MATH5945 Statistical Inference Assignment 1

Problem One

(a) Continuous Random Vector

$$f_{X,Y}(x,y) = 12xy^3, \quad 0 < x < y, \quad 0 < y < c$$

(i) Find c

Since total probability should integrate to 1:

$$\int \int f_{X,Y}(x,y)\ dx\ dy = 1$$

Further replacing the limits and function

$$\int_0^c \int_0^y 12xy^3\ dx\ dy = 1$$

Solving the inner integral:

$$\int_0^y 12xy^3\ dx = 6y^3 \left[x^2 \Big|_0^y \right] = 6y^5$$

Solving the outer integral using solve inner integral

$$\int_0^c 6y^5\ dy = 1$$

$$\frac{1}{6}y^6 \Big|_0^c = 1$$

$$c^6 = 1 \Rightarrow c = 1$$

Thus, $c = 1$.

(ii) Find Marginal Density $f_X(x)$ and Cumulative Distribution $F_X(x)$

Firstly, find Marginal Density $f_X(x)$

$$f_X(x) = \int_x^c 12xy^3 dy.$$

$$f_X(x) = 12x\frac{y^4}{4}\Big|_x^c = 3x(c^4 - x^4)$$

as c = 1

$$f_X(x) = 3x(c^4 - x^4)$$

Cumulative distribution function (CDF),

$$F_X(x) = P(X \leq x) = \int_0^x f_X(t)\, dt.$$

$$F_X(x) = \int_0^x 3t(c^4 - t^4)\, dt.$$

Splitting the integral using the difference rule

$$F_X(x) = 3\left[c^4 \int_0^x t\, dt - \int_0^x t^5\, dt\right].$$

solving

$$F_X(x) = 3\left(c^4\frac{t^2}{2}\Big|_0^x - \frac{t^6}{6}\Big|_0^x\right) = 3\left[\frac{c^4x^2}{2} - \frac{x^6}{6}\right].$$

$$F_X(x) = \frac{3}{2}c^4x^2 - \frac{x^6}{2}, \quad 0 < x < 1.$$

The marginal density and cumulative distribution function

$$f_X(x) = 3x(c^4 - x^4), \quad 0 < x < 1.$$

$$F_X(x) = \frac{3}{2}c^4x^2 - \frac{x^6}{2}, \quad 0 < x < 1.$$

(iii) Marginal Density $f_Y(y)$ and $F_Y(y)$

$$f_Y(y) = \int_0^y 12xy^3 dx.$$

$$f_Y(y) = 12y^3\frac{x^2}{2}\Big|_y^c = 6y^5$$

$$f_Y(y) = 6y^5$$

Cumulative distribution

$$F_Y(y) = \int_0^y 6t^5 dt = \frac{6y^6}{6}$$

$$F_Y(y) = y^6$$

(iv) Conditional Density $f_{Y|X}(y|x)$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{12xy^3}{3x(c^4 - x^4)}$$

$$f_{Y|X}(y|x) = \frac{4y^3}{(c^4 - x^4)}$$

(v) Conditional Expectation $E(Y|X = x)$

$$E(Y|X = x) = \int_x^c y f_{Y|X}(y|x)\, dy.$$

From part (iv), the cdf is:

$$f_{Y|X}(y|x) = \frac{4y^3}{c^4 - x^4}, \quad x < y < c$$

So,

$$E(Y|X = x) = \int_x^c y \cdot \frac{4y^3}{c^4 - x^4}\, dy$$

$$= \frac{4}{c^4 - x^4} \int_x^c y^4\, dy$$

Solving the integral

$$E(Y|X = x) = \frac{4}{c^4 - x^4} \times \left[\frac{y^5}{5}\right]_x^c$$

$$= \frac{4}{c^4 - x^4} \times \left(\frac{c^5}{5} - \frac{x^5}{5}\right).$$

$$= \frac{4}{5} \times \frac{c^5 - x^5}{c^4 - x^4}.$$

Thus, the conditional expectation is:

$$E(Y|X = x) = \frac{4}{5} \times \frac{c^5 - x^5}{c^4 - x^4}, \quad 0 < x < c$$

(b) Zoom Meeting Birthday Problem

$$P(\text{shared}) = 1 - P(\text{unique})$$

$n = 40$ participants. Probability $P(\text{shared})$ that no two share a birthday is:

$$P(\text{unique}) = \frac{365 \times 364 \times ... \times 365 - 39}{365}$$

We have to consider n-1 since the person themselves should not be considered

$$P(\text{unique}) = \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right)$$

Applying natural logarithm and using Taylor series for $\ln(1-x) \approx -x - \frac{x^2}{2} - \cdots$ (for a small x) and ignoring higher order terms:

$$\ln\left(1 - \frac{k}{365}\right) \approx -\frac{k}{365} - \frac{k^2}{2 \times 365^2}$$

Summing over all k

$$\ln(P(\text{unique})) \approx -\sum_{k=1}^{39} \left(\frac{k}{365} + \frac{k^2}{2 \times 365^2}\right)$$

Simplifying

$$\ln(P(\text{unique})) \approx -\left(\frac{\sum_{k=1}^{39} k}{365} + \frac{\sum_{k=1}^{39} k^2}{2 \times 365^2}\right)$$

Using $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ and $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$ and substituting $n = 40$

$$\ln(P(\text{unique})) \approx -\left(\frac{39(39+1)}{2 \times 365} + \frac{39(39+1)(2(39)+1}{6 \times 2 \times 365^2}\right)$$

$$\ln(P(\text{unique})) \approx -2.137 - 0.077 = -2.214$$

We can be rewrite $P(\text{unique})$:

$$P(\text{unique}) \approx e^{-2.214} \approx 0.109$$

$$P(\text{shared}) \approx 1 - 0.109 = 0.891 \quad (\text{or } \boxed{89.1\%}).$$

This confirms the probability is very close to 90%.

Problem Two

Given cdf:

$$F_X(x) = P(X \le x) = \begin{cases} 0, & x < 1 \\ 1 - \frac{1}{x^3}, & x \ge 1 \end{cases}$$

(i) Verify $F_X(x)$ as a CDF

A valid CDF must:

$$1. \quad \lim_{x \to -\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F_X(x) = 1 \tag{1}$$

$$2. \quad F_X(x) \text{ is non-decreasing} \tag{2}$$

$$3. \quad F_X(x) \text{ is right-continuous} \tag{3}$$

1. Limits boundaries

- **As $x \to -\infty$:** For $x < 1$, $F_X(x) = 0$. Thus,

$$\lim_{x\to-\infty} F_X(x) = 0.$$

- **As $x \to \infty$:** For $x \geq 1$,

$$\lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} \left(1 - \frac{1}{x^3}\right) = 1 - 0 = 1.$$

Both boundary conditions are satisfied.

2. Non-Decreasing

- **For $x < 1$:** $F_X(x) = 0$, which is constant (non-decreasing).
- **For $x \geq 1$:**

$$\frac{d}{dx} F_X(x) = \frac{d}{dx}\left(1 - \frac{1}{x^3}\right) = \frac{3}{x^4}.$$

  Since $\frac{3}{x^4} > 0$ for all $x \geq 1$, $F_X(x)$ is strictly increasing.

- **At $x = 1$:**

$$\text{Left-hand limit: } \lim_{x\to1^-} F_X(x) = 0,$$
$$\text{Right-hand value: } F_X(1) = 1 - \frac{1}{1^3} = 0.$$

  The function transitions from 0 to 0 at $x = 1$,

Thus, $F_X(x)$ is non-decreasing over $\mathbb{R}$.

3. Right-Continuity

- **For $x < 1$:** $F_X(x) = 0$, which is continuous (and hence right-continuous).
- **For $x \geq 1$:** $1 - \frac{1}{x^3}$ is continuous for $x > 1$
- **At $x = 1$:** Check right-continuity:

$$\text{Right-hand limit: } \lim_{x\to1^+} F_X(x) = \lim_{x\to1^+} \left(1 - \frac{1}{x^3}\right) = 0,$$
$$\text{Value at } x = 1\text{: } F_X(1) = 0.$$

  Since the right-hand limit equals $F_X(1)$, the function is right-continuous at $x = 1$.

$F_X(x)$ satisfies all three CDF properties. Thus, $F_X(x)$ is a valid cumulative distribution function.

(ii) Finding the density $f_X(x)$

The density function is given by:

$$f_X(x) = \frac{d}{dx} F_X(x).$$

Differentiating

$$f_X(x) = \frac{d}{dx}\left(1 - \frac{1}{x^3}\right)$$

$$f_X(x) = \frac{3}{x^4}, \quad x \geq 1$$

For all other x, $f_X(x) = 0$, ensuring a valid probability density.

(iii) Transforming to a new random variable Z

If the low-water mark is reset to 0 and use a unit of measurement that is $\frac{1}{10}$ of the previous one, then:

$$Z = 10(X - 1)$$

Thus, the new CDF is:

$$F_Z(z) = P(Z \leq z) = P(10(X-1) \leq z) = P(X \leq 1 + \frac{z}{10})$$

Substituting X:

$$F_Z(z) = F_X(1 + \frac{z}{10}) = 1 - \frac{1}{(1 + \frac{z}{10})^3}, \quad z \geq 0$$

To find the density:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{3}{10(1 + \frac{z}{10})^4}, \quad z \geq 0$$

Problem Three

(a) Spam Classification Probability

- $P(\text{spam}) = 0.3$, Probability an email is spam.
- $P(\text{predicted spam}|\text{spam}) = 0.9$, Probability the model correctly flags spam.
- $P(\text{predicted spam}|\text{not spam}) = 0.15$, Probability the model incorrectly flags non-spam.

Using Bayes' Theorem to find $P(\text{spam}|\text{predicted spam})$:

$$P(\text{spam}|\text{predicted spam}) = \frac{P(\text{predicted spam}|\text{spam}) \cdot P(\text{spam})}{P(\text{predicted spam})}$$

where the total probability of predicting spam is:

$$\begin{aligned}
P(\text{predicted spam}) &= P(\text{predicted spam}|\text{spam}) \cdot P(\text{spam}) + P(\text{predicted spam}|\text{not spam}) \cdot P(\text{not spa} \\
&= (0.9 \times 0.3) + (0.15 \times 0.7) \\
&= 0.27 + 0.105 = 0.375
\end{aligned}$$

Thus:

$$P(\text{spam}|\text{predicted spam}) = \frac{0.27}{0.375} = 0.72$$

If the model predicts spam, there is a **72% probability** the email is truly spam.

(b) Bayesian Estimation with Geometric Likelihood

(i) Interpretation of Conditional Distribution

The conditional distribution $f_{X_1|\Theta}(x|\theta) = \theta^x(1-\theta)$, for $x = 0, 1, 2, \ldots$, describes a **geometric distribution**. Specifically, X_1 given $\Theta = \theta$ represents x failures before the first success in a sequence of independent Bernoulli trials, where:

- θ is the probability of failure in a single trial
- $1 - \theta$ is the probability of success

For example, if $\theta = 0.6$, probability of $x = 2$ failures before the first success is $(0.6)^2(1 - 0.6) = 0.144$.

(ii) Posterior Distribution and Bayes Estimator

- **Prior**: $\tau(\theta) = 30\theta^4(1-\theta)$, which is a Beta distribution with parameters $\alpha = 5$ and $\beta = 2$, i.e., $\text{Beta}(5, 2)$.
- **Likelihood**: For n i.i.d. observations $X = (X_1, \ldots, X_n)$, the likelihood function is:

$$L(X|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta) = \theta^{\sum x_i}(1-\theta)^n.$$

By Bayes' theorem, the posterior distribution $h(\theta|X)$ is proportional to the product of the prior and likelihood:

$$h(\theta|X) \propto \tau(\theta) \cdot L(X|\theta) = 30\theta^4(1-\theta) \cdot \theta^{\sum x_i}(1-\theta)^n = 30\theta^{4+\sum x_i}(1-\theta)^{1+n}.$$

This is the kernel of a Beta distribution with parameters:

$$\alpha' = 5 + \sum_{i=1}^{n} x_i, \quad \beta' = 2 + n.$$

Thus, the posterior is $\text{Beta}(\alpha', \beta')$ or $\text{Beta}(5 + \sum_{i=1}^{n} x_i, 2 + n)$

The Bayes estimator for θ under quadratic loss is the posterior mean:

$$\hat{\theta}_{\text{Bayes}} = E(\theta|X) = \frac{\alpha'}{\alpha' + \beta'} = \frac{5 + \sum_{i=1}^{n} x_i}{7 + n + \sum_{i=1}^{n} x_i}$$

(iii) Hypothesis Testing with Zero-One Loss

Given the data $X = (2, 3, 5, 3, 5, 4, 2)$, we find the posterior parameters:

$$\sum x_i = 2 + 3 + 5 + 3 + 5 + 4 + 2 = 24, \quad n = 7 \implies \alpha' = 29, \quad \beta' = 9$$

The posterior is $\text{Beta}(29, 9)$. To test $H_0 : \theta \leq 0.80$ vs $H_1 : \theta > 0.80$, compute $P(\theta > 0.80|X)$:

$$P(\theta \leq 0.80|X) = \int_0^{0.80} \frac{\theta^{28}(1-\theta)^8}{B(29, 9)} d\theta$$

Using R code to evaluate the integral:

```
pbeta(0.80, 29, 9)  # Result: approx 0.686
```

Since $P(\theta \leq 0.80|X) \approx 68.6\% > 50\%$, we **accept** H_0.

Problem Four

(a) Loss Function Definition

Let the states of nature and actions be defined as:

- States of nature: $\Theta = \{\theta_0, \theta_1\}$ where
  - θ_0: Stock is profitable
  - θ_1: Stock is not profitable
- Actions: $\mathcal{A} = \{a_0, a_1\}$ where
  - a_0: Invest in the stock
  - a_1: Don't invest in the stock

$$L(\theta, a) = \begin{cases} 0 & \text{if } (\theta, a) = (\theta_0, a_0) \text{ (invest in profitable stock)} \\ 0 & \text{if } (\theta, a) = (\theta_1, a_1) \text{ (avoid unprofitable stock)} \\ 1 & \text{if } (\theta, a) = (\theta_0, a_1) \text{ (missed opportunity)} \\ 4 & \text{if } (\theta, a) = (\theta_1, a_0) \text{ (bad investment)} \end{cases}$$

	a_0 (Invest)	a_1 (Don't Invest)
θ_0 (Profitable)	0	1
θ_1 (Unprofitable)	4	0

(b) Probability Mass Function (PMF) for X

The random variable X = 0, 1, 2 follows a binomial distribution with n = 2 under both states of nature. Teams act independently, so for $n = 2$ trials:

Case 1: Stock is Profitable (θ_0) Each team recommends investing with probability $p = \frac{5}{6}$. The PMF is:

$$P(X = k \mid \theta_0) = \binom{2}{k} \left(\frac{5}{6}\right)^k \left(\frac{1}{6}\right)^{2-k}, \quad k = 0, 1, 2$$

$$P(X = 0 \mid \theta_0) = \binom{2}{0} \left(\frac{5}{6}\right)^0 \left(\frac{1}{6}\right)^2 = \frac{1}{36}$$
$$P(X = 1 \mid \theta_0) = \binom{2}{1} \left(\frac{5}{6}\right)^1 \left(\frac{1}{6}\right)^1 = \frac{10}{36}$$
$$P(X = 2 \mid \theta_0) = \binom{2}{2} \left(\frac{5}{6}\right)^2 \left(\frac{1}{6}\right)^0 = \frac{25}{36}$$

Case 2: Stock is Not Profitable (θ_1) Each team recommends investing with probability $p = \frac{1}{2}$. The PMF is:

$$P(X = k \mid \theta_1) = \binom{2}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{2-k} = \binom{2}{k} \left(\frac{1}{2}\right)^2, \quad k = 0, 1, 2$$

$$P(X = 0 \mid \theta_1) = \binom{2}{0} \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$
$$P(X = 1 \mid \theta_1) = \binom{2}{1} \left(\frac{1}{2}\right)^2 = \frac{1}{2}$$
$$P(X = 2 \mid \theta_1) = \binom{2}{2} \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

Table of PMFs

X	0	1	2
$P(X \mid \theta_0)$	$\frac{1}{36}$	$\frac{10}{36}$	$\frac{25}{36}$
$P(X \mid \theta_1)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(c) Risk Points for Non-Randomized Decision Rules

The risk function $R(\theta, d_i)$ represents the expected loss for decision rule d_i under state θ. Calculated as:

$$R(\theta_0, d_i) = \sum_{x=0}^{2} L(\theta_0, d_i(x)) \cdot P(X = x \mid \theta_0)$$

$$R(\theta_1, d_i) = \sum_{x=0}^{2} L(\theta_1, d_i(x)) \cdot P(X = x \mid \theta_1)$$

Risk Calculations

- **Loss Function**:

$$L(\theta, a) = \begin{cases} 0 & (\theta_0, a_0) \\ 1 & (\theta_0, a_1) \\ 4 & (\theta_1, a_0) \\ 1 & (\theta_1, a_1) \end{cases}$$

- **PMFs from Part (b)**:

X	0	1	2
$P(X \mid \theta_0)$	$\frac{1}{36}$	$\frac{10}{36}$	$\frac{25}{36}$
$P(X \mid \theta_1)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Risk Point Results

Rule	$R(\theta_0, d_i)$	$R(\theta_1, d_i)$
d_1	$0 \cdot \frac{1}{36} + 0 \cdot \frac{10}{36} + 0 \cdot \frac{25}{36} = \boxed{0}$	$4\left(\frac{1}{4} + \frac{1}{2} + \frac{1}{4}\right) = \boxed{4}$
d_2	$1 \cdot \frac{1}{36} + 0 \cdot \frac{10}{36} + 0 \cdot \frac{25}{36} = \boxed{\frac{1}{36}}$	$4\left(0 + \frac{1}{2} + \frac{1}{4}\right) = \boxed{3}$
d_3	$0 \cdot \frac{1}{36} + 1 \cdot \frac{10}{36} + 0 \cdot \frac{25}{36} = \boxed{\frac{10}{36}}$	$4\left(\frac{1}{4} + 0 + \frac{1}{4}\right) = \boxed{2}$
d_4	$1 \cdot \frac{1}{36} + 1 \cdot \frac{10}{36} + 0 \cdot \frac{25}{36} = \boxed{\frac{11}{36}}$	$4\left(0 + 0 + \frac{1}{4}\right) = \boxed{1}$
d_5	$0 \cdot \frac{1}{36} + 0 \cdot \frac{10}{36} + 1 \cdot \frac{25}{36} = \boxed{\frac{25}{36}}$	$4\left(\frac{1}{4} + \frac{1}{2} + 0\right) = \boxed{3}$
d_6	$1 \cdot \frac{1}{36} + 0 \cdot \frac{10}{36} + 1 \cdot \frac{25}{36} = \boxed{\frac{26}{36}}$	$4\left(0 + \frac{1}{2} + 0\right) = \boxed{2}$
d_7	$0 \cdot \frac{1}{36} + 1 \cdot \frac{10}{36} + 1 \cdot \frac{25}{36} = \boxed{\frac{35}{36}}$	$4\left(\frac{1}{4} + 0 + 0\right) = \boxed{1}$
d_8	$1 \cdot \frac{1}{36} + 1 \cdot \frac{10}{36} + 1 \cdot \frac{25}{36} = \boxed{1}$	$0 + 0 + 0 = \boxed{0}$

(d) Minimax Rule Determination

The minimax rule minimizes the maximum risk for all states of nature. For each decision rule d_i:

$$\text{Max Risk}(d_i) = \max\{R(\theta_0, d_i), R(\theta_1, d_i)\}$$

Compute Maximum Risks

Rule	Maximum Risk
d_1	$\max\{0, 4\} = 4$
d_2	$\max\left\{\frac{1}{36}, 3\right\} = 3$
d_3	$\max\left\{\frac{10}{36}, 2\right\} = 2$
d_4	$\max\left\{\frac{11}{36}, 1\right\} = 1$
d_5	$\max\left\{\frac{25}{36}, 3\right\} = 3$
d_6	$\max\left\{\frac{26}{36}, 2\right\} = 2$
d_7	$\max\left\{\frac{35}{36}, 1\right\} = 1$
d_8	$\max\{1, 0\} = 1$

Identify Minimax Rule(s)

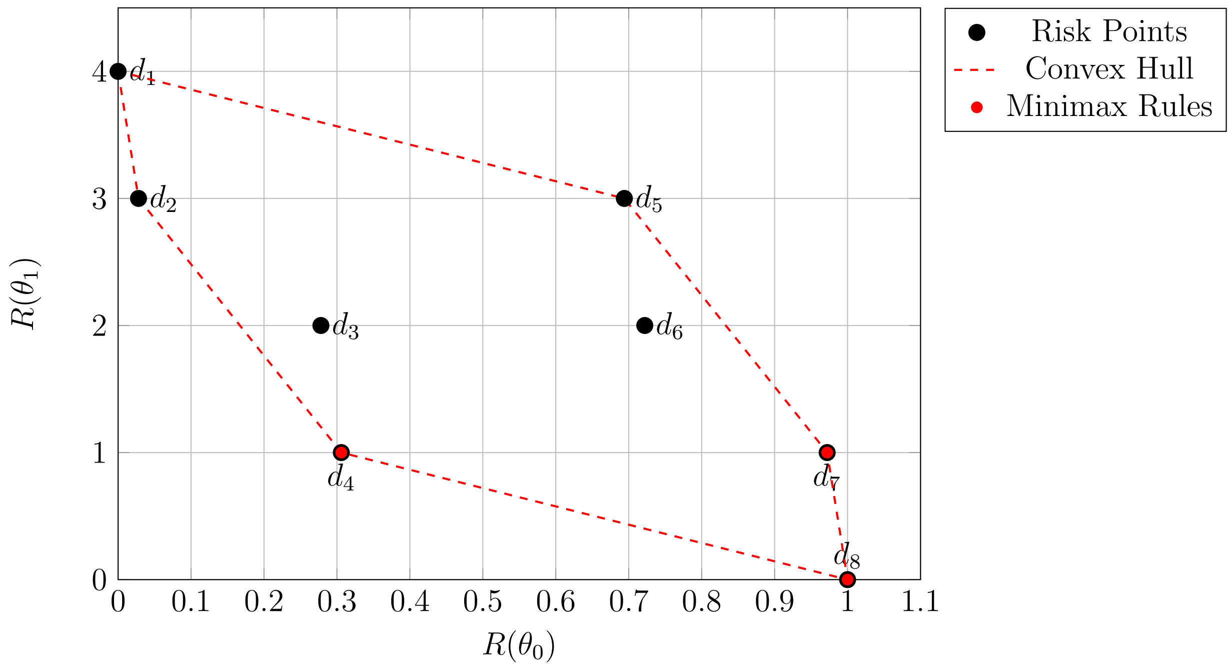
The smallest maximum risk is $\boxed{1}$, achieved by:

$$d_4, \quad d_7, \quad \text{and } d_8$$

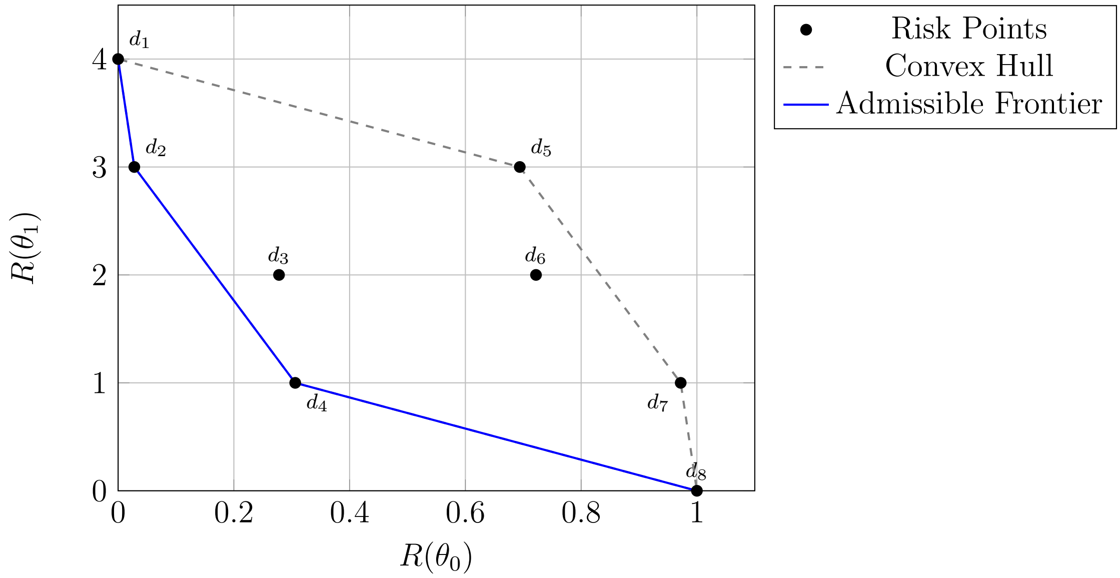
Thus, the minimax rule(s) are $\boxed{d_4}$, $\boxed{d_7}$, and $\boxed{d_8}$.

(e) Risk Set of Randomized Rules

Convex Hull and Risk Set Visualization



(f) Admissible Decision Rules.



- Blue is the lower boundary of the convex hull, connecting $d_1 \to d_2 \to d_4 \to d_8$. These rules are **not dominated** by any other rule.
- **Inadmissible Rules**: d_3, d_5, d_6, d_7 lie inside the convex hull and are strictly dominated by other rules

(g) Minimax Risk in Randomized Decision Rules

Minimax Risk in Non-Randomized Rules D

From part (d), the minimax rules in D are d_4, d_7, d_8, each with maximum risk:

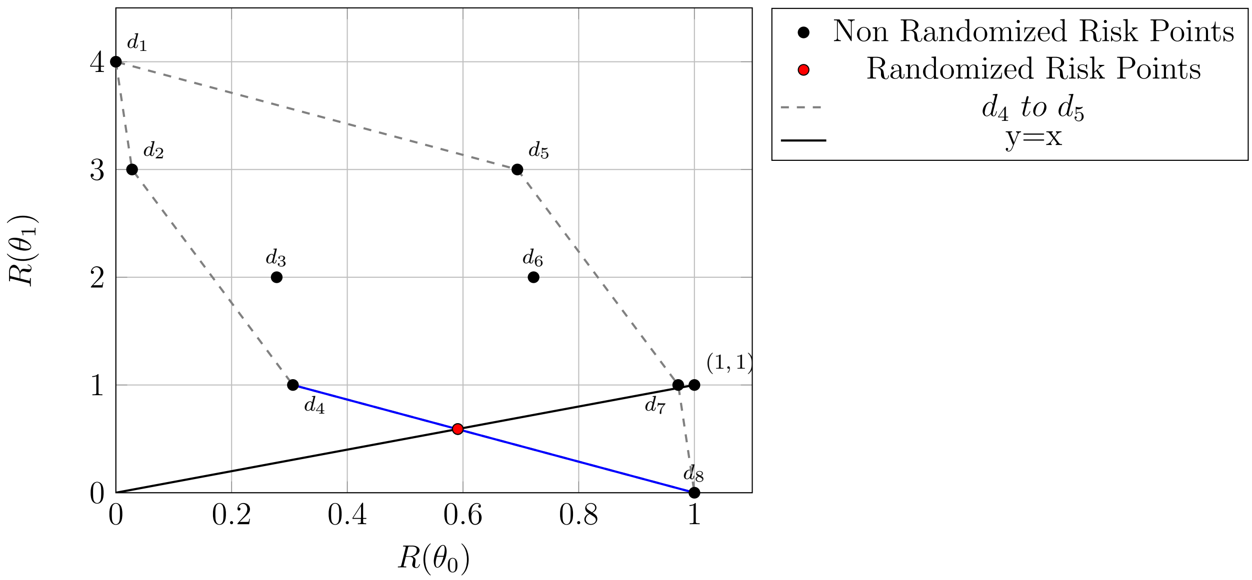
$$\max\{R(\theta_0, d_i), R(\theta_1, d_i)\} = 1$$

Thus, the minimax risk for non-randomized rules is $\boxed{1}$

Minimax Risk in Randomized Rules $\mathcal{D}$

The minimax risk corresponds to the smallest r such that the line $\max\{R(\theta_0), R(\theta_1)\} = r$ touches the convex hull. This occurs where $R(\theta_0) = R(\theta_1) = r$.

Intersection on the Admissible Frontier The admissible frontier is the line segment from d_4 to d_8. Parametrize this segment:



$$R(\theta_0) = \frac{11}{36} + t(1 - \frac{11}{36}) = 0.306 + 0.694t,$$
$$R(\theta_1) = 1 - t(1 - 0) = 1 - t, \quad t \in [0, 1]$$

Set $R(\theta_0) = R(\theta_1)$:

$$0.306 + 0.694t = 1 - t \implies 1.694t = 0.694 \implies t \approx 0.409$$

At $t \approx 0.409$:

$$R(\theta_0) = R(\theta_1) \approx 0.590$$

Thus, the minimax risk for randomized rules is:

$$\boxed{0.59}.$$

- The minimax risk decreases from 1.00 (non-randomized) to 0.59 (randomized).
- By mixing rules (e.g., combining d_4 and d_8), the fund manager can achieve a risk profile where neither $R(\theta_0)$ nor $R(\theta_1)$ dominates excessively.

(h) Minimax Rule in Randomized Decision Rules

The minimax rule in the set of randomized decision rules $\mathcal{D}$ is defined as a probability distribution over the non-randomized rules $\{d_1, d_2, \ldots, d_8\}$ that minimizing the worst-case risk across all states of nature θ. Formally, it is a convex combination:

$$d_{\text{minimax}} = \sum_{i=1}^{8} \lambda_i d_i$$

where $\lambda_i \geq 0$, $\sum_{i=1}^{8} \lambda_i = 1$

Expressing the Minimax Rule in Terms of Non-Randomized Rules : For the given problem, the minimax rule corresponds to a mixture of:

$$d_{\text{minimax}} = \lambda d_4 + (1-\lambda)d_8$$

where $\lambda \in [0,1]$ is chosen such that:

$$R(\theta_0, d_{\text{minimax}}) = R(\theta_1, d_{\text{minimax}}).$$

Risks of d_4 and d_8

$$d_4 : R(\theta_0, d_4) = \frac{11}{36} \approx 0.306, \quad R(\theta_1, d_4) = 1,$$
$$d_8 : R(\theta_0, d_8) = 1, \quad R(\theta_1, d_8) = 0.$$

Convex Combination

$$R(\theta_0, d_{\text{minimax}}) = \lambda \cdot \frac{11}{36} + (1-\lambda)\cdot 1$$

$$R(\theta_1, d_{\text{minimax}}) = \lambda \cdot 1 + (1-\lambda)\cdot 0 = \lambda$$

Equalize Risks

$$\lambda \cdot \frac{11}{36} + (1-\lambda) = \lambda$$

Solving for λ:

$$\lambda \left(\frac{11}{36} - 1\right) + 1 = \lambda \implies \lambda = \frac{36}{61} \approx 0.590$$

Resulting Risks

$$R(\theta_0, d_{\text{minimax}}) = R(\theta_1, d_{\text{minimax}}) = \frac{36}{61} \approx 0.59$$

The minimax rule in the randomized set $\mathcal{D}$ is explicitly defined as:

$$\boxed{d_{\text{minimax}} = \frac{36}{61}d_4 + \frac{25}{61}d_8}$$

(i) Prior for Which Minimax Rule is Bayes

A Bayes rule minimizes **Bayes risk** for a prior $\pi = (\pi_0, \pi_1)$:

$$\text{Bayes Risk} = \pi_0 R(\theta_0, d) + \pi_1 R(\theta_1, d)$$

For d_{minimax} to be a Bayes rule, there must exist a prior $\pi = (\pi_0, \pi_1)$:

$$\pi_0 R(\theta_0, d_{\text{minimax}}) + \pi_1 R(\theta_1, d_{\text{minimax}}) \leq \pi_0 R(\theta_0, d) + \pi_1 R(\theta_1, d), \quad \forall d \in \mathcal{D}.$$

Solving for the Prior

The minimax rule's equalized risks imply:

$$\pi_0 \cdot \frac{36}{61} + \pi_1 \cdot \frac{36}{61} \leq \pi_0 R(\theta_0, d) + \pi_1 R(\theta_1, d), \quad \forall d \in \mathcal{D}.$$

Simplifying:

$$\frac{36}{61}(\pi_0 + \pi_1) \leq \pi_0 R(\theta_0, d) + \pi_1 R(\theta_1, d).$$

Since $\pi_0 + \pi_1 = 1$, this reduces to:

$$\frac{36}{61} \leq \pi_0 R(\theta_0, d) + \pi_1 R(\theta_1, d), \quad \forall d \in \mathcal{D}.$$

$$\boxed{\pi(\theta_0) = \frac{36}{61}, \quad \pi(\theta_1) = \frac{25}{61}}$$

(j) Bayes Rule and Bayes Risk with Prior $\pi(\theta_0) = \frac{1}{2}$

The fund manager's prior belief is:

$$\pi(\theta_0) = \frac{1}{2}, \quad \pi(\theta_1) = \frac{1}{2}$$

where θ_0 = "stock is profitable" and θ_1 = "stock is not profitable."

Using the PMFs from part (b), compute the posterior probabilities $P(\theta \mid X)$:

For $X = 0$:

$$P(\theta_0 \mid X = 0) = \frac{\frac{1}{36} \cdot \frac{1}{2}}{\frac{1}{36} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2}} = \frac{1}{10}$$

$$P(\theta_1 \mid X = 0) = \frac{9}{10}$$

For $X = 1$:

$$P(\theta_0 \mid X = 1) = \frac{\frac{10}{36} \cdot \frac{1}{2}}{\frac{10}{36} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{5}{14}$$

$$P(\theta_1 \mid X = 1) = \frac{9}{14}$$

For $X = 2$:

$$P(\theta_0 \mid X = 2) = \frac{\frac{25}{36} \cdot \frac{1}{2}}{\frac{25}{36} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2}} = \frac{25}{34}$$

$$P(\theta_1 \mid X = 2) = \frac{9}{34}$$

Bayes Rule

For each X, the posterior expected loss for actions a_0 (invest) and a_1 (do not invest):

For $X = 0$:

$$\text{Expected loss for } a_0 = 4 \cdot \frac{9}{10} = 3.6$$
$$\text{Expected loss for } a_1 = 1 \cdot \frac{1}{10} = 0.1 \quad \Rightarrow \quad \text{Choose } a_1$$

For $X = 1$:

$$\text{Expected loss for } a_0 = 4 \cdot \frac{9}{14} \approx 2.57$$
$$\text{Expected loss for } a_1 = 1 \cdot \frac{5}{14} \approx 0.36 \quad \Rightarrow \quad \text{Choose } a_1$$

For $X = 2$:

$$\text{Expected loss for } a_0 = 4 \cdot \frac{9}{34} \approx 1.06$$
$$\text{Expected loss for } a_1 = 1 \cdot \frac{25}{34} \approx 0.74 \quad \Rightarrow \quad \text{Choose } a_1$$

The Bayes rule is to **always choose** a_1 (do not invest), regardless of X.

$$d_{\text{Bayes}}(X) = a_1 \quad \forall X \in \{0, 1, 2\}$$

Bayes Risk

The Bayes risk is the expected loss under the prior:

$$\text{Bayes Risk} = \frac{1}{2} \cdot L(\theta_0, a_1) + \frac{1}{2} \cdot L(\theta_1, a_1) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \boxed{\frac{1}{2}}$$

(k) ϵ-Minimax Rules on the Risk Set

For $\epsilon = 0.1$, we illustrate the risk points of all ϵ-minimax rules on the risk set.

The admissible frontier connects:

$$d_4 : \left(\frac{11}{36} \approx 0.306, 1\right) \quad \text{and} \quad d_8 : (1, 0).$$

The equation of this line can be written in the form $y = mx + c$ where: calculating m slope

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0 - 1}{1 - 0.306} \approx -1.44$$

calculating y-intercept c using (1,0)

$$0 = -1.44 \times 1 + c$$

$$c = 1.44$$

replacing m and c

$$y = -1.44x + 1.44$$

The ϵ-minimax rules satisfy:

$$\max\{R(\theta_0), R(\theta_1)\} \leq 0.69.$$

This defines following region

$$(0, 0.69), \quad (0.69, 0.69), \quad (0.69, 0), \quad (0, 0).$$

- **Intersection with** $R(\theta_1) = 0.69$:

  $$0.69 = -1.44x + 1.44 \implies x \approx 0.52$$

  Intersection point: $(0.521, 0.69)$.

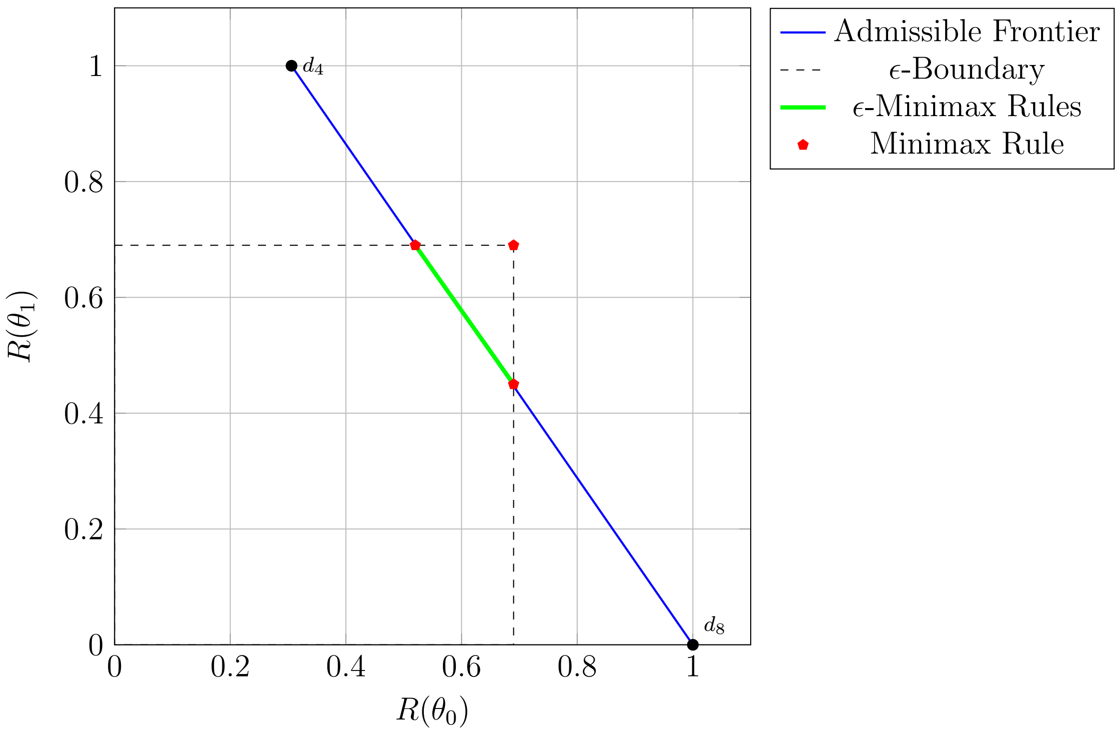
- **Intersection with** $R(\theta_0) = 0.69$:

  $$y = -1.44(0.69) + 1.44 \approx 0.45$$

  Intersection point: $(0.69, 0.45)$.

The risk points which are ϵ-minimax:

$$(0.52, 0.69) \quad , \quad (0.69, 0.45) \quad \text{and} \quad (0.69, 0.69)$$



Problem Five

(a) Proof that $E(T) = \int_0^\infty S(t)\,dt$

Given a continuous non-negative random variable T with survival function $S(t) = P(T > t)$, we aim to prove:

$$E(T) = \int_0^\infty S(t)\,dt.$$

The survival function:

$$S(t) = P(T > t)$$

where $P(T > t) = F(t)$

$$S(T) = P(T > t) = F(t) = \int_t^\infty f(t)\,dt.$$

differentiating both sides

$$\frac{d}{dt}S(t) = -f(t)$$

rearranging

$$f(t) = -\frac{d}{dt}S(t)$$

Expected Value for a continuous random variable:

$$E(T) = \int_0^\infty t f(t)\,dt$$

Applying integration by parts:$\int u\,dv = uv - \int v\,du$

$$E(T) = [-tS(t)]_0^\infty + \int_0^\infty S(t)\,dt.$$

Boundary Term Evaluation:

- As $t \to \infty$: $tS(t) \to 0$ (since $S(t) \to 0$ decreases faster than $t \to \infty$ increasing).
- As $t \to 0$: $tS(t) \to 0$.

Thus:

$$[-tS(t)]_0^\infty = 0 - 0 = 0.$$

Result:

$$\boxed{E(T) = \int_0^\infty S(t)\,dt}$$

(b)

(i) Hazard Function Relationships

Given the hazard function:

$$h_T(t) = \lim_{\eta \to 0} \frac{P(t \le T < t + \eta \mid T \ge t)}{\eta},$$

we show that:

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)} = -\frac{d}{dt} \log(1 - F_T(t)) = -\frac{d}{dt} \log(S(t)).$$

Proof. Express in terms of CDF and PDF:

$$P(t \le T < t + \eta \mid T \ge t) = \frac{P(t \le T < t + \eta)}{P(T \ge t)} = \frac{F_T(t + \eta) - F_T(t)}{1 - F_T(t)}.$$

Substituting into the hazard function:

$$h_T(t) = \lim_{\eta \to 0} \frac{F_T(t + \eta) - F_T(t)}{\eta \cdot (1 - F_T(t))}.$$

By definition of the PDF $f_T(t) = \lim_{\eta \to 0} \frac{F_T(t+\eta) - F_T(t)}{\eta}$, we get:

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)}.$$

Relate to the survival function $S(t)$: Since $S(t) = 1 - F_T(t)$, we have:

$$\frac{d}{dt} \log(S(t)) = \frac{1}{S(t)} \cdot \frac{d}{dt} S(t) = \frac{-f_T(t)}{S(t)} = -h_T(t).$$

Rearranging gives:

$$h_T(t) = -\frac{d}{dt} \log(S(t)) = -\frac{d}{dt} \log(1 - F_T(t)).$$

□

$$\boxed{h_T(t) = \frac{f_T(t)}{1 - F_T(t)} = -\frac{d}{dt} \log(1 - F_T(t)) = -\frac{d}{dt} \log(S(t))}$$

(ii) Prove that $S(t) = e^{-\int_0^t h_T(x)\, dx}$

We aim to prove that the survival function $S(t)$ is given by:

$$S(t) = e^{-\int_0^t h_T(x)\, dx}.$$

Proof. From part (b)(i), we have:

$$h_T(t) = -\frac{d}{dt}\log(S(t)).$$

Integrate both sides from 0 to t:

$$\int_0^t h_T(x)\,dx = -\int_0^t \frac{d}{dx}\log(S(x))\,dx$$

Simplifies to:

$$\int_0^t h_T(x)\,dx = -\left[\log(S(x))\right]_0^t$$

Apply boundary conditions:

- At $x = 0$, $S(0) = P(T > 0) = 1$, so $\log(S(0)) = \log(1) = 0$
- At $x = t$, $\log(S(t))$ remains

Thus:

$$\int_0^t h_T(x)\,dx = -\log(S(t)).$$

Exponentiate both sides to solve for $S(t)$:

$$\log(S(t)) = -\int_0^t h_T(x)\,dx \implies S(t) = e^{-\int_0^t h_T(x)\,dx}$$

□

$$\boxed{S(t) = e^{-\int_0^t h_T(x)\,dx}}$$

(iii) Verify the Hazard Function

Proof. Exponential distribution of the Survival Function $S(t)$:

$$S(t) = P(T > t) = \int_t^\infty \beta e^{-\beta x}\,dx = \left[-e^{-\beta x}\right]_t^\infty = e^{-\beta t}.$$

Substituting values into Hazard Function $h_T(t)$:

$$h_T(t) = \frac{f_T(t)}{S(t)} = \frac{\beta e^{-\beta t}}{e^{-\beta t}} = \beta.$$

The hazard function $h_T(t) = \beta$ is constant for all $t > 0$ where $\beta > 0$

□

$$\boxed{h_T(t) = \beta}$$

(iv) Hazard Function for Pareto Distribution

$$f_T(t \mid \theta) = \frac{\theta}{t^{\theta+1}}, \quad t \geq 1.$$

The survival function $S_T(t) = P(T > t)$ is derived from the cumulative distribution function (CDF):

$$F_T(t) = \int_1^t \frac{\theta}{x^{\theta+1}}\,dx = 1 - \frac{1}{t^\theta}, \quad t \geq 1.$$

Thus:

$$S_T(t) = 1 - F_T(t) = \frac{1}{t^\theta}, \quad t \geq 1$$

The hazard function is defined as:

$$h_T(t) = \frac{f_T(t)}{S_T(t)}$$

Substitute $f_T(t)$ and $S_T(t)$

$$h_T(t) = \frac{\frac{\theta}{t^{\theta+1}}}{\frac{1}{t^\theta}} = \frac{\theta}{t^{\theta+1}} \cdot t^\theta = \frac{\theta}{t}$$

The hazard function for the Pareto distribution with parameter $\theta > 0$ is:

$$\boxed{h_T(t) = \frac{\theta}{t}, \quad t \geq 1}$$