Probability Theory

Problem 1.1

a)

Let Ω be a sample space and $\mathcal{A}$ a σ-Algebra on that sample space. Then the mapping $P : \mathcal{A} \to [0, 1]$ is a probability measure iff:

1. $P(A) \geq 0 \forall A \in \mathcal{A}$
2. $P(\Omega) = 1$
3. $P\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

b)

$$
\begin{aligned}
\sigma(\mathcal{E}) &= \{\emptyset, \Omega, \{1\}, \{2,3,4\}, \{4\}, \{1,2,3\}, \{1,4\}, \{2,3\}\} \\
\mathcal{P}(\Omega) &= \{\emptyset, \Omega, \{1\}, \{2,3,4\}, \{4\}, \{1,2,3\}, \{1,4\}, \{2,3\}, \{1,3\}, \{1,2\}, \\
&\quad \{2,4\}, \{3,4\}, \{1,3,4\}, \{1,2,4\}, \{2\}, \{3\}\}
\end{aligned}
$$

c)

So we have been given:

$$P(A) = Q(A) = c \quad \forall A \in \mathcal{E}$$

Also this means that:

$$P(A^c) = Q(A^c) = 1 - Q(A) = 1 - c \quad \forall A \in \mathcal{E}$$

This means, that (using Lemma 1.2):

$$P(\{1,4\}) = Q(\{1,4\}) = 2c$$

Finally we can say :

$$P(\{1,4\}^c) = Q(\{1,4\}^c) = 1 - Q(\{1,4\}^c) = 1 - 2c$$

So it holds that:

$$P(A) = Q(A) \quad \forall A \in \sigma(\mathcal{E})$$

d)

Since $c = 0$ is allowed you can easily find a counter example with:

$$
\begin{aligned}
P(A) &= \delta_3(A) \\
Q(A) &= \delta_2(A)
\end{aligned}
$$

And it holds that:

$$
\begin{aligned}
&P(\{1\}) = Q(\{1\}) = 0 \\
&P(\{4\}) = Q(\{4\}) = 0 \\
\implies &P, Q \in \mathcal{M}
\end{aligned}
$$

But for $\{3\} \in \mathcal{P}(\Omega)$ it holds that:

$$
P(\{3\}) = 1 \neq 0 = Q(\{3\})
$$

So the statement does not hold for this example.

e)

$$
\begin{aligned}
\sum_{\omega \in \Omega} \omega P(\{\omega\}) &= \arg \max_{a,b,c \in [0,1]} 1c + 2a + 3b + 4c \\
&= \arg \max_{a,b,c \in [0,1]} 5c + 2a + 3b
\end{aligned}
$$

With the side condition that $2c + a + b = 1, c < \frac{1}{2}$. Because the optimization is linear in every input, we just have to just maximize the inputs, that maximize the expected value the most. This means:

$$
\begin{aligned}
c_{best} &= \frac{1}{2} - \epsilon \\
a_{best} &= 0 \\
b_{best} &= \frac{1}{2} + \epsilon
\end{aligned}
$$

With $\epsilon > 0$ but arbitrarily small. The expected value then becomes:

$$
\begin{aligned}
\sum_{\omega \in \Omega} \omega P(\{\omega\}) &= 5\left(\frac{1}{2} - \epsilon\right) + 3\left(\frac{1}{2} + \epsilon\right) \\
&= 2.5 - 5\epsilon + 1.5 + 3\epsilon \\
&= 4 - 2\epsilon
\end{aligned}
$$

But since you can't find the smallest element of $\epsilon > 0$ (because the real line is dense), there is no actual solution to this optimization.

Problem 1.2

a)

It trivially holds that:

$$
f_\lambda(x) = \lambda \exp(-\lambda x) \geq 0 \quad \forall x \in \mathbb{R}
$$

because $\lambda > 0$ and $\exp(x) > 0$ for all $x \in \mathbb{R}$.

Secondly:

$$
\begin{aligned}
\int_{\mathbb{R}} f_\lambda(x)dx &= \int_0^\infty \lambda \exp(-\lambda x)dx \\
&= \lambda \left[-\frac{1}{\lambda} \exp(-\lambda x) \right]_0^\infty \\
&= \lambda \left[\lim_{x\to\infty} -\frac{1}{\lambda}\exp(-\lambda x) - \left(\frac{1}{\lambda}\exp(0) \right) \right] \\
&= \lambda \left[0 - \left(-\frac{1}{\lambda} \right) \right] \\
&= \frac{\lambda}{\lambda} = 1
\end{aligned}
$$

b)

$$
\begin{aligned}
\varphi_X(t) = \mathbb{E}(\exp(itX)) \\
M_X(t) = \mathbb{E}(\exp(tX))
\end{aligned}
$$

These functions are **always** well defined, uniformly continuous and bounded. Also the characteristic function is unique for every probability distribution, and therefore fully *characterizes* the distribution.

c)

$$
\begin{aligned}
M_X(t) &= \mathbb{E}(\exp(tX)) \\
&= \int_0^\infty \exp(tx)\lambda \exp(-\lambda x)dx \\
&= \lambda \int_0^\infty \exp(x(t-\lambda))dx \\
&= \lambda \left[\frac{1}{t-\lambda} \exp(x(t-\lambda)) \right]_0^\infty \\
&= \lambda \left(\lim_{x\to\infty} \frac{1}{t-\lambda} \exp(x(t-\lambda)) \right) - \frac{\lambda}{t-\lambda}\exp(0) \\
&\underset{t<\lambda}{=} 0 - \frac{\lambda}{t-\lambda} = \frac{\lambda}{\lambda - t}
\end{aligned}
$$

d)

$$
\begin{aligned}
\mathbb{E}(X) &= M'_X(0) \\
&\underset{\text{chain rule}}{=} -1\frac{\lambda}{(\lambda-0)^2}(-1) \\
&= \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} \\
\mathbb{E}(X^2) &= M''_X(0) \\
&= -2\frac{\lambda}{(\lambda-0)^3}(-1) \\
&= 2\frac{1}{\lambda^2} \\
Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\
&= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}
$$

e)

$$
\begin{aligned}
P\left(\frac{X-\frac{1}{\lambda}}{\frac{1}{\lambda}} \leq t\right) &= P(\lambda X - 1 \leq t) \\
&= P(\lambda X \leq 1 + t) \\
&\leq \frac{\mathbb{E}(\lambda X)}{1+t} \\
&= \frac{\frac{\lambda}{\lambda}}{1+t} = \frac{1}{1+t}
\end{aligned}
$$

f)

$$
\begin{aligned}
\mathbb{E}(X-Y|X+Y) &= \mathbb{E}(X|X+Y) - \mathbb{E}(Y|X+Y) \\
&\underset{\text{independent}}{=} \frac{1}{\lambda} - \frac{1}{\lambda} = 0
\end{aligned}
$$

Problem 1.3

a)

i)

$$\forall \epsilon > 0 : P(|X_n - X| > \epsilon) \to 0$$

ii)

$$\mathbb{E}((X_n - X)^p) \to 0$$

b)

Let $\epsilon > 0$, then:

$$
\begin{aligned}
P(|X_n| > \epsilon) &\leq P(X_n = n^2) \\
&= \frac{1}{n} \to 0
\end{aligned}
$$

c)

$$
\begin{aligned}
\mathbb{E}(X_n) &= 0P(X_n = 0) + n^2 P(X_n = n^2) \\
&= n^2 \frac{1}{n} = n \to \infty \neq 0
\end{aligned}
$$

No no convergence in 1st mean given

d)

because of the statement written in the exercise we can state (because a normal distribution is stochastically bounded):

$$
\sqrt{n}\left(\begin{pmatrix} Y_n \\ Z_n \end{pmatrix} - \begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}\right) = \mathcal{O}_p(1)
$$

But also:

$$
n\left(\begin{pmatrix} Y_n \\ Z_n \end{pmatrix} - \begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}\right) = \mathcal{O}_p(\sqrt{n})
$$

e)

Using:

$$
\phi(x, y) = \begin{pmatrix} \ln x + \ln y \\ \ln x - \ln y \end{pmatrix}
$$

The derivative becomes:

$$
\phi'(x, y) = \begin{pmatrix} \frac{1}{x} & \frac{1}{y} \\ \frac{1}{x} & -\frac{1}{y} \end{pmatrix}
$$

So (inserting the mean values):

$$
\phi'(\mu_Y, \mu_Z) = \begin{pmatrix} \frac{1}{\mu_Y} & \frac{1}{\mu_Z} \\ \frac{1}{\mu_Y} & -\frac{1}{\mu_Z} \end{pmatrix}
$$

applying this to the given covariance structure yields:

$$\begin{pmatrix} \frac{1}{\mu_Y} & \frac{1}{\mu_Z} \\ \frac{1}{\mu_Y} & -\frac{1}{\mu_Z} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\mu_Y} & \frac{1}{\mu_Y} \\ \frac{1}{\mu_Z} & -\frac{1}{\mu_Z} \end{pmatrix} = \begin{pmatrix} \frac{1}{\mu_Y} & \frac{1}{\mu_Z} \\ \frac{1}{\mu_Y} & -\frac{1}{\mu_Z} \end{pmatrix} \begin{pmatrix} \frac{1}{\mu_Y} + \frac{\rho}{\mu_Z} & \frac{1}{\mu_Y} - \frac{\rho}{\mu_Z} \\ \frac{\rho}{\mu_Y} + \frac{1}{\mu_Z} & \frac{\rho}{\mu_Y} - \frac{1}{\mu_Z} \end{pmatrix}$$

This then leads to:

$$\begin{pmatrix} \frac{1}{\mu_Y}\left(\frac{1}{\mu_Y} + \frac{\rho}{\mu_Z}\right) + \frac{1}{\mu_Z}\left(\frac{\rho}{\mu_Y} + \frac{1}{\mu_Z}\right) & \frac{1}{\mu_Y}\left(\frac{1}{\mu_Y} - \frac{\rho}{\mu_Z}\right) + \frac{1}{\mu_Z}\left(\frac{\rho}{\mu_Y} - \frac{1}{\mu_Z}\right) \\ \frac{1}{\mu_Y}\left(\frac{1}{\mu_Y} + \frac{\rho}{\mu_Z}\right) - \frac{1}{\mu_Z}\left(\frac{\rho}{\mu_Y} + \frac{1}{\mu_Z}\right) & \frac{1}{\mu_Y}\left(\frac{1}{\mu_Y} - \frac{\rho}{\mu_Z}\right) - \frac{1}{\mu_Z}\left(\frac{\rho}{\mu_Y} - \frac{1}{\mu_Z}\right) \end{pmatrix} = \tilde{\Sigma}$$

Asymptotically it then holds that:

$$\sqrt{n}\left(\begin{pmatrix} \ln Y_n + \ln Z_n \\ \ln Y_n - \ln Z_n \end{pmatrix} - \begin{pmatrix} \ln \mu_Y + \ln \mu_Z \\ \ln \mu_Y - \ln \mu_Z \end{pmatrix}\right) \to \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tilde{\Sigma}\right)$$

And for asymptotic independence $\tilde{\Sigma}_{1,2} = \tilde{\Sigma}_{2,1} = 0$ must hold. The only example i found was:

$$\mu_Y = \mu_Z = \rho = 1$$

Decision Theory

Problem 2.1

a)

First let's compute the likelihood:

$$\begin{aligned} \mathcal{L}(X|\theta) &= \prod_{i=1}^{n} f(x_i|\theta) \\ &= \prod_{i=1}^{n} \frac{1}{\theta} \\ &= \frac{1}{\theta^n} \end{aligned}$$

So the likelihood is decreasing in θ. Therefore we need to take the smallest θ where all of realizations are still in the support of the resulting density.

$$\implies \hat{\theta}_{ML} = \max_{i \in \{1, \ldots, n\}} x_i$$

b)

From the way the cdf is defined, we already know that the support of f is the interval $[0, \theta)$ and there the pdf is just the derivative of the cdf:

$$\implies f(x) := \begin{cases} 0 & , x < 0 \\ \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} & , 0 \leq x < \theta \\ 0, & \text{else} \end{cases}$$

And for the expectation:

$$\begin{aligned} \mathbb{E}(\hat{\theta}_{ML}) &= \int_0^\theta x \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx \\ &= \int_0^\theta n\left(\frac{x}{\theta}\right)^n dx \\ &= n\int_0^1 u^n \theta du \\ &= n\theta\left[\frac{u^{n+1}}{n+1}\right]_0^1 \\ &= \theta\frac{n}{n+1} \end{aligned}$$

So the estimator is biased but asymptotically unbiased.

c)

We again are going to use bias variance decomposition:

$$\begin{aligned} \mathbb{E}((\theta - \tilde{\theta})^2) &= Var(\tilde{\theta}) + Bias(\tilde{\theta})^2 \\ &= \frac{n}{(n+2)(n+1)^2}\theta^2 + 0 \quad \text{unbiased} + \text{hint} \end{aligned}$$

For $\bar{\theta}$ we are going to first show, that it is unbiased:

$$\begin{aligned} \mathbb{E}(\tilde{\theta}) &= 2\frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) \\ &= 2\frac{1}{n}n\frac{\theta}{2} = \theta \end{aligned}$$

Then we continue with the loss:

$$\begin{aligned}
\mathbb{E}((\theta-\bar{\theta})^2) &= Var(\bar{\theta}) \quad \text{unbiased} \\
&= \frac{4}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{4}{n^2} \sum_{i=1}^{n} Var(X_i) \\
&= \frac{4}{n^2} \sum_{i=1}^{n} \frac{\theta^2}{12} \\
&= \frac{4}{n^2} n \frac{\theta^2}{12} \\
&= \frac{\theta^2}{3n}
\end{aligned}$$

Now that we have these two risks we can compare them:

$$\begin{aligned}
\frac{n}{(n+2)(n+1)^2}\theta^2 &\leq \frac{\theta^2}{3n} \\
\frac{1}{(n+2)(n+1)^2} &\leq \frac{1}{3} \\
(n+2)(n+1)^2 &\geq 3 \\
n &\geq 1
\end{aligned}$$

This means, that the corrected maximum likelihood estimator is R-better.

Problem 2.2

a)

The set is not **minimally** essentially complete, because the lower boundary is also essentially complete, we therefore have found a proper subset, that is essentially complete which proves the statement

b)

Because the half-circle is also a proper subset and complete. Therefore proving, that the proposed set is only complete but not minimal

c)

The half-circle of the lower border is a minimally complete set, because all rules are admissible meaning, that there is no R-better rule in the decision set for every element of the proposed set of the half-circle

d)

No, because the set is convex. If you take two points (or rules) of the set, their line will be above the proposed line (so all the randomized rules are R-worse than the rules from the set)

e)

Because all the rules are admissible, we only need to find an equalizer rule:

$$
\begin{aligned}
R(\theta_1, \delta_M) &= R(\theta_0, \delta_M) \\
y &= 1 - \sqrt{1 - (y-1)^2} \\
1 - y &= \sqrt{1 - (y-1)^2} \\
(1-y)^2 &= 1 - (y-1)^2 \\
(1-y)^2 + (y-1)^2 &= 1 \\
1 - 2y + y^2 + y^2 - 2y + 1 &= 1 \\
0 &= 2y^2 - 4y + 1 \\
0 &= y^2 - 2y + \frac{1}{2} \\
y_{1,2} &= 1 \pm \sqrt{1 - \frac{1}{2}} \\
y &= 1 - \sqrt{0.5}
\end{aligned}
$$

Problem 2.3

a)

$$
\begin{aligned}
\Theta &= \{0.3, 0.7\} \\
\mathscr{X} &= \{0,1\}^2 \\
D &= \{\delta_0, \delta_1\} = \begin{cases} H_0 \text{ is true} \\ H_1 \text{ is true} \end{cases} \\
L &= \text{0-1 loss}
\end{aligned}
$$

b)

We can say:

$$Y \sim \text{Binom}(2, p)$$

And this makes the calculation a lot easier:

$$
\begin{aligned}
&P_{p_0}(Y=0) = (1-p_0)^2 = 0.49 \\
&P_{p_0}(Y=1) = 2p_0(1-p_0) = 2 * 0.3 * 0.7 = 0.42 \\
&P_{p_0}(Y=2) = p_0^2 = 0.3^2 = 0.09
\end{aligned}
$$

On the other hand for $p_1 = 0.7$ we get:

$$P_{p_1}(Y=0) = (1-p_1)^2 = 0.09$$
$$P_{p_1}(Y=1) = 2p_1(1-p_1) = 2*0.7*0.3 = 0.42$$
$$P_{p_1}(Y=2) = p_1^2 = 0.7^2 = 0.49$$

For the accept and reject probabilities we get:

$$\begin{aligned}
P_{p_0}(\delta(Y,\{0.7\})=0) &= P_{p_0}(Y\neq 2) = 0.91\\
P_{p_0}(\delta(Y,\{0.7\})=1) &= P_{p_0}(Y=2) = 0.09\\
P_{p_1}(\delta(Y,\{0.7\})=0) &= P_{p_1}(Y\neq 2) = 0.51\\
P_{p_1}(\delta(Y,\{0.7\})=1) &= P_{p_1}(Y=2) = 0.49
\end{aligned}$$

So Type 1 (false rejection) is pretty low at 9% but false acceptance is even higher than correct acceptance at 49%.

c)

Like said in **b)** the test statistic is binomial distributed by definition:

$$Y \sim \text{Binom}(2,p)$$

And for any specific α for the size of the test, yes we can, by randomization:

$$\delta_\alpha(y,\{0.7\}) = \begin{cases} X \sim \text{Bernoulli}\left(\min\left\{1, \frac{0.09}{\alpha}\right\}\right) & , y=2 \\ 0 & , \text{else} \end{cases}$$

d)

The risk is the expected loss:

$$\mathbb{E}(L(\delta,p))$$

With the 0-K_i loss we get:

$$\mathbb{E}(L(\delta(y,\{0.7\},p))) = \begin{cases} P(\text{type 1 error})2K_0 & , \text{ if } p=0.3 \\ P(\text{type 2 error})K_0 & , \text{ if } p=0.7 \end{cases}$$

e)

To calculate the Bayes risk we have to compute:

$$B(\delta,P) = \sum_{i=0}^{1} R(\delta,p_i)P(p=p_i)$$

Now p is a random variable and now the computation changes to:

$$
\begin{aligned}
\sum_{i=0}^{1} R(\delta, p_i) P(p = p_i) &= P(p = p_0) P(\text{type 1 error}) 2K_0 \\
&+ P(p = p_1) P(\text{type 2 error}) K_0 \\
&= \frac{2}{3} 0.18 K_0 + \frac{1}{3} 0.51 K_0
\end{aligned}
$$

f)

Make it an equalizer rule:

$$
\begin{aligned}
R(\delta, p_0) &= R(\delta, p_1) \\
\frac{2}{3} 0.09 c K_0 &= \frac{1}{3} 0.51 K_0 \\
c &= \frac{1}{2} \frac{0.51}{0.09} \\
c &= 2.889
\end{aligned}
$$