme die verbesserte Schätzung $\tilde{T} = E[T|S]$.

c) Zeige, dass $\tilde{T}$ UMVU ist.

2. Wir betrachten das allgemeine lineare Modell

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \quad (i = 1, \ldots, n\,;\ p \le n)\,,$$

d. h. die n Beobachtungen y_i hängen – bis auf einen Fehler ε_i – linear von den p erklärenden Grössen $(x_{i1}, \ldots, x_{ip})$ ab. Die x_{ij} sind bekannt und fest, die y_i sind beobachtbar, und man möchte daraus $\beta = (\beta_1, \ldots, \beta_p)^\top$ schätzen. Wir nehmen an, dass die ε_i i.i.d. $\sim \mathcal{N}(0, \sigma^2)$ sind.

a) Seien $X = (x_{ij}) \in \mathbb{R}^{n \times p}$ und $Y = (y_i) \in \mathbb{R}^{n \times 1}$, und wir nehmen an, dass die Matrix X maximalen Rang p besitzt. Zeige, dass dann die Kleinste-Quadrate-Schätzer

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \Big(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2 = \arg\min_{\beta} (Y - X\beta)^\top (Y - X\beta)$$

geschrieben werden können als

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y\,.$$

b) Berechne die Fisher-Informationsmatrix für die Parameter $(\beta_1, \ldots, \beta_p)$.

c) Zeige, dass $\hat{\beta}$ erwartungstreu ist für β und die mehrdimensionale Cramér-Rao-Schranke erreicht, d. h. für beliebiges $(a_j) \in \mathbb{R}^p$ ist $\sum_{j=1}^{p} a_j \hat{\beta}_j$ UMVU für $\sum_{j=1}^{p} a_j \beta_j$.

3. Seien $X_1, \ldots, X_n$ i.i.d. $\sim p_\theta(x)dx$ mit $p_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$.

a) Zeige für $n = 1$, dass der Limes von $f_\Delta(x) = \frac{p_{\theta+\Delta}(x) - p_\theta(x)}{\Delta p_\theta(x)}$ für $\Delta \to 0$ existiert in $L_2(P_\theta)$, d.h. dass f existiert mit $\lim_{\Delta \to 0} \mathbf{E}[|f_\Delta - f|^2] = 0$.

b) Berechne die Cramér-Rao-Schranke für die Schätzgrösse $g(\theta) = \theta$ und berechne, wie weit $Var_\theta(\frac{1}{n} \sum X_i)$ von dieser Schranke entfernt ist.

c) Berechne für $n = 5$, wie weit $\mathrm{Var}_\theta(X_{(3)})$ von der Cramér-Rao-Schranke entfernt ist.
Hinweis: $\frac{5!}{2!2!} p_0(x) P_0(x)^2 (1 - P_0(x))^2 = \frac{15}{16} e^{-3|x|} (2 - e^{-|x|})^2$ liefert die Dichte von $X_{(3)} - \theta$, wobei $P_0(x)$ die Verteilungsfunktion sei.

Abgabe: Montag 7. Januar