Musterlösung zur Übung 5

1. **a)** Äquivarianz soll hier bedeuten, dass

$$T(U_1+c_1,\dots,U_n+c_1,V_1+c_2,\dots,V_m+c_2)=T(U_1,\dots,U_n,V_1,\dots,V_m)+(c_1-c_2)\,.$$

b) Analog zur Argumentation im Lokationsmodell gilt

$$\begin{aligned}
R(\theta,T) &= \mathbf{E}_\theta[L(\theta,T(U_1,\dots,U_n,V_1,\dots,V_m))]\\
&= \mathbf{E}_\theta[L(\theta,T(\mu+\varepsilon_1,\dots,\mu+\varepsilon_n,\nu+\eta_1,\dots,\nu+\eta_m))]\\
&= \mathbf{E}_\theta[L(\theta,T(\varepsilon_1,\dots,\varepsilon_n,\eta_1,\dots,\eta_m)+(\mu-\nu))]\\
&= \mathbf{E}_\theta[(T(\varepsilon_1,\dots,\varepsilon_n,\eta_1,\dots,\eta_m)+(\mu-\nu)-g(\theta))^2]\\
&= \mathbf{E}[(T(\varepsilon_1,\dots,\varepsilon_n,\eta_1,\dots,\eta_m))^2]\\
&= R((0,0),T)\text{ unabhängig von }\theta.
\end{aligned}$$

c) Beh.: T ist äquivariant genau dann, wenn $\exists\, v:\mathbb{R}^{n+m-2}\to\mathbb{R}$ mit $T=(u_n-v_m)+v(x_1,\dots,x_{n-1},y_1,\dots,y_{m-1})$, wobei $x_i=u_i-u_n$ und $y_j=v_j-v_m$.
Bew.: „$\Leftarrow$“ ist klar. Für „$\Rightarrow$“ benutzt man, dass

$$\begin{aligned}
&T(u_1,\dots,u_n,v_1,\dots,v_m)\\
&\quad=T(u_1-u_n,\dots,u_{n-1}-u_n,0,v_1-v_m,\dots,v_{m-1}-v_m,0)+(u_n-v_m)\,.
\end{aligned}$$

Also setzt man einfach

$$v(x_1,\dots,x_{n-1},y_1,\dots,y_{m-1})=T(x_1,\dots,x_{n-1},0,y_1,\dots,y_{m-1},0)\,.$$

$\square$

d) Der UMRE-Schätzer $\hat\mu$ für μ hat die Form

$$\hat\mu=U_n+v_1^*(X)\,,$$

mit $X=(U_1-U_n,\dots,U_{n-1}-U_n)$ und

$$v_1^*(X)=\arg\min_v \mathbf{E}_0[L(U_n+v)|X=(x_1,\dots,x_n)]=-\,\mathbf{E}_0[U_n|X=(x_1,\dots,x_{n-1})]\,.$$

Für den UMRE-Schätzer $\hat\nu$ von ν gilt analog

$$\hat\nu=V_m+v_2^*(Y)=V_m-\mathbf{E}_0[V_m|Y=(y_1,\dots,y_{m-1})]$$

mit $Y=(V_1-V_m,\dots,V_{m-1}-V_m)$.

Daraus folgt unter Verwendung der Unabhängigkeit der beiden Stichproben

$$\begin{aligned}
T &= \hat{\mu} - \hat{\nu} \\
&= U_n - \mathbf{E}_0[U_n | X = (x_1, \ldots, x_{n-1})] - (V_m - \mathbf{E}_0[V_m | Y = (y_1, \ldots, y_{m-1})]) \\
&= (U_n - V_m) - (\mathbf{E}_0[U_n | X = (x_1, \ldots, x_{n-1})] - \mathbf{E}_0[V_m | Y = (y_1, \ldots, y_{m-1})]) \\
&= (U_n - V_m) - (\mathbf{E}_0[U_n | X = (x_1, \ldots, x_{n-1}), Y = (y_1, \ldots, y_{m-1})] \\
&\qquad - \mathbf{E}_0[V_m | X = (x_1, \ldots, x_{n-1}), Y = (y_1, \ldots, y_{m-1})]) \\
&= (U_n - V_m) - \mathbf{E}_0[U_n - V_m | (X, Y) = (x_1, \ldots, x_{n-1}, y_1, \ldots, y_{m-1})] \\
&= (U_n - V_m) \\
&\qquad + \arg\min_v \mathbf{E}_{(0,0)}[L(U_n - V_m + v) | (X, Y) = (x_1, \ldots, x_{n-1}, y_1, \ldots, y_{m-1})] \,.
\end{aligned}$$

Derartiges punktweises Minimieren ergibt aber gerade den UMRE-Schätzer!

2. a) $R(\theta, T) = \mathbf{E}_\theta[(X - \theta)^2] = \mathbf{E}_\theta[(X - \mathbf{E}_\theta[\theta])^2] \equiv 1$. Nun nehmen wir an, es existiere ein Schätzer mit kleinerem maximalem Risiko als T, also T' mit $\sup_{\theta \geq 0} R(\theta, T') = 1 - \varepsilon$ für ein $\varepsilon > 0$. Einerseits

$$\begin{aligned}
R(\theta, T') &= \mathbf{E}_\theta[(T' - \theta)^2] = \mathbf{E}_\theta[(T' - \mathbf{E}_\theta[T'] + \mathbf{E}_\theta[T'] - \theta)^2] \\
&= \mathrm{Var}_\theta(T') + 2 \cdot 0 + (\mathbf{E}_\theta[T'] - \theta)^2 \\
&= \mathrm{Var}_\theta(T') + b(\theta)^2 \,,
\end{aligned}$$

andererseits

$$\mathrm{Var}_\theta(T') \geq \frac{(\frac{d}{d\theta}(b(\theta) + \theta))^2}{I(\theta)} = \frac{1}{I(\theta)}(b'(\theta) + 1)^2 \,,$$

also

$$R(\theta, T') - b(\theta)^2 \geq \frac{1}{I(\theta)}(b'(\theta) + 1)^2$$

beziehungsweise

$$1 - \varepsilon - b(\theta)^2 \geq \frac{1}{I(\theta)}(b'(\theta) + 1)^2$$

für alle $\theta \geq 0$. Die Fisherinformation ist

$$-\mathbf{E}_\theta \Big[\frac{d^2}{d\theta^2} \log \Big(\frac{1}{\sqrt{2\pi}} e^{-\frac{(X-\theta)^2}{2}}\Big)\Big] = -\mathbf{E}_\theta \Big[\frac{d}{d\theta}(X - \theta)\Big] = 1 \,.$$

Aus

$$1 - \varepsilon - b(\theta)^2 \geq (b'(\theta) + 1)^2 \tag{1}$$

erhält man

$$b(\theta)^2 + 2b'(\theta) \leq -\varepsilon - b'(\theta)^2 < 0$$

und

$$-\frac{b'(\theta)}{b(\theta)^2} > \frac{1}{2}$$

für $\theta \geq 0$. Für ein festes $\theta_0 \geq 0$ gilt

$$\begin{aligned} \frac{1}{b(\theta)} &= \frac{1}{b(\theta_0)} + \int_{\theta_0}^{\theta} \Big(\frac{1}{b(\tilde\theta)}\Big)' \, d\tilde\theta = \frac{1}{b(\theta_0)} + \int_{\theta_0}^{\theta} \underbrace{\Big(-\frac{b'(\tilde\theta)}{b(\tilde\theta)^2}\Big)}_{>\frac{1}{2}\,\forall \tilde\theta \geq \theta_0} \, d\tilde\theta \\ &> \frac{1}{b(\theta_0)} + \frac{1}{2}(\theta - \theta_0) \to \infty \ (\theta \to \infty)\,. \end{aligned}$$

Es folgt $b(\theta) \to 0$ $(\theta \to \infty)$. Da $b(\theta)$ streng monoton fallend ($b'(\theta) < -\frac{1}{2}b(\theta)^2 \leq 0$) gegen 0 konvergiert, muss auch $b'(\theta) \to 0$ $(\theta \to \infty)$. Der Grenzübergang $\theta \to \infty$ in Gleichung (1) führt zum Widerspruch

$$1 - \varepsilon \geq 1\,.$$

Also kann kein Schätzer mit kleinerem maximalem Risiko als $T = X$ existieren.

b) Sei $T' = 1_{[X \geq 0]} X$ der Kandidat für einen besseren Schätzer als $T = X$. Es gilt

$$\begin{aligned} R(\theta, T') &= \mathbf{E}_\theta[(\theta - 1_{[X\geq 0]}X)^2] = \int_{-\infty}^{\infty} (\theta - 1_{[x \geq 0]}x)^2 \phi(x-\theta)\,dx \\ &= \int_{-\infty}^{0} \underbrace{\theta^2}_{<(\theta+|x|)^2 = (\theta - x)^2} \phi(x-\theta)\,dx + \int_0^\infty (\theta - x)^2 \phi(x-\theta)\,dx \\ &< \int_{-\infty}^{0} (\theta - x)^2 \phi(x-\theta)\,dx + \int_0^\infty (\theta - x)^2 \phi(x-\theta)\,dx \\ &= \int_{-\infty}^{\infty} (\theta - x)^2 \phi(x-\theta)\,dx = R(\theta, T) \end{aligned}$$

für alle $\theta \geq 0$. Also ist T unzulässig.

3. a) Die folgenden Ungleichungen sind selbsterklärend:

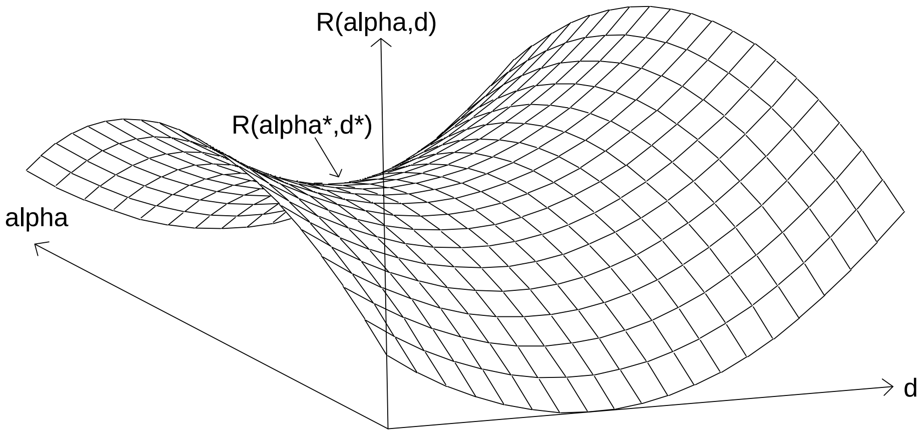
$$\begin{aligned} & \sup_\alpha R(\alpha, d) \geq R(\beta, d) \ \forall\, \beta, d \\ \Rightarrow\ & \inf_d \sup_\alpha R(\alpha, d) \geq \inf_d R(\beta, d) \ \forall\, \beta \\ \Rightarrow\ & \inf_d \sup_\alpha R(\alpha, d) \geq \sup_\beta \inf_d R(\beta, d) = \sup_\alpha \inf_d R(\alpha, d) \end{aligned}$$

b) Die Behauptung ergibt sich sofort aus a) und der folgenden Kette von Ungleichungen.

$$\sup_\alpha \inf_d R(\alpha, d) \geq \inf_d R(\alpha^*, d) \geq R(\alpha^*, d^*) \geq \sup_\alpha R(\alpha, d^*) \geq \inf_d \sup_\alpha R(\alpha, d)$$

Bemerkungen: Sattelpunkte sind eine Art Pattsituationen. In der Spieltheorie heissen sie auch „Nash-Gleichgewichte“. Wenn der Statistiker auf seiner Wahl d^* beharrt, dann maximiert die Natur wegen $R(\alpha, d^*) \leq R(\alpha^*, d^*)$ mit der Strategie α^* ihren Gewinn. (Wenn der Statistiker aber vom gewählten d^* abweicht, kann die Natur ihren Gewinn wegen $R(\alpha^*, d^*) \leq R(\alpha^*, d)$ unter Umständen noch vergrössern.) Genauso verhält es sich für den Statistiker. Solange die Natur an ihrer Wahl α^* festhält, lohnt es sich nicht, die Strategie zu wechseln.

Man kann sich die Situation anhand einer Abbildung veranschaulichen:



c) Wegen $R(\theta, T^*)$ konstant gilt $R(\alpha, T^*) = R(\alpha^*, T^*)$ und wegen T^* Bayes zur a priori Verteilung α^* gilt $R(\alpha^*, T^*) \leq R(\alpha^*, T)$, also sind die Voraussetzungen von b) erfüllt und (T^*, α^*) ist ein Sattelpunkt.

α^* heisst ungünstigste Verteilung, weil der erwartete Verlust des (vorsichtigen, sich an die Minimax-Strategie haltenden) Statistikers bei dieser Verteilung maximal ist. Bei der Bezeichnung wird davcon ausgegangen, dass $R(\theta, T_\alpha)$ für die Bayes-Schätzer T_α anderer a priori-Verteilungen α nicht konstant ist.

d) Die untenstehende Auszahlungsmatrix zeigt den Verlust von Sara:

		N		
		1	2	max
S	1	-2	3	3
	2	3	-4	3
min		-2	-4	

Durch Zeigen von $d_N = 1$ Fingern erreicht Nina also ihren Minimax-"Gewinn" von -2. Sara's maximaler Verlust ist 3 unabhängig von d_S, der Anzahl Finger, die sie zeigt. Also gilt:

$$\inf_{d_S} \sup_{d_N} R(d_S, d_N) = 3 > -2 = \sup_{d_N} \inf_{d_S} R(d_S, d_N)$$

Es gibt also keine Pattsituation.

e) Sara's erwarteter Verlust lautet:

$$\begin{aligned} R &= -2p_N p_S - 4(1-p_N)(1-p_S) + 3(p_N(1-p_S) + (1-p_N)p_S) \\ &= -4 + 7p_N + 7p_S - 12p_N p_S \\ &= -4 + 7p_N + p_S \cdot (7 - 12p_N) = -4 + 7p_S + p_N \cdot (7 - 12p_S) \end{aligned}$$

Aus obiger Gleichung ersehen wir, dass für $p_N = \frac{7}{12}$ resp. $p_S = \frac{7}{12}$ der erwartete Verlust unabhängig ist von der Strategie der Gegnerin, das Risiko also konstant ist.
Die Voraussetzungen von b) sind erfüllt, da $R(p_N, \frac{7}{12}) = R(\frac{7}{12}, \frac{7}{12}) = R(\frac{7}{12}, p_S)$. Also handelt es sich beim Strategiepaar $(p_N = \frac{7}{12}, p_S = \frac{7}{12})$ um einen Sattelpunkt.
Nebenbemerkung: Aus der Warte der ersten Teilaufgaben betrachtet entspricht dieses Beispiel der Statistikerin Sara, die den Bayes-Schätzer

$$d = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p_S = \frac{7}{12} \\ 2 & \text{mit Wahrscheinlichkeit } p_S = \frac{5}{12} \end{cases}$$

wählt zur ungünstigsten a priori Verteilung der Natur "Nina":

$$\alpha = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p_N = \frac{7}{12} \\ 2 & \text{mit Wahrscheinlichkeit } p_N = \frac{5}{12} \end{cases}$$

f) Eingesetzt ergibt der Sattelpunkt ein Risiko $R = \frac{1}{12}$. Es ist also sinnvoll, an Stelle von Nina zu spielen. Das ist auf den ersten Blick ein erstaunliches Resultat!