## Problem 3.1

**1.**

First state the general problem:

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} \mathcal{L}(x|\theta) = \arg\max_{\theta\in\Theta} \ln \mathcal{L}(x|\theta)$$

Maximum likelihood function:

$$\begin{aligned}
\mathcal{L}(x|\theta) &= \prod_{i=1}^{n} f(x_i|\theta) \\
&= \prod_{i=1}^{n} \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right) \\
&= \theta^{-n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right)
\end{aligned}$$

Now log likelihood:

$$\ln \mathcal{L}(x|\theta) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

FOC:

$$\begin{aligned}
\frac{\partial \ln \mathcal{L}(x|\theta)}{\partial\theta} &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0 \\
\frac{1}{\theta}\sum_{i=1}^{n} x_i &= n \\
\theta_{ML} &= \bar{X}_n
\end{aligned}$$

SOC:

$$\begin{aligned}
\ln \mathcal{L}''(x|\theta) &= \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n} x_i \\
&= \frac{n}{\theta^2} - \frac{2n}{\theta^3}\bar{X}_n
\end{aligned}$$

Means, if we check at $x = \bar{X}_n$:

$$
\begin{aligned}
\ln \mathcal{L}''(\bar{X}_n|\theta) &= \frac{n}{(\bar{X}_n)^2} - \frac{2n}{(\bar{X}_n)^3}\bar{X}_n \\
&= \frac{n-2n}{\bar{X}_n^2} < 0
\end{aligned}
$$

**2.**

Unbiasedness $\mathbb{E}(\bar{X}_n) = \theta$:

$$
\begin{aligned}
\mathbb{E}(\bar{X}_n) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\theta \\
&= \frac{n}{n}\theta = \theta
\end{aligned}
$$

Variance:

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

**3.**

Bias variance decomposition:

$$
\begin{aligned}
R(\delta,\theta) &= \mathbb{E}((\delta(X)-\theta)^2) \\
&= Bias(\delta)^2 + Var(\delta)
\end{aligned}
$$

But as we look only at unbiased estimators our risk becomes:

$$
R(\delta,\theta) = Var(\delta)
$$

And because $Var(\bar{X}_n) = \frac{\theta^2}{n}$ and therefore has the lowest variance, it also has the lowest risk and is therefore admissible.

**4.**

First compute the bias:

$$
\begin{aligned}
\mathbb{E}(\hat{\theta}_2) &= \mathbb{E}\left(\frac{1}{n+1}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n+1}\sum_{i=1}^{n}\mathbb{E}(X_i) \\
&= \frac{n}{n+1}\theta \\
\implies Bias(\hat{\theta}_2) &= \frac{n}{n+1}\theta - \theta \\
&= -\frac{1}{n+1}\theta
\end{aligned}
$$

Then compute the variance

$$
\begin{aligned}
Var(\hat{\theta}_2) &= \frac{1}{(n+1)^2}\sum_{i=1}^{n} Var(X_i) \\
&= \frac{n}{(n+1)^2}\theta^2
\end{aligned}
$$

And finally combine in the Bias variance decomposition:

$$
\begin{aligned}
R(\hat{\theta}_2, \theta) &= Bias(\hat{\theta}_2)^2 + Var(\hat{\theta}_2) \\
&= \frac{1}{(n+1)^2}\theta^2 + \frac{n}{(n+1)^2}\theta^2 \\
&= \frac{(n+1)}{(n+1)^2}\theta^2 \\
&= \frac{1}{n+1}\theta^2 < \frac{1}{n}\theta^2 = R(L(\hat{\theta}_{ML}, \theta))
\end{aligned}
$$

So even-though the max. likelihood est. is an admissible unbiased estimator, the estimator is not overall admissible.

## Problem 3.2

**1.**

$$
\begin{aligned}
&\Theta = (0,1) \\
&\mathscr{X} = \{0, .., k\}^n \\
&D = \Theta \\
&L = L(\theta, d) = \min\left\{2, \left(\frac{d-\theta}{\theta}\right)^2\right\}
\end{aligned}
$$

**2.**

Let $\delta_M \in D$ a decision rule, then $\delta_M$ is minimax iff:

$$\sup_{\theta \in \Theta} R(\delta, \theta) = \inf_{\delta \in D} \sup_{\theta \in \Theta} R(\delta, \theta)$$

Where $R$ is the expected loss under parameter $\theta$.

**3.**

$$L(\theta, \delta_0) = \min\left\{2, \left(-\frac{\theta}{\theta}\right)^2\right\} = 1 \quad \forall \theta \in \Theta$$

If we were to choose another randomized rule, this rule would need to be positive $\delta(X) = c > 0$.

In that case it would hold, that:

$$g(x, \theta) = \left(\frac{c - \theta}{\theta}\right)^2 = \left(\frac{c}{\theta} - \frac{\theta}{\theta}\right)^2 = \left(\frac{c}{\theta} - 1\right)^2$$

In that case we can choose $\theta \to 0$ and get $g(x, \theta) \to \infty$ and therefore a loss $R(\theta, \delta) = L(\theta, \delta) = 2$.

So there is no other rule, with a smaller "worst case" scenario.

## Problem 3.3

**a)**

$$\begin{aligned} \Theta &= H_0 \cup H_1 = \mathbb{R}_+ \\ \mathscr{X} &= \mathbb{R}^n \\ D &= \{\delta : \mathscr{X} \to \{0, 1\}\} \\ L &= \text{0-1 loss} \end{aligned}$$

Because $\sigma$ is given, we can assume, that the parameter space is just the space of $\mu$.

**b)**

Do we need to show this?

Assuming iid we can say (because any linear combination of normally distributed random variables is also normal):

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Likewise we can say:

$$\bar{X}_n - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

And finally:

$$Z_n = \frac{(\bar{X}_n - \mu)\sqrt{n}}{\sigma} \sim \mathcal{N}(0,1)$$

Because:

$$\begin{aligned} Var(Z_n) &= Var\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)\right) \\ &= \frac{n}{\sigma^2} Var(\bar{X}_n - \mu) \\ &= \frac{n}{\sigma^2} Var(\bar{X}_n) \\ &= \frac{n}{\sigma^2} \cdot \frac{\sigma^2}{n} = 1 \end{aligned}$$

since we know, that $\sigma^2 = 4$ we can now make the statement:

$$\begin{aligned} P\left(Z_n \geq q_{0.975}\right) &= 2.5\% \\ \iff \quad P\left(Z_n \geq 1.96\right) &= 2.5\% \\ \iff \quad P\left(\frac{(\bar{X}_n - \mu)\sqrt{n}}{2} \geq 1.96\right) &= 2.5\% \\ \iff \quad P_{H_0}\left(\frac{\bar{X}_n\sqrt{n}}{2} \geq 1.96\right) &= 2.5\% \end{aligned}$$

Because this exactly equivalent to the critical region $C_n$ of our test we have now computed the probability for a type 1 error, i.e. the significance level. This means, that the significance level of the test is equal to 2.5%.

**c)**

We are looking the solution to the following term:

$$P_{H_1}(t(\vec{X}) \geq 1.96) = 1 - P_{H_1}\left(\frac{\bar{X}_n\sqrt{n}}{2} < 1.96\right)$$

Meaning the test statistic realizes a value outside of the critical region under the alternative. Under the alternative the term has the distribution:

$$t(\vec{X}) \sim \mathcal{N}(\mu, 1)$$

meaning that we can describe the probability of a type 2 error like so:

$$1 - \Phi\left(t(\vec{X}) - \frac{\mu\sqrt{n}}{2}\right)$$

So the power is dependent on the size of $\mu$. The higher $\mu$ the higher the power.

**d)**

$$\begin{aligned}
\lambda(\vec{X}) &= \frac{\mathcal{L}(\vec{X}, 0)}{\sup_{\mu\in(0,\infty)} \mathcal{L}(\vec{X}, \mu)} \\
&= \frac{\left[\frac{1}{\sqrt{2\pi\sigma}}\right]^n}{\left[\frac{1}{\sqrt{2\pi\sigma}}\right]^n} \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right) \\
&= \exp\left(2\sum_{i=1}^{n} x_i \bar{x}_n - n\right)
\end{aligned}$$

**e)**

Question: is the likelihood ratio montonic?

**f)**

We can express the risk depending on weither $H_0$ is true, or the alternative:

$$R(t, \mu) = \begin{cases} 2.5\% & \text{if } \mu = 0 \\ 1 - \Phi\left(t(\vec{X}) - \frac{\mu\sqrt{n}}{2}\right) & \text{else} \end{cases}$$

and yes the function is piecewise monotonic, because it jumps to 97,5% at $\mu \to 0$ and then decreases. For $\mu = 0$ it is a single value.