Sheet 5

Problem 1

We begin by computing the posterior distribution of θ by use of Bayes-rule:

$$\begin{aligned}
f(\theta|X) &\propto f(X|\theta)\pi(\theta) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\theta)^2\right) \frac{1}{\sqrt{2\pi} m} \exp\left(-\frac{1}{2m}\theta^2\right) \\
&\propto \exp\left(-\frac{1}{2}\left((x-\theta)^2 + \frac{1}{m}\theta^2\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left((x-\theta)^2 + \frac{1}{m}\theta^2 - \frac{1}{1+m}x^2\right)\right)
\end{aligned}$$

In the last step, is a proportional transformation, because $\exp(-0.5(1/(1+m))x^2)$ is a constant, because the function is dependent on θ and not x.

only considering the term inside of the brackets of the exponent, we want to get to a form, where we can read the variance and expected value of this distribution from the term:

$$\begin{aligned}
(x-\theta)^2 + \frac{1}{m}\theta^2 - \frac{1}{1+m}x^2 &= x^2 - 2\theta x + \theta^2 + \frac{1}{m}\theta^2 - \frac{1}{1+m}x^2 \\
&= \frac{m}{m+1}x^2 - 2x\theta + \frac{1+m}{m}\theta^2 \\
&= \left(\frac{1+m}{m}\right)^2 \left(\frac{m}{1+m}\right)^2 \left(\frac{m}{m+1}x^2 - 2x\theta + \frac{1+m}{m}\theta^2\right) \\
&= \left(\frac{1+m}{m}\right) \left(\left(\frac{m}{m+1}\right)^2 x^2 - 2x\left(\frac{m}{1+m}\right)\theta + \theta^2\right) \\
&= \left(\frac{1+m}{m}\right) \left(\frac{m}{m+1}x - \theta\right)^2 = \frac{(\theta-\mu)^2}{\sigma^2}
\end{aligned}$$

From this we know that:

$$\mu = \frac{m}{m+1}x \tag{1}$$
$$\sigma^2 = \frac{m}{m+1} \tag{2}$$

This leads to:

$$\theta|X \sim \mathcal{N}\left(\frac{m}{m+1}x, \frac{m}{m+1}\right)$$

Now we can calculate the Baysian-Risk for $\delta_m(x)$:

$$
\begin{aligned}
B(\pi_m, \delta_m) &= \mathbb{E}((\theta - \delta_m(x))^2) \\
&= Var(\theta) + Bias(\theta)^2 \\
&\underset{(2)}{=} \frac{m}{m+1} + \left(\mu - \frac{m}{m+1}x\right)^2 \\
&\underset{(1)}{=} \frac{m}{m+1} + \left(\frac{m}{m+1}x - \frac{m}{m+1}x\right)^2 \\
&= \frac{m}{m+1}
\end{aligned}
$$

Problem 2

In order to get the ML estimator we need to maximize the log-likelihood function. The likelihood function is given by:

$$
L(X, \theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left[-\frac{1}{\theta}x\right]
$$

The log-likelihood function $\mathcal{L}$ is given by $\log(L(X, \theta))$:

$$
\mathcal{L}(X, \theta) = \sum_{i=1}^{n} \log\left(\frac{1}{\theta}\right) - \frac{1}{\theta}x_i = n \log\left(\frac{1}{\theta}\right) - \frac{1}{\theta}\sum_{i=1}^{n} x_i
$$

Deriving by θ then yields (by use of chain rule):

$$
\frac{\partial \mathcal{L}}{\partial \theta} = n\frac{1}{\frac{1}{\theta}}\left(-\frac{1}{\theta^2}\right) + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i - n\frac{1}{\theta}
$$

FOC condition yields:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \theta} &\overset{!}{=} 0 \\
\Leftrightarrow \frac{1}{\theta^2}\sum_{i=1}^{n} x_i - n\frac{1}{\theta} &= 0 \\
\Leftrightarrow \frac{1}{\theta}\sum_{i=1}^{n} x_i - n &= 0 \\
\theta &= \frac{1}{n}\sum_{i=1}^{n} x_i := \bar{X}_n
\end{aligned}
$$

We can easily show, that δ_{ML} is unbiased:

$$
\mathbb{E}(\delta_{ML}) = \mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(x_i) = \frac{n}{n}\theta = \theta
$$

Also the Variance of δ_{ML} is:

$$
\begin{aligned}
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\
&\underset{iid}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\
&\underset{iid}{=} \frac{1}{n^2}\sum_{i=1}^{n} \theta^2 \\
&= \frac{1}{n^2} n\theta^2 = \frac{1}{n}\theta^2 (*)
\end{aligned}
$$

We are going to show that $R(\theta, \delta_{ML}) = 1$ holds for all $n \in \mathbb{N}_0$:

$$
\begin{aligned}
R(\theta, \delta_{ML}) &= \mathbb{E}(L(\theta, \delta_{ML})) \\
&= \mathbb{E}\left(\frac{n(\theta - \delta_{ML})^2}{\theta^2}\right) \\
&= \frac{n}{\theta^2}\mathbb{E}\left(\left(\mathbb{E}(\bar{X}_n) - \bar{X}_n\right)^2\right) \\
&= \frac{n}{\theta^2} MSE(\bar{X}_n) \\
&= \frac{n}{\theta^2}(Var(\bar{X}_n) + Bias(\bar{X}_n)^2) \\
&\underset{\text{unbiased}}{=} \frac{n}{\theta^2} Var(\bar{X}_n) \\
&\underset{(*)}{=} \frac{n}{\theta^2}\frac{\theta^2}{n} = 1
\end{aligned}
$$

Because the risk is constant regardless of the θ, and the risk is the lowest possible, we can say that δ_{ML} is an **admissible equalizer rule** and therefore a mini-max rule.

Problem 3

There seems to be no random element in the risk. For example $R(\theta_1, d_1) = 7 = L(\theta_1, d_1)$. Bayes-Risk is given by:

$$
B(\delta, \pi) = \sum_{i=1}^{3} \pi(\theta_i) R(\theta_i, \delta) = \sum_{i=1}^{3} \pi(\theta_i) L(\theta_i, \delta)
$$

The prior π is given by π_*. The Bayes-Risk is therefore:

$$
B(\delta, \pi_*) = \frac{7}{49} L(\theta_1, \delta) + \frac{10}{49} L(\theta_2, \delta) + \frac{32}{49} L(\theta_3, \delta)
$$

It turns out, that the Bayes-Risk under this prior is the same for every non-random rule:

$$
\begin{aligned}
B(d_1, \pi_*) &= 1.653061 \\
B(d_2, \pi_*) &= 1.653061 \\
B(d_3, \pi_*) &= 1.653061
\end{aligned}
$$

For a randomized rule $\delta_R = p_1 d_1 + p_2 d_2 + p_3 d_3$ where $p_1 + p_2 + p_3 = 1$ the Bayes-Risk can be computed by:

$$
\begin{aligned}
B(\delta_R, \pi) &= \sum_{i=1}^{3} \pi_*(\theta_i) \left(\sum_{j=1}^{3} p_j R(\theta_i, d_j) \right) \\
&= \sum_{i=1}^{3} \sum_{j=1}^{3} \pi_*(\theta_i) p_j R(\theta_i, d_j) \\
&= \sum_{j=1}^{3} p_j \sum_{i=1}^{3} \pi_*(\theta_i) R(\theta_i, d_j) \\
&= \sum_{j=1}^{3} p_j B(d_j, \pi_*)
\end{aligned}
$$

But because $B(d_1, \pi_*) = B(d_2, \pi_*) = B(d_3, \pi_*) = C$ and $p_1 + p_2 + p_3 = 1$ it holds that:

$$
B(\delta_R, \pi_*) = \sum_{j=1}^{3} p_j B(d_j, \pi_*) = (p_1 + p_2 + p_3) C = C
$$

So every randomized rule will have a Bayes-risk of C. Therefore δ_* is a Bayes-rule under the prior π_* as well as every random and non-random rule. Choice simply does not matter.

To show that δ_* is also a mini-max rule, we first need to show that π_* is the **least favorable prior-distribution.**

If we consider the the agent being indifferent between the decision, every decision needs to yield the same risk under the prior π.

For this we need to solve the linear system:

$$
\begin{bmatrix} 3 & 6 & 0 \\ 1 & 1 & 2 \\ 7 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} C \\ C \\ C \end{bmatrix}
$$

It turns out, that π_* solves this linear system.

b)

We have discussed before, that $R(\theta, \delta) = L(\theta, \delta)$. So the table given before, is at the same time a table for the loss. Therefore d_2 is the non-random mini-max rule.