Musterlösung zur Übung 4

1. a)

$$\begin{aligned}
P[X_i = x_i] &= \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \\
&= \binom{n}{x_i} e^{x_i \log(\frac{p}{1-p}) + n\log(1-p)} \qquad \text{(exponentielle Familie)} \\
P[X_1 = x_1, \ldots, X_r = x_r] &= \prod_{i=1}^{r} P[X_i = x_i] \\
&= e^{\sum_{i=1}^{r} x_i \log(\frac{p}{1-p}) + nr\log(1-p)} \cdot \prod_{i=1}^{r} \binom{n}{x_i}
\end{aligned}$$

Also ist $S = \sum X_i$ suffizient. Da S nichts anderes darstellt als die Anzahl defekter Stücke in einer Stichprobe des Umfangs $n \cdot r$, ist $S \sim \text{Bin}(nr, p)$.

b)

$$\begin{aligned}
\tilde{T} &= \mathbf{E}[T|S=s] = \mathbf{E}[I_{\{X_1 \leq 1\}}|S=s] = P[X_1 \leq 1|S=s] \\
&= \frac{P[X_1 \leq 1 \cap S = s]}{P[S=s]} \\
&= \frac{P[X_1 = 0] \cdot P[\sum_{i=2}^{r} X_i = s] + P[X_1 = 1] \cdot P[\sum_{i=2}^{r} X_i = s-1]}{P[S=s]} \\
&= \frac{(1-p)^n \cdot \binom{n(r-1)}{s} p^s (1-p)^{n(r-1)-s} + np(1-p)^{n-1} \cdot \binom{n(r-1)}{s-1} p^{s-1}(1-p)^{n(r-1)-s+1}}{\binom{nr}{s} p^s (1-p)^{nr-s}} \\
&= \frac{\binom{n(r-1)}{s} + n\binom{n(r-1)}{s-1}}{\binom{nr}{s}}
\end{aligned}$$

c) $T = I_{\{X_1 \leq 1\}}$ ist erwartungstreu für $P[X_1 \leq 1]$, da $\mathbf{E}[T] = \mathbf{E}[I_{\{X_1 \leq 1\}}] = P[X_1 \leq 1]$.
Gemäss Satz 3.2 ist S vollständig, da $c(p) = \log(\frac{p}{1-p})$ eine offene "Kugel" in $\mathbb{R}$ enthält.
Aus Korollar 3.1 (Lehmann-Scheffé) folgt, dass $\tilde{T} = \mathbf{E}[T|S]$ UMVU für $P[X_1 \leq 1]$ ist.

2. a) Durch Ableiten und Nullsetzen des Ausdrucks $(Y - X\beta)^\top(Y - X\beta)$ erhält man $(-2)X^\top(Y - X\beta) = 0$ beziehungsweise die p Normalgleichungen

$$X^\top X\beta = X^\top Y\,.$$

Die Behauptung folgt, da die $p \times p$-Matrix $X^\top X$ regulär ist.

b) Das Modell $y_i = \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i$, ε_i i.i.d. $\sim \mathcal{N}(0,\sigma^2)$ bedeutet nichts anderes als $Y \sim \mathcal{N}_n(X\beta, \Sigma)$ mit Kovarianzmatrix $\Sigma = \mathrm{diag}(\sigma^2, \dots, \sigma^2)$. Die Dichte des Beobachtungsvektors ist also (nach Definition A.1)

$$p_\beta(Y) = (2\pi)^{-\frac{1}{2}}(\det\Sigma)^{-\frac{1}{2}} \exp\big(-\frac{1}{2}(Y-X\beta)^\top\Sigma^{-1}(Y-X\beta)\big).$$

Mit der Bezeichnung $\frac{\partial}{\partial\beta} = (\frac{\partial}{\partial\beta_1}, \dots, \frac{\partial}{\partial\beta_p})^\top$ folgt

$$\frac{\partial}{\partial\beta}\log p_\beta(Y) = X^\top\Sigma^{-1}(Y-X\beta)$$

und

$$\begin{aligned} I(\beta) &= \mathbf{E}_\beta\big[\frac{\partial}{\partial\beta}\log p_\beta(Y)\,\big(\frac{\partial}{\partial\beta}\log p_\beta(Y)\big)^\top\big] \\ &= \mathbf{E}_\beta[X^\top\Sigma^{-1}(Y-X\beta)\,(X^\top\Sigma^{-1}(Y-X\beta))^\top] \\ &= \mathbf{E}_\beta[X^\top\Sigma^{-1}(Y-X\beta)\,(Y-X\beta)^\top\Sigma^{-1}X] \\ &= X^\top\Sigma^{-1}\underbrace{\mathbf{E}_\beta[(Y-X\beta)\,(Y-X\beta)^\top]}_{\Sigma}\Sigma^{-1}X \\ &= \Sigma^{-1}X^\top X\,. \end{aligned}$$

c) $\hat\beta$ ist erwartungstreu wegen

$$\mathbf{E}_\beta[\hat\beta] = \mathbf{E}_\beta[(X^\top X)^{-1}X^\top Y] = (X^\top X)^{-1}X^\top\,\mathbf{E}_\beta[Y] = (X^\top X)^{-1}X^\top X\beta = \beta\,.$$

Ausserdem erreicht $\hat\beta$ die mehrdimensionale Cramér-Rao-Schranke:

$$\begin{aligned} \mathrm{Var}_\beta(\hat\beta) &= \mathrm{Var}_\beta((X^\top X)^{-1}X^\top Y) = (X^\top X)^{-1}X^\top\,\mathrm{Var}_\beta(Y)((X^\top X)^{-1}X^\top)^\top \\ &= (X^\top X)^{-1}X^\top\Sigma((X^\top X)^{-1}X^\top)^\top = (X^\top X)^{-1}X^\top X(X^\top X)^{-1}\Sigma \\ &= (X^\top X)^{-1}\Sigma = I(\beta)^{-1} \end{aligned}$$

Wegen $\mathrm{Var}_\beta(a^\top\hat\beta) = a^\top\,\mathrm{Var}_\beta(\hat\beta)\,a = a^\top I(\beta)^{-1}a$ und der Schranke aus Korollar 3.2 folgt daraus sofort, dass für alle $a = (a_1, \dots, a_p)$ der Schätzer $a^\top\hat\beta = \sum_{j=1}^p a_j\hat\beta_j$ UMVU ist für $a^\top\beta = \sum_{j=1}^p a_j\beta_j$.

3. a) Wir zeigen zuerst f.s.-Konvergenz und Beschränktheit der betrachteten Folge, dann, dass daraus Konvergenz in $L_2(P_\theta)$ folgt.

i)

Sei ein P_θ *fix*. Für alle $x \neq \theta$ ist $p_\theta(x)$ nach θ differenzierbar, d. h. die Folge $\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta}$ konvergiert für $\Delta \to 0$ gegen $\frac{\partial}{\partial\theta}p_\theta(x)$. Folglich konvergiert die Folge $\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)} = \frac{1}{p_\theta(x)} \cdot \frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta}$ gegen

$$\frac{\frac{\partial}{\partial\theta}p_\theta(x)}{p_\theta(x)} = \frac{\partial}{\partial\theta}\log p_\theta(x) = \begin{cases} 1 & \text{falls } x > \theta \\ -1 & \text{falls } x < \theta \end{cases}.$$

Da $\{x|x=\theta\}$ bezüglich P_θ eine Nullmenge ist, konvergiert also $\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}$ P_θ-f.s. für $\Delta \to 0$.

ii)

Beh.: Für jedes Δ mit $|\Delta| \leq \Delta_0$ ist $|\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}| \leq \frac{e^{|\Delta|}-1}{|\Delta|} \leq \frac{e^{\Delta_0}-1}{\Delta_0} =: c$.

Bew.: Sei zuerst $\Delta > 0$. Wir unterscheiden drei Fälle:

- $\theta < \theta + \Delta < x$:

$$|\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}| = |\frac{e^{-(x-(\theta+\Delta))}-e^{-(x-\theta)}}{\Delta e^{-(x-\theta)}}| = |\frac{e^{\Delta}-1}{\Delta}| = \frac{|e^{|\Delta|}-1|}{|\Delta|} = \frac{e^{|\Delta|}-1}{|\Delta|}$$

- $\theta < x < \theta + \Delta$:

$$\begin{aligned} |\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}| &= |\frac{e^{-(\theta+\Delta-x)}-e^{-(x-\theta)}}{\Delta e^{-(x-\theta)}}| = |\frac{e^{2x-2\theta-\Delta}-1}{\Delta}| \\ &= \frac{|e^{\overbrace{2x-2\theta-\Delta}^{\in]-\Delta,\Delta[}}-1|}{|\Delta|} \end{aligned}$$

Für $y \in [0,\Delta[$ ist $|e^y-1| = e^y - 1 \leq e^\Delta - 1$, da e^y-1 monoton wächst. Für $y \in]-\Delta, 0[$ ist $|e^y-1|$ ebenfalls $\leq e^\Delta - 1$, denn e^y bewegt sich in positiver Richtung schneller von der 1 weg als in negativer. Also gilt oben weiter

$$|\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}| = \frac{|e^{2x-2\theta-\Delta}-1|}{|\Delta|} \leq \frac{e^{|\Delta|}-1}{|\Delta|}.$$

- $x < \theta < \theta + \Delta$:

$$|\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}| = |\frac{e^{-(\theta+\Delta-x)}-e^{-(\theta-x)}}{\Delta e^{-(\theta-x)}}| = |\frac{e^{-\Delta}-1}{\Delta}| = \frac{|e^{-\Delta}-1|}{|\Delta|} \leq \frac{e^{|\Delta|}-1}{|\Delta|}$$

Für $\Delta < 0$ erhält man die Aussage durch analoge Überlegungen in den Fällen $x < \theta+\Delta < \theta$, $\theta + \Delta < x < \theta$ und $\theta+\Delta < \theta < x$. □

iii)

Sei $f_\Delta(x) := \frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}$. Die $L_2(P_\theta)$-Konvergenz folgt dann aus dem folgenden Lemma.

Lemma: $f_\Delta \to f \ P_\theta - \text{f.s.}, \quad \sup_x |f_\Delta(x)| \leq c\, P_\theta - \text{f.s.}\, \forall\Delta$
$\Longrightarrow f_\Delta \to f$ in $L_p(P_\theta) (\forall p \geq 1)$.

Bew.: Sei $\delta > 0$.

$P_\theta[|f(x)| > c+\delta] \leq P_\theta[|f(x)| > |f_\Delta(x)| + \delta] \leq P_\theta[|f_\Delta(x) - f(x)| > \delta] \to 0,$ denn f.s.-Konvergenz impliziert Konvergenz in Wahrscheinlichkeit.

Es folgt $P_\theta[|f(x)| > c+\delta] = 0$ bzw. $|f(x)| \leq c + \delta\ \mathcal{P}_\theta -$ f.s.. Da dies für alle δ gilt, folgt $|f(x)| \leq c\ \mathcal{P}_\theta -$ f.s.. Daraus folgt $|f_\Delta - f| \leq |f_\Delta| + |f| \leq 2c\ \mathcal{P}_\theta -$ f.s. oder $g_\Delta := |f_\Delta - f|^p \leq (2c)^p\ \mathcal{P}_\theta -$ f.s.. Zusammen mit $g_\Delta := |f_\Delta - f|^p \to 0\ \mathcal{P}_\theta -$ f.s. sind somit die Voraussetzungen für den Satz von der majorisierten Konvergenz (H. Lebesgue) für die Folge g_Δ erfüllt. Also

$$\lim_{\Delta\to 0} \mathbf{E}[|f_\Delta - f|^p] = \mathbf{E}[\lim_{\Delta\to 0} |f_\Delta - f|^p] = 0\,.$$

□

Bemerkungen: Der Nachweis von $|f| \leq c$ f.s. im Beweis des Lemmas wäre hier nicht nötig gewesen, da in unserem Fall sowieso $|f| \leq 1$ f.s. gilt. Die Aussage des Lemmas gilt auch unter den schwächeren Voraussetzungen „Konvergenz in Wahrscheinlichkeit" und einer zufälligen L_p-integrierbaren Schranke. Für Beziehungen zwischen den verschiedenen Konvergenzbegriffen siehe z. B. Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

b) Aus $I(\theta) = \mathbf{E}_\theta[(\frac{\partial}{\partial\theta}\log p_\theta(x))^2] = \mathbf{E}_\theta[(\text{sign}(x-\theta))^2] = 1$ folgt die Cramér-Rao-Schranke

$$\frac{g'(\theta)^2}{I_n(\theta)} = \frac{g'(\theta)^2}{nI(\theta)} = \frac{1}{n}.$$

$$\begin{aligned}
\text{Var}_\theta\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) &= \frac{1}{n}\text{Var}_\theta(X_1) = \frac{1}{n}\text{Var}_0(X_1) = \frac{1}{n}\mathbf{E}_0[X_1^2] \\
&= \frac{1}{n}\int_{-\infty}^{\infty} x^2\frac{1}{2}e^{-|x|}\,dx = \frac{1}{n}\int_0^\infty x^2e^{-x}\,dx = \frac{2}{n}\int_0^\infty xe^{-x}\,dx \\
&= \frac{2}{n}\int_0^\infty e^{-x}\,dx = -\frac{2}{n}e^{-x}\Big|_0^\infty = \frac{2}{n}
\end{aligned}$$

$\text{Var}_\theta(\frac{1}{n}\sum_{i=1}^n X_i)$ liegt also um $\frac{1}{n}$ über der Cramér-Rao-Schranke.

c) Wir geben zuerst eine heuristische Herleitung, weshalb der Hinweis korrekt ist. (Eine saubere Herleitung wäre, zuerst die Verteilungsfunktion von $X_{(3)}$ herzuleiten und dann diese abzuleiten.) OEA sei $\theta = 0$. Das Ereignis $x \leq X_{(3)} \leq x + dx$ tritt genau dann auf, wenn 2 Beobachtungen kleiner als x sind, eine im Intervall $[x, x+dx]$ liegt und 2 Beobachtungen grösser als x sind. Ein Anordnung dieser Art hat Wahrscheinlichkeit $p_0(x)P_0(x)^2(1-P_0(x))^2$ und es gibt $\frac{5!}{2!2!}$ solcher Anordnungen, womit der Hinweis bewiesen ist.

Nun zur Berechnung von $\text{Var}_\theta(X_{(3)})$:

$$\begin{aligned}
\text{Var}_\theta(X_{(3)}) &= \mathbf{E}[(X_{(3)}-\theta)^2] = \frac{15}{16}\int_{-\infty}^{\infty} x^2e^{-3|x|}(2-e^{-|x|})^2dx \\
&= 2\cdot\frac{15}{16}\int_0^\infty x^2e^{-3x}(4-4e^{-x}+e^{-2x})dx
\end{aligned}$$

Unter Verwendung von $\int_0^\infty x^2e^{-kx}dx = \frac{2}{k^3}$ für $k \in \mathbb{N}$ (via partielle Integration) ergibt sich

$$\text{Var}_\theta(X_{(3)}) = \frac{30}{16}\Big(\frac{8}{3^3} - \frac{8}{4^3} + \frac{2}{5^3}\Big) \approx 0.35.$$

Für $n = 5$ ist die Varianz des Maximum Likelihood Schätzer $X_{(3)}$ also ungefähr 0.15 von der Cramér-Rao-Schranke entfernt. Dies ist besser als das arithmetische Mittel, dessen Varianz 0.2 von der Cramér-Rao-Schranke entfernt ist.

<u>Ausblick:</u> In Kapitel 6 wird gezeigt, dass asymptotisch (d.h. für $n \to \infty$) gilt $\text{med}(X_i) \sim \mathcal{N}(\theta, \frac{1}{nI(\theta)}) = \mathcal{N}(\theta, \frac{1}{n})$ und $\overline{X} \sim \mathcal{N}(\theta, \frac{2}{n})$. Asymptotisch erreicht also der Median die Cramér-Rao-Schranke. Dies gilt allgemein für den Maximum Likelihood Schätzer (siehe Abschnitt 6.3.1), aber (wie dieses Beispiel zeigt) nicht für Momentenschätzer.

Wir begnügen uns hier mit einer Grafik, die die Varianzen von $\text{med}(X_i)$ und $\overline{X}$, sowie die Cramér-Rao-Schranke für n zwischen 10 und 100 zeigt. Weil sich die Varianzen von $\text{med}(X_i)$ nicht ohne weiteres exakt berechnen lassen, wurden sie durch Simulation bestimmt.