Musterlösung zur Übung 3

1. a) Die erste Tabelle enthält die neun möglichen Verfahren.

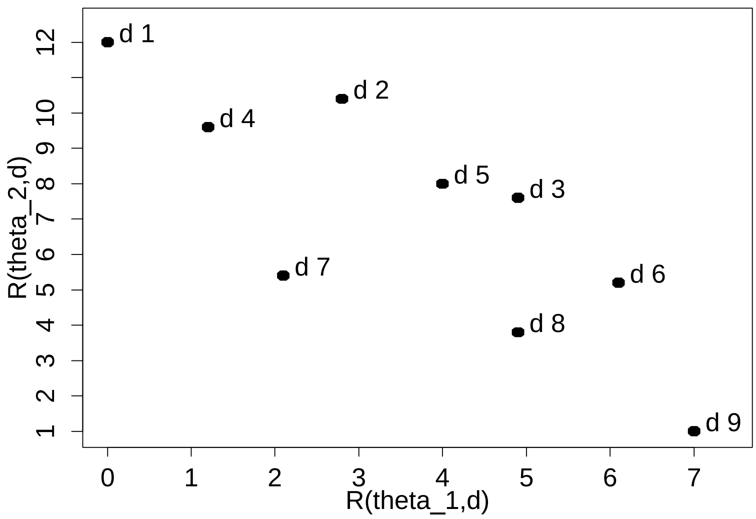
Verfahren	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
Entscheid bei $X = 0$	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3
Entscheid bei $X = 1$	a_1	a_2	a_3	a_1	a_2	a_3	a_1	a_2	a_3

Die zweite Tabelle enthält die Risiken aller neun Verfahren in Abhängigkeit des Parameters gemäss

$$R(\theta_i, d_j) = \int_{\{0,1\}} L(\theta_i, d_j(x))\, P_{\theta_i}(dx) = L(\theta_i, d_j(0))P_{\theta_i}(0) + L(\theta_i, d_j(1))P_{\theta_i}(1)\,.$$

Verfahren	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
$R(\theta_1, d_j)$	0	2.8	4.9	1.2	4	6.1	2.1	4.9	7
$R(\theta_2, d_j)$	12	10.4	7.6	9.6	8	5.2	5.4	3.8	1

Aus der zweiten Tabelle oder der folgenden Figur kann man die Antworten zu den Aufgaben b) und c) ohne weiteres ablesen.



b) Entscheidungsregel d_8 ist minimax.

c) Zulässig sind die Verfahren d_1, d_4, d_7, d_8 und d_9.

d) Nach dem Satz von Bayes gilt

$$P[\theta_i|X] = \frac{P_{\theta_i}(X)\,\alpha(\theta_i)}{P_{\theta_1}(X)\,\alpha(\theta_1) + P_{\theta_2}(X)\,\alpha(\theta_2)} \quad (i = 1, 2)\,,$$

also $P[\theta_1|X = 0] = \frac{0.3\cdot 0.6}{0.3\cdot 0.6+0.6\cdot 0.4} \approx 0.43$, $P[\theta_2|X = 0] = 1 - P[\theta_1|X = 0] \approx 0.57$, $P[\theta_1|X = 1] = \frac{0.7\cdot 0.6}{0.7\cdot 0.6+0.4\cdot 0.4} \approx 0.72$ und $P[\theta_2|X = 1] = 1 - P[\theta_1|X = 1] \approx 0.28$.

Nach Satz 2.1 findet man das Bayesverfahren durch punktweises Minimieren des erwarteten Verlusts unter der Posterioriverteilung, d. h.

$$d(x) = \arg\min_a \mathbf{E}[L(\theta, a)|X = x]\,.$$

Diese bedingten Erwartungswerte kann man für jedes a (und für beide X) wie folgt berechnen:

$$\begin{aligned}
\mathbf{E}[L(\theta, a_1)|X = 0] &= L(\theta_1, a_1)\,P[\theta_1|X = 0] + L(\theta_2, a_1)\,P[\theta_2|X = 0]\\
&\approx 0\cdot 0.43 + 12\cdot 0.57 = 6.84\\
\mathbf{E}[L(\theta, a_2)|X = 0] &\approx 4\cdot 0.43 + 8\cdot 0.57 = 6.28\\
\mathbf{E}[L(\theta, a_3)|X = 0] &\approx 7\cdot 0.43 + 1\cdot 0.57 = 3.58\\
\mathbf{E}[L(\theta, a_1)|X = 1] &= L(\theta_1, a_1)\,P[\theta_1|X = 1] + L(\theta_2, a_1)\,P[\theta_2|X = 1]\\
&\approx 0\cdot 0.72 + 12\cdot 0.28 = 3.36\\
\mathbf{E}[L(\theta, a_2)|X = 1] &\approx 4\cdot 0.72 + 8\cdot 0.28 = 5.12\\
\mathbf{E}[L(\theta, a_3)|X = 1] &\approx 7\cdot 0.72 + 1\cdot 0.28 = 5.32
\end{aligned}$$

Das Verfahren d_7 mit $d_7(0) = a_3$ und $d_7(1) = a_1$ ist Bayes bezüglich α.

2. a) Es seien $\alpha(\theta)$ die Dichte der Prioriverteilung, $p_\theta(\underline{X}) = p_\theta(X_1, \ldots, X_n)$ die Dichte der Likelihood und $p(\theta|\underline{X}) = p(\theta|X_1, \ldots, X_n)$ die Dichte der Posterioriverteilung. Der Bezeichner "$\propto$" bedeutet "ist proportional zu", d.h. die Terme unterscheiden sich nur um den Regularitätsfaktor, um eine Wahrscheinlichkeits-Dichte zu erzeugen. Mit diesen Bezeichnungen gilt

$$\begin{aligned}
p(\theta|\underline{X}) &= \frac{p_\theta(\underline{X})\alpha(\theta)}{\int p_\theta(\underline{X})\alpha(\theta)\,d\theta} \propto p_\theta(\underline{X})\alpha(\theta)\\
&= (2\pi)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2\Big)\,(2\pi)^{-\frac{1}{2}}\tau^{-1}\exp\Big(-\frac{1}{2}\Big(\frac{\theta-\xi}{\tau}\Big)^2\Big)\\
&\propto \exp\Big(-\frac{1}{2}\Big[\theta^2\underbrace{\Big(n + \frac{1}{\tau^2}\Big)}_{=:a^2} - 2\theta\underbrace{\Big(\sum_{i=1}^{n} X_i + \frac{\xi}{\tau^2}\Big)}_{=:ab} + \sum_{i=1}^{n} X_i^2 + \frac{\xi^2}{\tau^2}\Big]\Big)\\
&\propto \exp\Big(-\frac{1}{2}(a\theta - b)^2\Big) = \exp\Bigg(-\frac{1}{2}\bigg(\frac{\theta - \frac{b}{a}}{\frac{1}{a}}\bigg)^2\Bigg)\,.
\end{aligned}$$

Dieser Ausdruck ist proportional zur Dichte der Normalverteilung $\mathcal{N}(b/a, 1/a^2)$, also

$$\theta|X_1,\ldots,X_n \sim \mathcal{N}\Big(\frac{\sum X_i + \frac{\xi}{\tau^2}}{n+\frac{1}{\tau^2}}, \frac{1}{n+\frac{1}{\tau^2}}\Big) = \mathcal{N}\Big(\frac{n\tau^2}{1+n\tau^2}\,\overline{X} + \frac{1}{1+n\tau^2}\,\xi, \frac{\tau^2}{1+n\tau^2}\Big)\,.$$

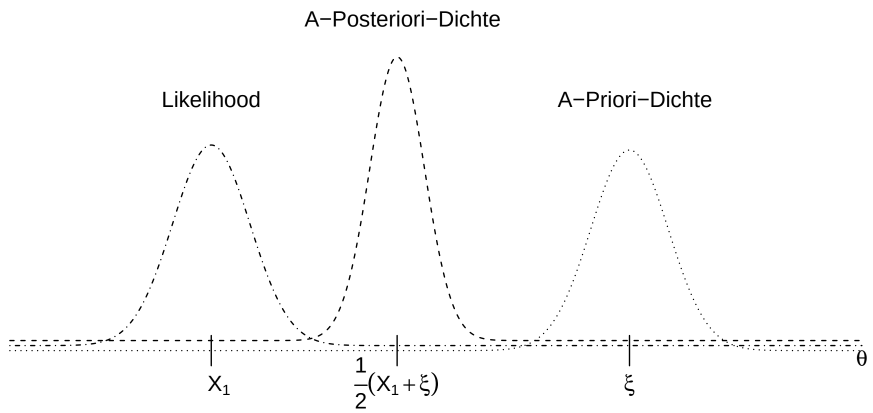
b)

$$\begin{aligned}\hat{\theta}_{Bayes} &= \arg\min_a \mathbf{E}[L(\theta,a)|\underline{X}] = \arg\min_a \mathbf{E}[(\theta-a)^2|\underline{X}] = \mathbf{E}[\theta|\underline{X}] \\ &= \frac{n\tau^2}{1+n\tau^2}\,\overline{X} + \frac{1}{1+n\tau^2}\,\xi\end{aligned}$$

Der Bayesschätzer ist eine konvexe Kombination aus dem Erwartungswert der Prioriverteilung und dem Stichprobenmittelwert. Je grösser der Stichprobenumfang, desto mehr Gewicht erhalten die Beobachtungen.

c) $|X_1 - \xi| >> 1$ bedeutet, dass die Beobachtung vom a-priori angenommenen Erwartungswert stark abweicht. Die Prioriverteilung ist also a-posteriori (d. h. nachdem man $\underline{X}$ beobachtet hat) nicht mehr plausibel. Intuitiv kann ein Kompromiss zwischen Prioriverteilung und Beobachtung, also eine zwischen ξ und der Beobachtung zentrierte Posterioriverteilung erwartet werden.

Das ist auch der Fall. Der Erwartungswert der Posterioriverteilung liegt bei $\frac{1}{2}X_1 + \frac{1}{2}\xi$. Überraschen kann jedoch, dass ihre Standardabweichung $\frac{1}{\sqrt{2}}$ ist, also kleiner als diejenige der Prioriverteilung. Die Posterioriverteilung hat also den Großteil ihrer Masse dort, wo weder Likelihood-, noch Prioriverteilung nennenswert Masse haben. Betrachte folgende Skizze:



3. a) Zuerst zum Hinweis: Bei gegebenen θ_j sind die X_j unabhängig $\mathcal{N}(\theta_j, 1)$-verteilt, und lineare Transformationen $\varepsilon_j := X_j - \theta_j$ der X_j sind dann unabhängig $\mathcal{N}(0,1)$-verteilt. Insbesondere sind die ε_j gar nicht mehr von den θ_j abhängig. Man hat also $X_j = \theta_j + \varepsilon_j$ wie im Hinweis angegeben.

Nun sind Summen von unabhängigen Normalverteilungen wiederum Normalverteilungen (siehe A.4). Wegen $\mathbf{E}[\theta_j + \varepsilon_j] = \mathbf{E}[\theta_j] + \mathbf{E}[\varepsilon_j] = \xi + 0 = \xi$ und $\mathrm{Var}(\theta_j + \varepsilon_j) = \mathrm{Var}(\theta_j) + \mathrm{Var}(\varepsilon_j) = \tau^2 + 1$, und weil $(\theta_1, \varepsilon_1), \ldots, (\theta_k, \varepsilon_k)$ unabhängig sind, hat man also

$$X_j \text{ i.i.d. } \sim \mathcal{N}(\xi, 1+\tau^2) \quad (j = 1,\ldots,k)\,.$$

b) Die ML-Schätzungen für Erwartungswert und Varianz der normalverteilten Zufallsvariablen X_j sind

$$\hat{\xi}_{MLE} = \overline{X} \quad \text{und} \quad 1 + \hat{\tau}^2_{MLE} = \frac{1}{k}\sum_{j=1}^{k}(X_j - \overline{X})^2 =: s_X^2 \, .$$

Diese Schätzungen braucht man nur noch in den Bayesschätzer $\hat{\theta}_{j\,Bayes} = \frac{\tau^2}{1+\tau^2}\,X_j + \frac{1}{1+\tau^2}\,\xi$ (s. 1.b) mit n=1, da pro θ_j ein X_j beobachtet wurde) einzusetzen. Man erhält

$$\hat{\theta}_{j\,emp.\ Bayes} = \frac{s_X^2 - 1}{s_X^2}\,X_j + \frac{1}{s_X^2}\,\overline{X} \, .$$