# 24 Problems

Aayush Bajaj

December 15, 2025

..............................................................................................

## Contents

..............................................................................................

## Rules

As is tradition, the prize pool has increased (to $300 this year).

I have collapsed first and second place into a winner-takes-all arrangement

(c'est la vie).

Furthermore, there are additional changes to the structure of this Game:

1. you must now *pass* the problem set to be awarded the prize money;

2. you may submit your solutions to the problem set at any point in the future;

3. if you plagiarise work, I reserve the right to ban you from all subsequent competitions — grim trigger

4. the problem set is also available on my website

   - there you will find an alternative presentation of these problems upheld with MathJaX, TikZ and JavaScript.
   - the url for the problem set is https://abaj.ai/projects/bday-problems/24th/problems
   - my solutions will be available from the start of 2026; by viewing them you *forfeit the prize money* (obviously)
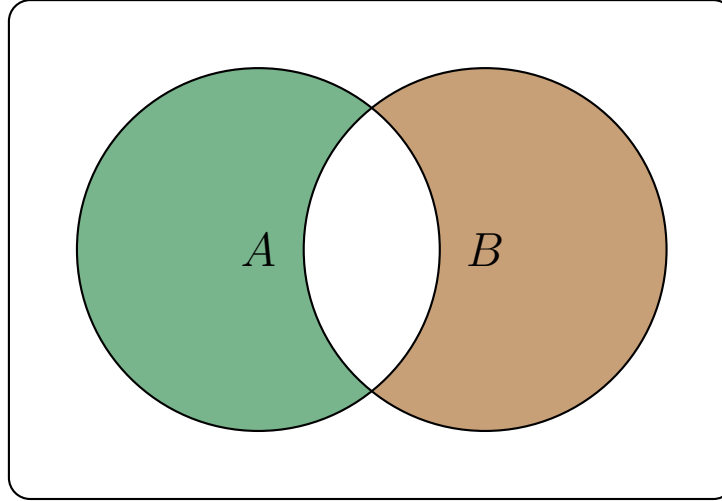
5. Good luck!

..............................................................................................

Submit

# § Problems

1. Given that $A$ and $B$ are sets like so: $\boxed{4}$

   $A$ and $B$ are non-empty sets

   

   (a) Give an expression for the conditional probability $P(A|B)$ $\boxed{1}$

   > **Solution:**
   > $$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

   (b) Hence or otherwise derive Bayes' (first) rule for conditional probability: $\boxed{1}$

   $$P(A|B) = P(B|A) \times \frac{P(A)}{P(B)}$$

   > **Solution:**
   >
   > $$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{as (1)}$$
   > $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$
   > $$\implies P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$$
   > $$\implies P(A|B) = P(B|A) \times \frac{P(A)}{P(B)}$$

   (c) Prove that the symmetric difference $A \Delta B = (A - B) \cup (B - A)$ is the same as $(A \cup B) - (A \cap B)$. $\boxed{2}$

**Solution:**

Let $x \in A \Delta B$.

Then:

- $x \in A - B \implies x \in A$ and $x \notin B$

OR:

- $x \in B - A \implies x \in B$ and $x \notin A$
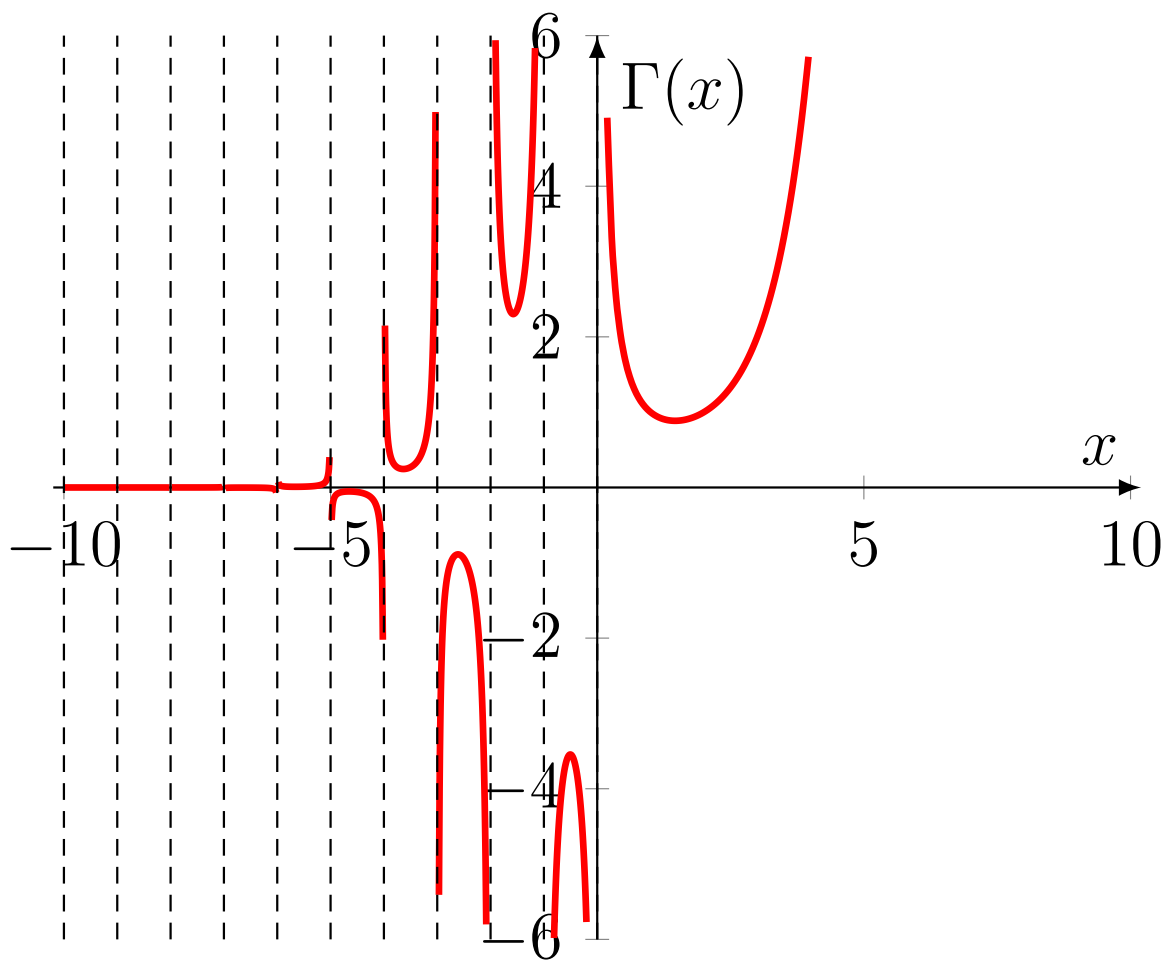
So:

- $x \in A$ or $x \in B \implies x \in (A \cup B)$
- $x \notin (A \cap B) \implies x \in (A \cap B)^c$

Therefore:

$$x \in (A \cup B) \cap (A \cap B)^c = (A \cup B) - (A \cap B)$$

2. The Gamma Function ⁹

(a) Recall the integral of

$$\int e^{-x} \mathrm{d}x$$

<span style="float:right">1</span>

**Solution:**

$$\int e^{-x} \mathrm{d}x = -e^{-x} \tag{2}$$

(b) Apply integration by parts on

$$\int x e^{-x} \mathrm{d}x.$$

<span style="float:right">2</span>

Factorise your result.

**Solution:**

$$= -xe^{-x} + \int e^{-x}\mathrm{d}x$$

$$= -e^{-x}(1+x)$$

(c) These above are special cases of $\alpha = 1$ and $\alpha = 2$. More generally we can compute the integration of

$$\int x^{\alpha-1}e^{-x}\mathrm{d}x.$$

[3]

**Solution:**

$$= -x^{\alpha-1}e^{-x} + (\alpha - 1)\int x^{\alpha-2}e^{-x}\mathrm{d}x \tag{3}$$

$$\tag{4}$$

(d) Fixing the limits of integration leads to the Gamma function, defined by

[3]

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x, \quad \text{for } \alpha > 0$$

Show that with (3), it follows that

$$\Gamma(\alpha) = (\alpha - 1) \times \Gamma(\alpha - 1).$$

Therefore, we have for any integral $k$: $\Gamma(k) = (k-1) \times (k-2) \times \ldots \times 2 \times \Gamma(1)$, and $\Gamma(1) = 1$ due to (2).

**Solution:** From (3), we have:

$$\int x^{\alpha-1}e^{-x}\mathrm{d}x = -x^{\alpha-1}e^{-x} + (\alpha - 1)\int x^{\alpha-2}e^{-x}\mathrm{d}x$$

Evaluating with limits from 0 to $\infty$:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x$$

$$= \left[-x^{\alpha-1}e^{-x}\right]_0^\infty + (\alpha - 1)\int_0^\infty x^{\alpha-2}e^{-x}\mathrm{d}x$$

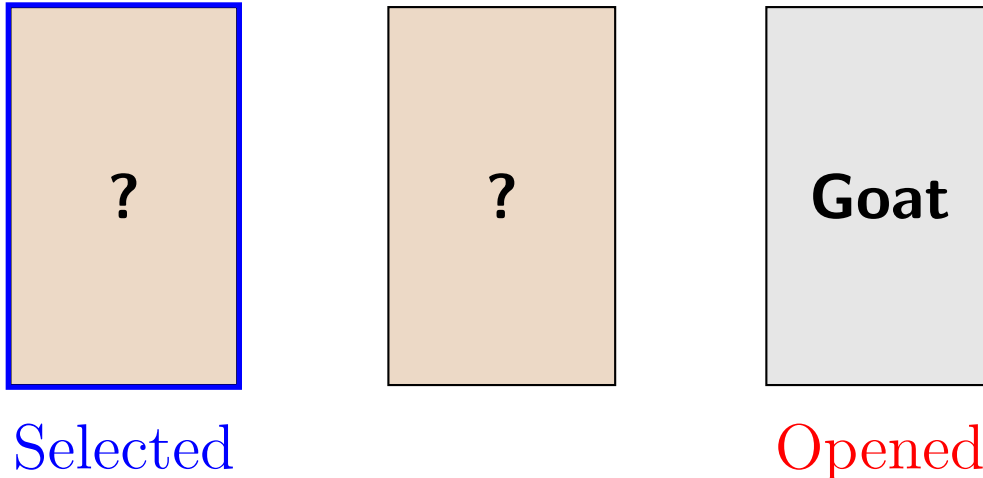For $\alpha > 1$, as $x \to \infty$: $x^{\alpha-1}e^{-x} \to 0$ (exponential dominates polynomial). At $x = 0$: $0^{\alpha-1}e^0 = 0$ for $\alpha > 1$.

Therefore:

$$\Gamma(\alpha) = [0 - 0] + (\alpha - 1) \int_0^\infty x^{\alpha-2} e^{-x} \mathrm{d}x$$
$$= (\alpha - 1)\Gamma(\alpha - 1)$$

3. Monty Hall Problem:

<div style="text-align: center;">

?     ?     **Goat**

Selected     Opened

</div>

5

(a) Suppose there are three curtains. Behind one curtain there is a nice prize, while behind the other two there are worthless prizes (perhaps 1 car and 2 goats). A contestant selects one curtain at random, and then Monte Hall opens one of the other two curtains to reveal a worthless prize (goat). Hall then expresses the willingness to trade the curtain that the contestant has chosen for the other curtain that has not been opened. Should the contentestant switch curtains or stick or stick with the one that she has?

2

**Solution:** She should **switch**.

Let $C$ denote the event "car is behind chosen door" and $S$ denote "car is behind switch door."

Initially: $P(C) = \frac{1}{3}$ and $P(S) = \frac{2}{3}$.

Monty **knows** where the car is and **always** opens a door with a goat. This action does not change the initial probabilities:

- If the car was behind the contestant's door (prob $\frac{1}{3}$), switching loses.

- If the car was behind one of the other two doors (prob $\frac{2}{3}$), Monty must reveal the goat from the remaining door, so switching wins.

6

Using Bayes' theorem: Let $M$ be the event "Monty opens door 3 (revealing goat)." Assuming the contestant chose door 1:

$$P(\text{Car at 2}|M) = \frac{P(M|\text{Car at 2})P(\text{Car at 2})}{P(M)}$$

$$= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

where $P(M) = P(M|\text{Car at 1})P(\text{Car at 1}) + P(M|\text{Car at 2})P(\text{Car at 2}) + P(M|\text{Car at 3})P(\text{Car at 3}) = \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}$

Therefore, switching gives probability $\frac{2}{3}$ of winning vs $\frac{1}{3}$ for staying.

(b) Now consider the variant where the host does not know which door hides the car. What is the **posterior** probability of winning the prize / car if our contestant switches? $\boxed{3}$

**Solution:** When the host **does not know** where the car is and randomly opens a door, we must carefully condition on observing a goat.

Let contestant choose door 1, host randomly opens door 3 revealing a goat. Let $C_i =$ "car behind door $i$", $G_3 =$ "host opens door 3 *and* reveals goat."

$$P(C_2|G_3) = \frac{P(G_3|C_2)P(C_2)}{P(G_3)}$$

Computing the probabilities (host picks uniformly from doors 2 and 3):

- $P(G_3|C_1) = \frac{1}{2}$ (picks door 3 with prob $\frac{1}{2}$, finds goat)

- $P(G_3|C_2) = \frac{1}{2}$ (picks door 3 with prob $\frac{1}{2}$, finds goat)

- $P(G_3|C_3) = 0$ (if picks door 3, finds car, not goat)

Therefore:

$$P(G_3) = P(G_3|C_1)P(C_1) + P(G_3|C_2)P(C_2) + P(G_3|C_3)P(C_3)$$
$$= \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}$$
$$= \frac{1}{3}$$

Thus:

$$P(C_2|G_3) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{3}} = \frac{1}{2}$$

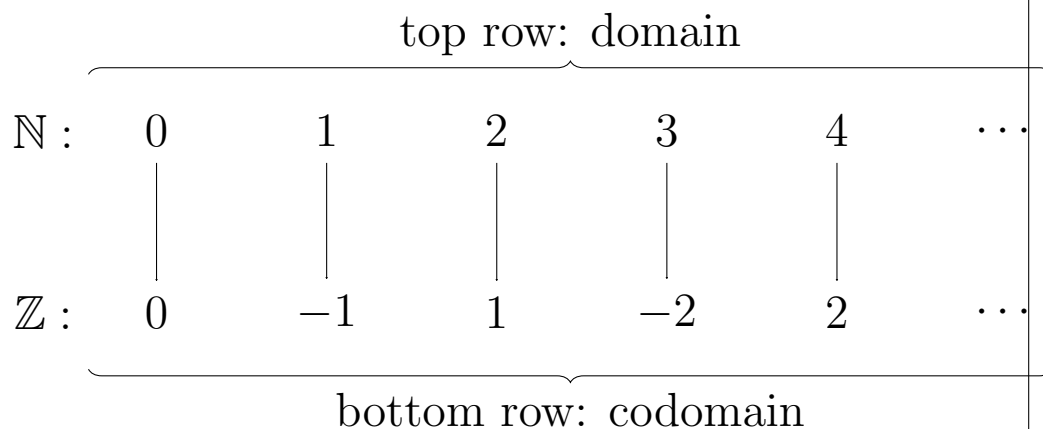> **Answer:** The posterior probability of winning by switching is
>
> $$\boxed{\dfrac{1}{2}}$$
>
> The key insight: when the host doesn't know, revealing a goat provides no information to distinguish between doors 1 and 2, so they become equally likely.

4. (a) Prove that there are equally as many **natural numbers** as there are **integers**, i.e. that $|\mathbb{Z}| = |\mathbb{N}|$. $\boxed{2}$

> **Solution:** Infinite sets are equal in cardinality if there is a bijection between both sets:
>
> Consider $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(2n) = n$ and $f(2n+1) = -(n+1)$. Such a function is injective and surjective. The correspondence is illustrated below.
>
> <div align="center">
>
> top row: domain
>
> | $\mathbb{N}:$ | 0 | 1 | 2 | 3 | 4 | $\cdots$ |
> |---|---|---|---|---|---|---|
> | | $\vert$ | $\vert$ | $\vert$ | $\vert$ | $\vert$ | |
> | $\mathbb{Z}:$ | 0 | $-1$ | 1 | $-2$ | 2 | $\cdots$ |
>
> bottom row: codomain
>
> </div>

(b) Prove that there are equally as many **integers** as there are **rational numbers**, i.e. that $|\mathbb{Z}| = |\mathbb{Q}|$. $\boxed{2}$

> **Solution:**
>
> Every rational number $q \in \mathbb{Q}$ has a unique **reduced** representation
>
> $$q = \frac{a}{b} \quad \text{with } a \in \mathbb{Z},\ b \in \mathbb{N}_{>0},\ \gcd(|a|, b) = 1.$$
>
> Define an injection $\iota : \mathbb{Q} \to \mathbb{Z} \times \mathbb{N}_{>0}$ by
>
> $$\iota(0) = (0, 1), \qquad \iota\left(\frac{a}{b}\right) = (a, b) \quad \text{(in lowest terms, } b > 0\text{)}.$$
>
> This is injective because the reduced representation is unique.

8

Next, $\mathbb{Z} \times \mathbb{N}_{>0}$ is countable: let $g : \mathbb{Z} \to \mathbb{N}$ be the bijection

$$g(z) = \begin{cases} 2z, & z \geq 0, \\ -2z - 1, & z < 0, \end{cases}$$

and let $\pi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be Cantor's pairing function

$$\pi(x, y) = \frac{(x + y)(x + y + 1)}{2} + y.$$

Then
$$h : \mathbb{Z} \times \mathbb{N}_{>0} \to \mathbb{N}, \qquad h(z, b) = \pi\big(g(z), b - 1\big)$$

is a bijection, so $\mathbb{Z} \times \mathbb{N}_{>0}$ is countable.

Since $\mathbb{Q}$ injects into a countable set, $\mathbb{Q}$ is countable. Also $\mathbb{Z} \subseteq \mathbb{Q}$, so $\mathbb{Q}$ is infinite. Hence $\mathbb{Q}$ is **countably infinite**, and therefore

$$|\mathbb{Q}| = |\mathbb{N}| = |\mathbb{Z}|.$$

(c) Prove that the *real numbers* are *uncountable*; i.e. that $|\mathbb{R}| \neq |\mathbb{N}|$.  $\boxed{3}$

**Solution:**

It suffices to show that $[0, 1] \subseteq \mathbb{R}$ is uncountable. Suppose for contradiction that $[0, 1]$ is countable. Then we can list all its elements as a sequence

$$x_1, \ x_2, \ x_3, \ \ldots$$

Write each $x_n$ in a decimal expansion

$$x_n = 0.a_{n1}a_{n2}a_{n3} \cdots$$

choosing the representation that does **not** end in repeating 9s (so the expansion is unique).

Now define a new real number $y \in [0, 1]$ by specifying its digits:

$$y = 0.b_1 b_2 b_3 \cdots, \qquad b_n = \begin{cases} 1, & a_{nn} \neq 1, \\ 2, & a_{nn} = 1. \end{cases}$$

By construction, $b_n \neq a_{nn}$ for every $n$. Hence $y$ differs from $x_n$ in the $n$-th decimal place, so $y \neq x_n$ for all $n$. This contradicts the assumption that $(x_n)$ lists **all** of $[0, 1]$.

Therefore $[0, 1]$ is uncountable, and since $[0, 1] \subseteq \mathbb{R}$, the real numbers are uncountable. In particular,
$$|\mathbb{R}| \neq |\mathbb{N}|.$$

$$
\begin{array}{rcl}
x_1 & = & 0.\ \boxed{a_{11}}\ \ a_{12}\ \ a_{13}\ \ a_{14}\ \ \cdots \\
x_2 & = & 0.\ \ a_{21}\ \ \boxed{a_{22}}\ \ a_{23}\ \ a_{24}\ \ \cdots \\
x_3 & = & 0.\ \ a_{31}\ \ a_{32}\ \ \boxed{a_{33}}\ \ a_{34}\ \ \cdots \\
x_4 & = & 0.\ \ a_{41}\ \ a_{42}\ \ a_{43}\ \ \boxed{a_{44}}\ \ \cdots \\
\vdots & & \\
y & = & 0.\ \ b_1\ \ \ b_2\ \ \ b_3\ \ \ b_4\ \ \cdots
\end{array}
$$

$b_n \neq a_{nn}$ for all $n$
so $y \neq x_n$ for all $n$.

5. (a) What does the sequence $1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \dots$ converge to? (If at all)  $\boxed{2}$

**Solution:** The sequence is $a_n = \frac{1}{n^2}$ for $n = 1, 2, 3, \dots$

To show convergence to 0, we use the $\epsilon$-$N$ definition:

For any $\epsilon > 0$, we need $N$ such that for all $n \geq N$:

$$
\left| \frac{1}{n^2} - 0 \right| < \epsilon
$$

This is equivalent to $\frac{1}{n^2} < \epsilon$, or $n^2 > \frac{1}{\epsilon}$, or $n > \frac{1}{\sqrt{\epsilon}}$.

Choose $N = \left\lceil \frac{1}{\sqrt{\epsilon}} \right\rceil$. Then for all $n \geq N$, we have $|a_n - 0| < \epsilon$.

Therefore:

$$
\boxed{\lim_{n \to \infty} \frac{1}{n^2} = 0}
$$

(b) What does the sequence $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$ converge to? (If at all)  $\boxed{2}$

**Solution:** The sequence is $a_n = \frac{1}{n}$ for $n = 1, 2, 3, \dots$

To show convergence to 0:

For any $\epsilon > 0$, we need $N$ such that for all $n \geq N$:

$$
\left| \frac{1}{n} - 0 \right| < \epsilon
$$

This requires $\frac{1}{n} < \epsilon$, or $n > \frac{1}{\epsilon}$.

Choose $N = \left\lceil \frac{1}{\epsilon} \right\rceil$. Then for all $n \geq N$, we have $|a_n - 0| < \epsilon$.

Therefore:

$$
\boxed{\lim_{n \to \infty} \frac{1}{n} = 0}
$$

> **Note:** While the sequence converges, the corresponding series $\sum_{n=1}^{\infty} \frac{1}{n}$ (the harmonic series) *diverges.*

6. Which is larger asymptotically as $n \to \infty$?

$$2^n \ll n!$$

OR

$$2^n \gg n!$$

Give a proof by induction.

---

**Solution:** We claim that

$$\boxed{2^n \ll n!}$$

for sufficiently large $n$.

**Proof by induction:**

We will prove that $n! > 2^n$ for all $n \geq 4$.

*Base case:* $n = 4$

$$4! = 24 > 16 = 2^4$$

*Inductive step:* Assume $k! > 2^k$ for some $k \geq 4$. We must show $(k+1)! > 2^{k+1}$.

$$\begin{aligned}
(k+1)! &= (k+1) \cdot k! \\
&> (k+1) \cdot 2^k \qquad \text{(by inductive hypothesis)} \\
&> 2 \cdot 2^k \qquad \text{(since } k+1 > 2 \text{ for } k \geq 4) \\
&= 2^{k+1}
\end{aligned}$$

Therefore, by induction, $n! > 2^n$ for all $n \geq 4$.

**Asymptotic analysis:** The ratio $\frac{n!}{2^n}$ grows without bound:

$$\frac{n!}{2^n} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdots n}{2^n} = \frac{1}{2} \cdot \frac{2}{2} \cdot \frac{3}{2} \cdot \frac{4}{2} \cdots \frac{n}{2}$$

For $n \geq 5$, each factor $\frac{k}{2}$ for $k \geq 3$ is $> 1$, and we have infinitely many such factors as $n \to \infty$. Thus $\frac{n!}{2^n} \to \infty$, confirming $n! \gg 2^n$.

---

7. (a) Find the eigenspaces of the following matrices:

   1. $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

2. $\begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 2 & 3 & 0 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

**Solution: (a)** $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

Characteristic polynomial: $\det(A - \lambda I) = (1 - \lambda)^2 = 0$

Eigenvalue: $\lambda = 1$ (multiplicity 2)

Eigenspace for $\lambda = 1$: Solve $(A - I)\mathbf{v} = 0$:

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \implies x = 0$$

Eigenspace: $E_1 = \text{span}\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

**(b)** $B = \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix}$

Characteristic polynomial: $\det(B - \lambda I) = (-2 - \lambda)(1 - \lambda) - 4 = \lambda^2 + \lambda - 6 = (\lambda + 3)(\lambda - 2) = 0$

Eigenvalues: $\lambda_1 = -3$, $\lambda_2 = 2$

For $\lambda_1 = -3$: $(B + 3I)\mathbf{v} = 0$:

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \implies x + 2y = 0$$

$E_{-3} = \text{span}\left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}$

For $\lambda_2 = 2$: $(B - 2I)\mathbf{v} = 0$:

$$\begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \implies -4x + 2y = 0$$

$E_2 = \text{span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$

**(c)** $C = \begin{bmatrix} 2 & 3 & 0 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix}$

Characteristic polynomial: $\det(C - \lambda I) = (1 - \lambda)[(2 - \lambda)(4 - \lambda) - 3] = (1 - \lambda)(\lambda^2 - 6\lambda + 5) = (1 - \lambda)(\lambda - 1)(\lambda - 5) = 0$

Eigenvalues: $\lambda_1 = 1$ (multiplicity 2), $\lambda_2 = 5$

For $\lambda = 1$: $(C - I)\mathbf{v} = 0$:

$$\begin{bmatrix} 1 & 3 & 0 \\ 1 & 3 & 3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

From row 1: $x + 3y = 0$, from row 2: $x + 3y + 3z = 0 \implies z = 0$

$$E_1 = \text{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} \right\}$$

For $\lambda = 5$: $(C - 5I)\mathbf{v} = 0$:

$$\begin{bmatrix} -3 & 3 & 0 \\ 1 & -1 & 3 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

From row 3: $z = 0$, from row 1: $-3x + 3y = 0 \implies x = y$

$$E_5 = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

**(d)** $D = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ (assuming this is $3 \times 3$)

Characteristic polynomial: $\det(D - \lambda I) = (1 - \lambda)(-\lambda)^2 = -\lambda^2(1 - \lambda) = 0$

Eigenvalues: $\lambda_1 = 0$ (multiplicity 2), $\lambda_2 = 1$

For $\lambda = 0$: $D\mathbf{v} = 0$:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \implies x + y = 0$$

$$E_0 = \text{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

For $\lambda = 1$: $(D - I)\mathbf{v} = 0$:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \implies y = z = 0$$

$$E_1 = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

(b) Determine if the following matrices are diagonalisable. If so, determine their [8]
diagonal form and a basis with respect to which the transformation matrices
are diagonal.

If they are not diagonal, give reasons why not.

1. $\begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$

4. $\begin{bmatrix} 5 & 4 & 2 & 1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{bmatrix}$

**Solution: (a)** $A = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix}$

Characteristic polynomial: $\det(A - \lambda I) = -\lambda(4 - \lambda) + 8 = \lambda^2 - 4\lambda + 8 = 0$

$\lambda = \frac{4 \pm \sqrt{16-32}}{2} = \frac{4 \pm 4i}{2} = 2 \pm 2i$

**Not diagonalizable over** $\mathbb{R}$ (complex eigenvalues). However, it is diagonalizable over $\mathbb{C}$.

For $\lambda_1 = 2 + 2i$: $(A - \lambda_1 I)\mathbf{v} = 0$:

$$\begin{bmatrix} -2 - 2i & 1 \\ -8 & 2 - 2i \end{bmatrix} \mathbf{v} = 0$$

Eigenvector: $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 + 2i \end{bmatrix}$

For $\lambda_2 = 2 - 2i$: $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 - 2i \end{bmatrix}$

Over $\mathbb{C}$: $D = \begin{bmatrix} 2 + 2i & 0 \\ 0 & 2 - 2i \end{bmatrix}$, $P = \begin{bmatrix} 1 & 1 \\ 2 + 2i & 2 - 2i \end{bmatrix}$

**(b)** $B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

Characteristic polynomial: Notice rank$(B) = 1$, so $\lambda = 0$ has multiplicity at least 2. tr$(B) = 3$, so eigenvalues sum to 3.

Eigenvalues: $\lambda_1 = 3$, $\lambda_2 = \lambda_3 = 0$

For $\lambda = 3$: $(B - 3I)\mathbf{v} = 0$:

$$\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \mathbf{v} = 0$$

All rows are equivalent to $x + y + z = 0$ after reduction. Eigenvector:
$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

For $\lambda = 0$: $B\mathbf{v} = 0 \implies x+y+z = 0$ Eigenspace: $E_0 = \text{span} \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}$

(dimension 2)

**Diagonalizable:** $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Basis: $P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$

**(c)** $C = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$

Characteristic polynomial (expanding along suitable row/column): $\det(C - \lambda I) = -\lambda^3 + 5\lambda^2 - 8\lambda + 4 = -(\lambda - 1)(\lambda - 2)^2 = 0$

Eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 2$ (multiplicity 2)

For $\lambda = 1$: $(C - I)\mathbf{v} = 0$:
$$\begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \mathbf{v} = 0$$

Solving: $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$

For $\lambda = 2$: $(C - 2I)\mathbf{v} = 0$:
$$\begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix} \mathbf{v} = 0$$

Row reduces to: $x - 2y - 2z = 0$ Eigenspace: $\dim(E_2) = 2$, spanned by
$$\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

**Diagonalizable:** $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Basis: $P = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

**(d)** This appears to be a $3 \times 4$ matrix (non-square).

**Not diagonalizable** - only square matrices can be diagonalized in the standard sense.

If intended as $3 \times 3$ (ignoring last column), one would need to compute eigenvalues/eigenvectors accordingly.

8. Consider the following bivariate distribution $p(x, y)$ of two discrete random variables $X$ and $Y$.

<div style="text-align:right">4</div>

|   | $y_1$ | 0.01 | 0.02 | 0.03 | 0.1 | 0.1 |
|---|-------|------|------|------|------|------|
| $Y$ | $y_2$ | 0.05 | 0.1 | 0.05 | 0.07 | 0.2 |
|   | $y_3$ | 0.1 | 0.05 | 0.03 | 0.05 | 0.04 |
|   |   | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |

$$X$$

(a) Compute the marginal distributions $p(x)$ and $p(y)$. 

<div style="text-align:right">2</div>

**Solution: Marginal distribution $p(x)$:** Sum over all values of $y$:

$$p(x_1) = 0.01 + 0.05 + 0.10 = 0.16$$
$$p(x_2) = 0.02 + 0.10 + 0.05 = 0.17$$
$$p(x_3) = 0.03 + 0.05 + 0.03 = 0.11$$
$$p(x_4) = 0.10 + 0.07 + 0.05 = 0.22$$
$$p(x_5) = 0.10 + 0.20 + 0.04 = 0.34$$

Therefore: $p(x) = (0.16, 0.17, 0.11, 0.22, 0.34)$

**Marginal distribution $p(y)$:** Sum over all values of $x$:

$$p(y_1) = 0.01 + 0.02 + 0.03 + 0.10 + 0.10 = 0.26$$
$$p(y_2) = 0.05 + 0.10 + 0.05 + 0.07 + 0.20 = 0.47$$
$$p(y_3) = 0.10 + 0.05 + 0.03 + 0.05 + 0.04 = 0.27$$

Therefore: $p(y) = (0.26, 0.47, 0.27)$

Verification: $\sum p(x) = \sum p(y) = 1.00$

(b) Compute the conditional distributions $p(x|Y = y_1)$ and $p(y|X = x_3)$. 

<div style="text-align:right">2</div>

**Solution: Conditional distribution $p(x|Y = y_1)$:**

Using $p(x|y) = \frac{p(x,y)}{p(y)}$, with $p(y_1) = 0.26$:

$$p(x_1|y_1) = \frac{0.01}{0.26} \approx 0.0385$$
$$p(x_2|y_1) = \frac{0.02}{0.26} \approx 0.0769$$
$$p(x_3|y_1) = \frac{0.03}{0.26} \approx 0.1154$$
$$p(x_4|y_1) = \frac{0.10}{0.26} \approx 0.3846$$
$$p(x_5|y_1) = \frac{0.10}{0.26} \approx 0.3846$$

Or exactly: $p(x|y_1) = \left(\frac{1}{26}, \frac{2}{26}, \frac{3}{26}, \frac{10}{26}, \frac{10}{26}\right)$

**Conditional distribution $p(y|X = x_3)$:**

Using $p(y|x) = \frac{p(x,y)}{p(x)}$, with $p(x_3) = 0.11$:

$$p(y_1|x_3) = \frac{0.03}{0.11} = \frac{3}{11} \approx 0.2727$$
$$p(y_2|x_3) = \frac{0.05}{0.11} = \frac{5}{11} \approx 0.4545$$
$$p(y_3|x_3) = \frac{0.03}{0.11} = \frac{3}{11} \approx 0.2727$$

Therefore: $p(y|x_3) = \left(\frac{3}{11}, \frac{5}{11}, \frac{3}{11}\right)$

9. Consider two random variables, $x$, $y$ with joint distribution $p(x, y)$. Show that $\boxed{3}$

$$\mathbb{E}_X[x] = \mathbb{E}_Y[\mathbb{E}_X[x|y]]$$

Here, $\mathbb{E}_X[x|y]$ denotes the expected value of $x$ under the conditional distribution $p(x|y)$.

**Solution: Discrete case:**

Starting with the right-hand side:

$$\mathbb{E}_Y[\mathbb{E}_X[x|y]] = \sum_y p(y)\mathbb{E}_X[x|y]$$

$$= \sum_y p(y) \sum_x x \cdot p(x|y)$$

$$= \sum_y \sum_x x \cdot p(x|y) \cdot p(y)$$

$$= \sum_y \sum_x x \cdot p(x,y) \qquad (\text{since } p(x,y) = p(x|y)p(y))$$

$$= \sum_x x \sum_y p(x,y)$$

$$= \sum_x x \cdot p(x)$$

$$= \mathbb{E}_X[x]$$

**Continuous case:**

$$\mathbb{E}_Y[\mathbb{E}_X[x|y]] = \int_{-\infty}^{\infty} \mathbb{E}_X[x|y] \cdot p(y)\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} x \cdot p(x|y)\,\mathrm{d}x \right) p(y)\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot p(x|y) \cdot p(y)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot p(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} x \left( \int_{-\infty}^{\infty} p(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} x \cdot p(x)\,\mathrm{d}x$$

$$= \mathbb{E}_X[x]$$

Therefore, $\mathbb{E}_X[x] = \mathbb{E}_Y[\mathbb{E}_X[x|y]]$.   $\square$

10. Last year we qualitatively described a number of Probability Distributions, this year we shall discover more results. $\boxed{7}$

    To begin, find the Probability Mass/Density Functions for the following distributions:

    1. Binomial

18

2. Bernoulli

3. Geometric

4. Poisson

5. Exponential

6. Gamma

7. Normal

---

**Solution: (a) Binomial:** $X \sim \text{Bin}(n, p)$, where $n \in \mathbb{N}$, $p \in [0, 1]$

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

**(b) Bernoulli:** $X \sim \text{Ber}(p)$, where $p \in [0, 1]$

$$P(X = k) = p^k (1 - p)^{1-k}, \quad k \in \{0, 1\}$$

Equivalently: $P(X = 1) = p$, $P(X = 0) = 1 - p$

**(c) Geometric:** $X \sim \text{Geo}(p)$, where $p \in (0, 1]$

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

(Number of trials until first success)

Alternative parameterisation (number of failures before first success):

$$P(X = k) = (1 - p)^k p, \quad k = 0, 1, 2, \ldots$$

**(d) Poisson:** $X \sim \text{Pois}(\lambda)$, where $\lambda > 0$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

**(e) Exponential:** $X \sim \text{Exp}(\lambda)$, where $\lambda > 0$

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

**(f) Gamma:** $X \sim \text{Gamma}(\alpha, \beta)$, where $\alpha, \beta > 0$

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt$

**(g) Normal (Gaussian):** $X \sim \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$, $\sigma > 0$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

11. (a) What is the formula for the Moment Generating Function? What about the $k$th (central) MGF? **23**

**Hint.**

$$\varphi_X(s) = \mathbb{E}\,(?) = \int ?$$

---

**Solution:**

$$\varphi_X(s) = \mathbb{E}\left(e^{sX}\right) \tag{5}$$

$$= \int_{-\infty}^{\infty} \mathrm{d}F_X(x) \tag{6}$$

$$\varphi_X^{(k)}(0) = \mathbb{E}(X^k) \tag{7}$$

---

(b) Hence, or otherwise derive the **Expectation**, **Variance** and **Moment Generating Function** for all 7 distributions in Q10. Show minimal working. **21**

---

**Solution: 1. Binomial** $X \sim \mathrm{Bin}(n, p)$:

- $\mathbb{E}[X] = np$

- $\mathrm{Var}(X) = np(1 - p)$

- $M_X(t) = (pe^t + (1 - p))^n$

*Derivation:* $X = \sum_{i=1}^{n} X_i$ where $X_i \sim \mathrm{Ber}(p)$ are independent. By linearity: $\mathbb{E}[X] = \sum \mathbb{E}[X_i] = np$, $\mathrm{Var}(X) = \sum \mathrm{Var}(X_i) = np(1 - p)$. For MGF: $M_X(t) = \prod_{i=1}^{n} M_{X_i}(t) = (pe^t + 1 - p)^n$.

**2. Bernoulli** $X \sim \mathrm{Ber}(p)$:

- $\mathbb{E}[X] = p$

- $\mathrm{Var}(X) = p(1 - p)$

- $M_X(t) = pe^t + (1 - p)$

*Derivation:* $\mathbb{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p$. $\mathbb{E}[X^2] = p$, so $\mathrm{Var}(X) = p - p^2 = p(1 - p)$. $M_X(t) = \mathbb{E}[e^{tX}] = (1 - p)e^0 + pe^t$.

**3. Geometric** $X \sim \mathrm{Geo}(p)$ (trials until first success):

- $\mathbb{E}[X] = \frac{1}{p}$

- $\mathrm{Var}(X) = \frac{1-p}{p^2}$

- $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$ for $t < -\ln(1 - p)$

---

*Derivation:*

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

$$M_X(t) = \sum_{k=1}^{\infty} e^{tk}(1-p)^{k-1}p = \frac{pe^t}{1 - (1-p)e^t}$$

**4. Poisson** $X \sim \text{Pois}(\lambda)$:

- $\mathbb{E}[X] = \lambda$

- $\text{Var}(X) = \lambda$

- $M_X(t) = e^{\lambda(e^t - 1)}$

*Derivation:*

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk}\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{\lambda(e^t - 1)}$$

For variance: $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda$, so $\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

**5. Exponential** $X \sim \text{Exp}(\lambda)$:

- $\mathbb{E}[X] = \frac{1}{\lambda}$

- $\text{Var}(X) = \frac{1}{\lambda^2}$

- $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$

*Derivation:*

$$\mathbb{E}[X] = \int_0^{\infty} x\lambda e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}$$

$$M_X(t) = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\mathrm{d}x = \lambda \int_0^{\infty} e^{-(\lambda - t)x}\mathrm{d}x = \frac{\lambda}{\lambda - t}$$

**6. Gamma** $X \sim \text{Gamma}(\alpha, \beta)$:

- $\mathbb{E}[X] = \frac{\alpha}{\beta}$

- $\text{Var}(X) = \frac{\alpha}{\beta^2}$

- $M_X(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$ for $t < \beta$

*Derivation:* Using $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$:

$$\mathbb{E}[X] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\beta x} \mathrm{d}x = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} = \frac{\alpha}{\beta}$$

$$M_X(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x} \mathrm{d}x = \left(\frac{\beta}{\beta-t}\right)^\alpha$$

7. **Normal** $X \sim \mathcal{N}(\mu, \sigma^2)$:

- $\mathbb{E}[X] = \mu$

- $\mathrm{Var}(X) = \sigma^2$

- $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

*Derivation:* For standard normal $Z \sim \mathcal{N}(0, 1)$:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{tz} e^{-z^2/2} \mathrm{d}z = e^{t^2/2}$$

$$\text{For } X = \mu + \sigma Z: \quad M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t + \sigma^2 t^2/2}$$

12. Shortest distance between an arbitrary point and hyperplane given by $wx + b = 0$    $\boxed{4}$

**Solution:** Let $\mathbf{x}_0 \in \mathbb{R}^n$ be an arbitrary point, and consider the hyperplane $H = \{\mathbf{x} : \mathbf{w}^T\mathbf{x} + b = 0\}$ where $\mathbf{w} \neq \mathbf{0}$.

The shortest distance from $\mathbf{x}_0$ to $H$ is the perpendicular distance.

**Derivation:**

1. The normal vector to the hyperplane is $\mathbf{w}$.

2. Any point on the hyperplane can be written as $\mathbf{x}_H$ where $\mathbf{w}^T\mathbf{x}_H + b = 0$.

3. The projection of $\mathbf{x}_0$ onto the hyperplane is found by moving in the direction of $-\mathbf{w}$ (toward the hyperplane):

$$\mathbf{x}_{\text{proj}} = \mathbf{x}_0 - \alpha\mathbf{w}$$

where $\alpha$ is chosen so that $\mathbf{x}_{\text{proj}} \in H$:

$$\mathbf{w}^T(\mathbf{x}_0 - \alpha\mathbf{w}) + b = 0 \implies \mathbf{w}^T\mathbf{x}_0 - \alpha\|\mathbf{w}\|^2 + b = 0$$

$$\alpha = \frac{\mathbf{w}^T\mathbf{x}_0 + b}{\|\mathbf{w}\|^2}$$

4. The distance is:

$$d(\mathbf{x}_0, H) = \|\mathbf{x}_0 - \mathbf{x}_{\text{proj}}\| = \|\alpha\mathbf{w}\| = |\alpha|\|\mathbf{w}\| = \frac{|\mathbf{w}^T\mathbf{x}_0 + b|}{\|\mathbf{w}\|}$$

**Answer:**

$$\boxed{d(\mathbf{x}_0, H) = \frac{|\mathbf{w}^T\mathbf{x}_0 + b|}{\|\mathbf{w}\|}}$$

In the case where $\|\mathbf{w}\| = 1$ (normalized), the distance simplifies to $|\mathbf{w}^T\mathbf{x}_0 + b|$.

13. Consider minimising the strictly convex quadratic function $\boxed{4}$

$$q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} + \mathbf{d}^T\mathbf{x} + c$$

on $\mathbb{R}^n$, where $n > 1$, $\mathbf{d}$ is a constant $n \times 1$ vector, $\mathbf{G}$ is a constant $n \times n$ symmetric positive definite matrix and $c$ is a scalar. Let $\mathbf{x}^* \in \mathbb{R}^n$ be the global minimiser for $q(\mathbf{x})$, $\mathbf{x}^{(1)} \in \mathbb{R}^n$ and $\mathbf{x}^{(1)} \neq \mathbf{x}^*$.

Consider applying **Newton's method** to $q(\mathbf{x})$ starting at $\mathbf{x}^{(1)}$

(a) Write down the Newton direction at $\mathbf{x}^{(1)}$ and show that it is a descent direction. $\boxed{2}$

---

**Solution:** For the quadratic $q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} + \mathbf{d}^T\mathbf{x} + c$:

**Gradient:** $\nabla q(\mathbf{x}) = \mathbf{G}\mathbf{x} + \mathbf{d}$

**Hessian:** $\nabla^2 q(\mathbf{x}) = \mathbf{G}$

The **Newton direction** at $\mathbf{x}^{(1)}$ is:

$$\mathbf{p}^{(1)} = -[\nabla^2 q(\mathbf{x}^{(1)})]^{-1}\nabla q(\mathbf{x}^{(1)}) = -\mathbf{G}^{-1}(\mathbf{G}\mathbf{x}^{(1)} + \mathbf{d})$$

**Showing it's a descent direction:**

A direction $\mathbf{p}$ is a descent direction if $\nabla q(\mathbf{x}^{(1)})^T\mathbf{p} < 0$.

$$\begin{aligned}
\nabla q(\mathbf{x}^{(1)})^T\mathbf{p}^{(1)} &= (\mathbf{G}\mathbf{x}^{(1)} + \mathbf{d})^T\left[-\mathbf{G}^{-1}(\mathbf{G}\mathbf{x}^{(1)} + \mathbf{d})\right] \\
&= -(\mathbf{G}\mathbf{x}^{(1)} + \mathbf{d})^T\mathbf{G}^{-1}(\mathbf{G}\mathbf{x}^{(1)} + \mathbf{d}) \\
&= -\|\nabla q(\mathbf{x}^{(1)})\|^2_{\mathbf{G}^{-1}}
\end{aligned}$$

Since $\mathbf{G}$ is positive definite, $\mathbf{G}^{-1}$ is also positive definite. Therefore, the quadratic form $\mathbf{v}^T\mathbf{G}^{-1}\mathbf{v} > 0$ for all $\mathbf{v} \neq \mathbf{0}$.

---

Since $\mathbf{x}^{(1)} \neq \mathbf{x}^*$, we have $\nabla q(\mathbf{x}^{(1)}) \neq \mathbf{0}$ (as $\mathbf{x}^*$ is the unique stationary point).

Thus: $\nabla q(\mathbf{x}^{(1)})^T \mathbf{p}^{(1)} < 0$, confirming $\mathbf{p}^{(1)}$ is a descent direction.

(b) How many iteration(s) will Newton's method take to reach the minimiser $\mathbf{x}^*$ of $q(\mathbf{x})$. Give reasons for your answer.  ☐2

**Solution: Exactly 1 iteration.**

Newton's method update is:

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \mathbf{p}^{(1)} = \mathbf{x}^{(1)} - \mathbf{G}^{-1}(\mathbf{G}\mathbf{x}^{(1)} + \mathbf{d})$$

The minimiser $\mathbf{x}^*$ satisfies the first-order optimality condition:

$$\nabla q(\mathbf{x}^*) = \mathbf{0} \implies \mathbf{G}\mathbf{x}^* + \mathbf{d} = \mathbf{0} \implies \mathbf{x}^* = -\mathbf{G}^{-1}\mathbf{d}$$

Substituting into the Newton update:

$$\begin{aligned}
\mathbf{x}^{(2)} &= \mathbf{x}^{(1)} - \mathbf{G}^{-1}\mathbf{G}\mathbf{x}^{(1)} - \mathbf{G}^{-1}\mathbf{d} \\
&= \mathbf{x}^{(1)} - \mathbf{x}^{(1)} - \mathbf{G}^{-1}\mathbf{d} \\
&= -\mathbf{G}^{-1}\mathbf{d} \\
&= \mathbf{x}^*
\end{aligned}$$

Therefore, Newton's method reaches the minimiser in **exactly 1 iteration** for any quadratic function, regardless of the starting point $\mathbf{x}^{(1)}$.

This is because the Hessian of a quadratic is constant, and Newton's method is *exact* for quadratic functions—it finds the point where the quadratic approximation (which is the function itself) is minimised.

14. Consider the following inequality constrained optimisation problem  ☐14

$$\begin{aligned}
\text{(IP)} \quad &\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimise}} \quad \mathbf{a}^T\mathbf{x} \\
&\text{subject to} \quad \mathbf{x}^T Q\mathbf{x} - 1 \leq 0,
\end{aligned} \qquad (8)$$

where $n \geq 1$, $Q$ is a symmetric and **positive definite** $n \times n$ matrix, $\mathbf{a} \in \mathbb{R}^n, \mathbf{a} \neq \mathbf{0}$ and the constraint function $\mathbf{c}(\mathbf{x}) = \mathbf{x}^T Q\mathbf{x} - 1$.

(a) Write down the gradient $\nabla c(\mathbf{x})$ and the Hessian $\nabla^2 c(\mathbf{x})$.  ☐2

**Solution:** For $c(\mathbf{x}) = \mathbf{x}^T Q\mathbf{x} - 1$:
**Gradient:**
$$\nabla c(\mathbf{x}) = 2Q\mathbf{x}$$

> (Since $Q$ is symmetric: $\nabla(\mathbf{x}^T Q \mathbf{x}) = (Q + Q^T)\mathbf{x} = 2Q\mathbf{x}$)
> **Hessian:**
> $$\nabla^2 c(\mathbf{x}) = 2Q$$

(b) Show that $c(\mathbf{x})$ is a convex function. $\boxed{2}$

> **Solution:** A twice-differentiable function $c : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its Hessian is positive semidefinite everywhere.
> From part (a): $\nabla^2 c(\mathbf{x}) = 2Q$
> Since $Q$ is given to be **positive definite**, we have for all $\mathbf{v} \neq \mathbf{0}$:
> $$\mathbf{v}^T Q \mathbf{v} > 0$$
> Therefore:
> $$\mathbf{v}^T (\nabla^2 c(\mathbf{x}))\mathbf{v} = 2\mathbf{v}^T Q \mathbf{v} > 0 \quad \forall \mathbf{v} \neq \mathbf{0}$$
> Thus $\nabla^2 c(\mathbf{x})$ is positive definite (hence positive semidefinite), which implies $c(\mathbf{x})$ is **strictly convex**.

(c) Show that the problem (IP) is a convex optimisation problem. $\boxed{2}$

> **Solution:** A problem is a **convex optimisation problem** if:
>
> 1. The objective function is convex
>
> 2. The feasible region is convex (i.e., all inequality constraints $g_i(\mathbf{x}) \leq 0$ have convex $g_i$)
>
> **1. Objective function:** $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x}$ is linear, hence convex.
> **2. Constraint:** $c(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x} - 1 \leq 0$
> From part (b), $c(\mathbf{x})$ is (strictly) convex.
> The feasible region is $\Omega = \{\mathbf{x} : \mathbf{x}^T Q \mathbf{x} \leq 1\}$, which is a convex set (it's a sublevel set of a convex function).
> Alternatively: for $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and $\lambda \in [0, 1]$:
> $$c(\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \leq \lambda c(\mathbf{x}_1) + (1 - \lambda)c(\mathbf{x}_2) \quad \text{(by convexity of } c\text{)}$$
> $$\leq \lambda \cdot 0 + (1 - \lambda) \cdot 0 = 0$$
> So $\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2 \in \Omega$, confirming $\Omega$ is convex.
> Therefore, (IP) is a **convex optimisation problem**.

(d) Show that $\mathbf{x}^* = \dfrac{-Q^{-1}\mathbf{a}}{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}}$ is a constrained stationary point for the problem (IP). $\boxed{2}$

**Solution:** A point $\mathbf{x}^*$ is a constrained stationary point if it satisfies the **KKT conditions**:

1. **Feasibility:** $c(\mathbf{x}^*) \leq 0$

2. **Stationarity:** $\nabla f(\mathbf{x}^*) + \mu \nabla c(\mathbf{x}^*) = \mathbf{0}$ for some $\mu$

3. **Complementarity:** $\mu c(\mathbf{x}^*) = 0$

4. **Dual feasibility:** $\mu \geq 0$

Let's verify:

**1. Feasibility:**

$$
\begin{aligned}
c(\mathbf{x}^*) &= (\mathbf{x}^*)^T Q \mathbf{x}^* - 1 \\
&= \frac{(-Q^{-1}\mathbf{a})^T Q(-Q^{-1}\mathbf{a})}{\mathbf{a}^T Q^{-1}\mathbf{a}} - 1 \\
&= \frac{\mathbf{a}^T Q^{-1} Q Q^{-1}\mathbf{a}}{\mathbf{a}^T Q^{-1}\mathbf{a}} - 1 \\
&= \frac{\mathbf{a}^T Q^{-1}\mathbf{a}}{\mathbf{a}^T Q^{-1}\mathbf{a}} - 1 = 0
\end{aligned}
$$

So $\mathbf{x}^*$ lies on the boundary of the constraint (constraint is active).

**2. Stationarity:** We need $\mathbf{a} + \mu \cdot 2Q\mathbf{x}^* = \mathbf{0}$ for some $\mu > 0$.

$$
\begin{aligned}
\mathbf{a} + 2\mu Q\mathbf{x}^* &= \mathbf{a} + 2\mu Q \cdot \frac{-Q^{-1}\mathbf{a}}{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}} \\
&= \mathbf{a} - \frac{2\mu \mathbf{a}}{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}}
\end{aligned}
$$

Setting this to $\mathbf{0}$:

$$
\mathbf{a}\left(1 - \frac{2\mu}{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}}\right) = \mathbf{0}
$$

Since $\mathbf{a} \neq \mathbf{0}$, we need:

$$
\mu = \frac{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}}{2} > 0
$$

**3 & 4:** With $\mu > 0$ and $c(\mathbf{x}^*) = 0$, complementarity and dual feasibility are satisfied.

Therefore, $\mathbf{x}^*$ is a constrained stationary point.

(e) Determine whether $\mathbf{x}^* = \frac{-Q^{-1}\mathbf{a}}{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}}$ is a local minimiser, global minimiser or $\boxed{2}$

neither for the problem (IP).

Solution: Answer: $\mathbf{x}^*$ is a global minimiser.

**Proof:**

Since (IP) is a convex optimisation problem (from part c) with:

- Convex objective function $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$

- Convex feasible region $\Omega$

Any local minimiser is also a **global minimiser** for convex problems.

We showed in part (d) that $\mathbf{x}^*$ satisfies the KKT conditions. For convex problems, the KKT conditions are **sufficient** for global optimality.

**Alternative geometric argument:**

The problem minimises a linear function over an ellipsoid. The minimum occurs where the level set $\{\mathbf{x} : \mathbf{a}^T\mathbf{x} = t\}$ (a hyperplane orthogonal to $\mathbf{a}$) is tangent to the ellipsoid $\{\mathbf{x} : \mathbf{x}^T Q\mathbf{x} \leq 1\}$.

This tangency occurs at:

$$\mathbf{x}^* = \frac{-Q^{-1}\mathbf{a}}{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}}$$

The direction $-Q^{-1}\mathbf{a}$ points from the origin toward decreasing values of $\mathbf{a}^T\mathbf{x}$, and we scale to hit the boundary of the ellipsoid.

**Verification:** The objective value at $\mathbf{x}^*$ is:

$$f(\mathbf{x}^*) = \mathbf{a}^T\mathbf{x}^* = -\frac{\mathbf{a}^T Q^{-1}\mathbf{a}}{\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}} = -\sqrt{\mathbf{a}^T Q^{-1}\mathbf{a}}$$

This is the **minimum** value over the feasible region.

Therefore, $\mathbf{x}^*$ is a **global minimiser**.

(f) Write down the Wolfe dual problem for (IP). ⬜2

Solution: The **Wolfe dual** for the problem

$$\min_{\mathbf{x}} \quad \mathbf{a}^T\mathbf{x}$$
$$\text{s.t.} \quad \mathbf{x}^T Q\mathbf{x} - 1 \leq 0$$

is constructed from the Lagrangian:

$$L(\mathbf{x}, \mu) = \mathbf{a}^T\mathbf{x} + \mu(\mathbf{x}^T Q\mathbf{x} - 1)$$

The Wolfe dual is:

$$(\text{WD}) \quad \max_{\mathbf{x}, \mu} \quad L(\mathbf{x}, \mu) = \mathbf{a}^T \mathbf{x} + \mu(\mathbf{x}^T Q \mathbf{x} - 1)$$

$$\text{s.t.} \quad \nabla_{\mathbf{x}} L(\mathbf{x}, \mu) = \mathbf{a} + 2\mu Q \mathbf{x} = \mathbf{0}$$

$$\mu \geq 0$$

Equivalently, eliminating $\mathbf{x}$ using the stationarity condition $\mathbf{x} = -\frac{1}{2\mu} Q^{-1} \mathbf{a}$:

$$(\text{WD}) \quad \max_{\mu \geq 0} \quad -\frac{\mathbf{a}^T Q^{-1} \mathbf{a}}{4\mu} - \mu$$

(g) What is the optimal objective value of the Wolfe dual problem in part f)? Give reasons for your answer. You do not need to solve the Wolfe dual problem. $\boxed{2}$

**Solution:** The optimal objective value of the Wolfe dual is $-\sqrt{\mathbf{a}^T Q^{-1} \mathbf{a}}$. For convex optimisation problems with differentiable objective and constraints, **strong duality** holds under Slater's condition (which is satisfied here since the constraint is strictly feasible at the origin if $0 < 1$, or at any interior point).

By strong duality:

$$\text{Primal optimal value} = \text{Dual optimal value}$$

From part (e), we found that the primal optimal value is:

$$f(\mathbf{x}^*) = \mathbf{a}^T \mathbf{x}^* = -\sqrt{\mathbf{a}^T Q^{-1} \mathbf{a}}$$

Therefore, the Wolfe dual optimal value is also:

$$\boxed{-\sqrt{\mathbf{a}^T Q^{-1} \mathbf{a}}}$$

**Verification:** At the optimal $\mu^* = \frac{\sqrt{\mathbf{a}^T Q^{-1} \mathbf{a}}}{2}$ from part (d), the dual objective is:

$$L(\mathbf{x}^*, \mu^*) = \mathbf{a}^T \mathbf{x}^* + \mu^* \cdot 0$$
$$= -\sqrt{\mathbf{a}^T Q^{-1} \mathbf{a}}$$

confirming strong duality with zero duality gap.

15. Consider the following equality constrained optimisation problem $\boxed{15}$

28

$$\text{(EP)} \quad \min_{\mathbf{x} \in \mathbb{R}^n} \quad \mathbf{a}^T \mathbf{x}$$

$$\text{s.t.} \quad \mathbf{x}^T \mathbf{x} - 1 = 0,$$

where $n \geq 1$, $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq \mathbf{0}$ and $\|\mathbf{a}\|_2 = \sqrt{\mathbf{a}^T \mathbf{a}}$. Let $\mathbf{z}^* = \frac{-\mathbf{a}}{\|\mathbf{a}\|_2}$.

where $n \geq 1$, $Q$ is a symmetric and **positive definite** $n \times n$ matrix, $\mathbf{a} \in \mathbb{R}^n, \mathbf{a} \neq \mathbf{0}$ and the constraint function $\mathbf{c}(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x} - 1$.

(a) Show that global minima of (EP) must exist. $\boxed{2}$

> **Solution:** We use the **Extreme Value Theorem**: A continuous function on a compact set attains its minimum.
>
> **1. Objective is continuous:** $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x}$ is linear, hence continuous.
>
> **2. Feasible region is compact:** The constraint set is $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{x} = 1\}$, which is the unit sphere $S^{n-1}$.
>
> The unit sphere is:
>
> - **Closed:** It's the preimage of the closed set $\{1\}$ under the continuous function $\mathbf{x} \mapsto \|\mathbf{x}\|^2$
>
> - **Bounded:** $\|\mathbf{x}\| = 1$ for all $\mathbf{x} \in \Omega$
>
> By the Heine-Borel theorem, a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded. Therefore $\Omega$ is compact.
>
> Since $f$ is continuous and $\Omega$ is compact and non-empty, by the Extreme Value Theorem, $f$ attains its minimum on $\Omega$.
>
> Therefore, **global minima of (EP) must exist**.

(b) Show that $\mathbf{z}^*$ is a regular feasible point for (EP). $\boxed{2}$

> **Solution:** A point is **regular** (satisfies LICQ - Linear Independence Constraint Qualification) if the gradients of the active constraints are linearly independent.
>
> For (EP), the constraint is $h(\mathbf{x}) = \mathbf{x}^T \mathbf{x} - 1 = 0$.
>
> **Gradient:**
> $$\nabla h(\mathbf{x}) = 2\mathbf{x}$$
>
> At $\mathbf{z}^* = \frac{-\mathbf{a}}{\|\mathbf{a}\|_2}$:
> $$\nabla h(\mathbf{z}^*) = 2\mathbf{z}^* = \frac{-2\mathbf{a}}{\|\mathbf{a}\|_2}$$
>
> Since $\mathbf{a} \neq \mathbf{0}$, we have $\nabla h(\mathbf{z}^*) \neq \mathbf{0}$.

A single non-zero vector is linearly independent, so the LICQ condition is satisfied.

**Verification of feasibility:**

$$(\mathbf{z}^*)^T \mathbf{z}^* = \frac{\mathbf{a}^T \mathbf{a}}{\|\mathbf{a}\|_2^2} = \frac{\|\mathbf{a}\|_2^2}{\|\mathbf{a}\|_2^2} = 1$$

Therefore, $\mathbf{z}^*$ is a **regular feasible point** for (EP).

(c) Show rigorously that the feasible region $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{x} - 1 = 0\}$ of (EP) $\boxed{2}$ is not a convex set.

**Solution: Proof by counterexample:**

Consider $n = 2$. Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$.

Both points are in $\Omega$:

$$\mathbf{x}_1^T \mathbf{x}_1 = 1, \quad \mathbf{x}_2^T \mathbf{x}_2 = 1$$

For a convex set, the line segment between any two points must lie entirely in the set. Consider the midpoint with $\lambda = \frac{1}{2}$:

$$\mathbf{x}_{\text{mid}} = \frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2 = \frac{1}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Check if $\mathbf{x}_{\text{mid}} \in \Omega$:

$$\mathbf{x}_{\text{mid}}^T \mathbf{x}_{\text{mid}} = 0 \neq 1$$

Therefore $\mathbf{x}_{\text{mid}} \notin \Omega$.

Since we found $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ but their convex combination $\frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2 \notin \Omega$, the set $\Omega$ is **not convex**.

**General argument:** For any $n \geq 1$, the unit sphere $S^{n-1}$ is not convex because it does not contain the line segments between antipodal points (or any pair of distinct points on the sphere).

(d) Verify that $\mathbf{z}^*$ is a constrained stationary point for (EP). $\boxed{3}$

**Solution:** For an equality-constrained problem, a point $\mathbf{z}^*$ is a constrained stationary point if there exists a Lagrange multiplier $\lambda$ such that:

**Lagrange condition:**

$$\nabla f(\mathbf{z}^*) + \lambda \nabla h(\mathbf{z}^*) = \mathbf{0}$$

where $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x}$ and $h(\mathbf{x}) = \mathbf{x}^T \mathbf{x} - 1$.

We have:

- $\nabla f(\mathbf{x}) = \mathbf{a}$

- $\nabla h(\mathbf{x}) = 2\mathbf{x}$

At $\mathbf{z}^* = \frac{-\mathbf{a}}{\|\mathbf{a}\|_2}$:

$$\nabla h(\mathbf{z}^*) = 2\mathbf{z}^* = \frac{-2\mathbf{a}}{\|\mathbf{a}\|_2}$$

The Lagrange condition becomes:

$$\mathbf{a} + \lambda \cdot \frac{-2\mathbf{a}}{\|\mathbf{a}\|_2} = \mathbf{0}$$

$$\mathbf{a}\left(1 - \frac{2\lambda}{\|\mathbf{a}\|_2}\right) = \mathbf{0}$$

Since $\mathbf{a} \neq \mathbf{0}$, we need:

$$1 - \frac{2\lambda}{\|\mathbf{a}\|_2} = 0 \implies \lambda = \frac{\|\mathbf{a}\|_2}{2}$$

With this value of $\lambda$, the Lagrange condition is satisfied.

Therefore, $\mathbf{z}^*$ is a **constrained stationary point** with Lagrange multiplier $\lambda^* = \frac{\|\mathbf{a}\|_2}{2}$.

(e) Using the second-order sufficient optimality conditions, show that $\mathbf{z}^*$ is a strict local minimiser for (EP). $\boxed{3}$

**Solution:** The **second-order sufficient condition** states: If $\mathbf{z}^*$ is a regular stationary point and

$$\mathbf{d}^T \nabla^2_{xx} \mathcal{L}(\mathbf{z}^*, \lambda^*)\mathbf{d} > 0$$

for all $\mathbf{d} \neq \mathbf{0}$ satisfying $\nabla h(\mathbf{z}^*)^T\mathbf{d} = 0$ (tangent to the constraint), then $\mathbf{z}^*$ is a strict local minimiser.

The Lagrangian is:

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{a}^T\mathbf{x} + \lambda(\mathbf{x}^T\mathbf{x} - 1)$$

The Hessian with respect to $\mathbf{x}$:

$$\nabla^2_{xx}\mathcal{L}(\mathbf{x}, \lambda) = 2\lambda I$$

At $\mathbf{z}^*$ with $\lambda^* = \frac{\|\mathbf{a}\|_2}{2} > 0$ (since $\mathbf{a} \neq \mathbf{0}$):

$$\nabla^2_{xx}\mathcal{L}(\mathbf{z}^*, \lambda^*) = 2 \cdot \frac{\|\mathbf{a}\|_2}{2}I = \|\mathbf{a}\|_2 I$$

The tangent space at $\mathbf{z}^*$ is:

$$T_{\mathbf{z}^*} = \{\mathbf{d} : \nabla h(\mathbf{z}^*)^T \mathbf{d} = 0\} = \{\mathbf{d} : 2(\mathbf{z}^*)^T \mathbf{d} = 0\} = \{\mathbf{d} : \mathbf{d} \perp \mathbf{z}^*\}$$

For any $\mathbf{d} \in T_{\mathbf{z}^*}$ with $\mathbf{d} \neq \mathbf{0}$:

$$\mathbf{d}^T \nabla_{xx}^2 \mathcal{L}(\mathbf{z}^*, \lambda^*) \mathbf{d} = \mathbf{d}^T (\|\mathbf{a}\|_2 I)\mathbf{d} = \|\mathbf{a}\|_2 \|\mathbf{d}\|^2 > 0$$

Since the second-order sufficient condition is satisfied, $\mathbf{z}^*$ is a **strict local minimiser** for (EP).

(f) Show that $\mathbf{z}^*$ is a global minimiser for (EP). $\boxed{3}$

**Solution:** We'll show that $f(\mathbf{z}^*) \leq f(\mathbf{x})$ for all feasible $\mathbf{x}$.
For any $\mathbf{x} \in \Omega$ (i.e., $\|\mathbf{x}\| = 1$), by the **Cauchy-Schwarz inequality**:

$$\mathbf{a}^T \mathbf{x} \geq -|\mathbf{a}^T \mathbf{x}| \geq -\|\mathbf{a}\|_2 \|\mathbf{x}\|_2 = -\|\mathbf{a}\|_2$$

Equality holds when $\mathbf{x}$ and $\mathbf{a}$ point in opposite directions:

$$\mathbf{x} = \frac{-\mathbf{a}}{\|\mathbf{a}\|_2} = \mathbf{z}^*$$

At $\mathbf{z}^*$:
$$f(\mathbf{z}^*) = \mathbf{a}^T \mathbf{z}^* = \mathbf{a}^T \cdot \frac{-\mathbf{a}}{\|\mathbf{a}\|_2} = -\frac{\|\mathbf{a}\|_2^2}{\|\mathbf{a}\|_2} = -\|\mathbf{a}\|_2$$

Therefore:
$$f(\mathbf{z}^*) = -\|\mathbf{a}\|_2 \leq \mathbf{a}^T \mathbf{x} = f(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega$$

This shows that $\mathbf{z}^*$ achieves the minimum value of the objective function over the feasible region.
Therefore, $\mathbf{z}^*$ is a **global minimiser** for (EP).
**Note:** The maximum is achieved at $-\mathbf{z}^* = \frac{\mathbf{a}}{\|\mathbf{a}\|_2}$ with value $\|\mathbf{a}\|_2$.

16. (a) Prove that $\boxed{0}$
**Proposition.**

$$|a| \leq b \iff -b \leq a \leq b$$

*Proof.* **Solution: Proof:**
($\Rightarrow$) Assume $|a| \leq b$.

By definition of absolute value:

$$|a| = \begin{cases} a, & \text{if } a \geq 0 \\ -a, & \text{if } a < 0 \end{cases}$$

**Case 1:** If $a \geq 0$, then $|a| = a$, so $a \leq b$. Also, since $a \geq 0$, we have $a \geq -b$ (as $-b \leq 0 \leq a$ when $b \geq 0$). Thus $-b \leq a \leq b$.

**Case 2:** If $a < 0$, then $|a| = -a$, so $-a \leq b$, which gives $a \geq -b$. Also, $a < 0 \leq b$ (assuming $b \geq 0$, which must hold since $|a| \geq 0$). Thus $-b \leq a < 0 \leq b$.

In both cases, $-b \leq a \leq b$.

($\Leftarrow$) Assume $-b \leq a \leq b$.

**Case 1:** If $a \geq 0$, then $|a| = a \leq b$ by assumption.

**Case 2:** If $a < 0$, then $|a| = -a$. From $-b \leq a$, we get $-a \leq b$, so $|a| = -a \leq b$.

In both cases, $|a| \leq b$. $\quad\square$

$\square$

(b) Prove that $\boxed{2}$

**Proposition.**

$$\forall a, b \in \mathbb{R}, \quad |a| \cdot |b| = |a \cdot b|$$

*Proof.* **Solution: Proof:**

We consider four cases based on the signs of $a$ and $b$:

**Case 1:** $a \geq 0$ and $b \geq 0$

Then $ab \geq 0$, so:
$$|a| \cdot |b| = a \cdot b = |ab|$$

**Case 2:** $a \geq 0$ and $b < 0$

Then $ab \leq 0$, so:
$$|a| \cdot |b| = a \cdot (-b) = -ab = |ab|$$

**Case 3:** $a < 0$ and $b \geq 0$

Then $ab \leq 0$, so:
$$|a| \cdot |b| = (-a) \cdot b = -ab = |ab|$$

**Case 4:** $a < 0$ and $b < 0$

Then $ab > 0$, so:

$$|a| \cdot |b| = (-a) \cdot (-b) = ab = |ab|$$

In all cases, $|a| \cdot |b| = |ab|$.  $\square$

**Alternative proof:** Note that $|a| = \sqrt{a^2}$ and $|b| = \sqrt{b^2}$. Then:

$$|a| \cdot |b| = \sqrt{a^2} \cdot \sqrt{b^2} = \sqrt{a^2 b^2} = \sqrt{(ab)^2} = |ab|$$

$\square$

(c) Complete the following $\boxed{2}$

**Proposition.**

Two real numbers, $a$ and $b$ are equal if and only if, for every $\epsilon > 0$, $|a - b| < \epsilon$.

*Proof.* **Solution: Proof:**

($\Rightarrow$) Assume $a = b$.

Then $|a - b| = |0| = 0 < \epsilon$ for all $\epsilon > 0$.

($\Leftarrow$) Assume that for every $\epsilon > 0$, $|a - b| < \epsilon$.

Suppose, for contradiction, that $a \neq b$. Then $|a - b| = d > 0$ for some $d \in \mathbb{R}$.

Choose $\epsilon_0 = \frac{d}{2} > 0$. By assumption, $|a - b| < \epsilon_0 = \frac{d}{2}$.

But this gives $d < \frac{d}{2}$, which implies $2d < d$, hence $d < 0$. This contradicts $d > 0$.

Therefore, $a = b$.  $\square$

**Interpretation:**

This states that two real numbers are equal if and only if the distance between them is arbitrarily small. This is a fundamental characterisation of equality in terms of the metric structure of $\mathbb{R}$.

17. Complete the following $\boxed{3}$

**Proposition.**

Let $\mathcal{A} \subseteq \mathbb{R}$ and $m \in \mathcal{A}$ be the minimum of $\mathcal{A}$, then $\inf \mathcal{A} = m$.

*Proof.* **Solution: Proof:**

Recall that $m$ is the **minimum** of $\mathcal{A}$ if:

1. $m \in \mathcal{A}$

2. $m \leq a$ for all $a \in \mathcal{A}$

Recall that $\inf \mathcal{A}$ (the **infimum** or greatest lower bound) satisfies:

1. $\inf \mathcal{A} \leq a$ for all $a \in \mathcal{A}$ (it is a lower bound)

2. If $\ell$ is any lower bound of $\mathcal{A}$, then $\ell \leq \inf \mathcal{A}$ (it is the greatest lower bound)

**We must show** $\inf \mathcal{A} = m$:

**Part 1:** Show $m$ is a lower bound of $\mathcal{A}$.

Since $m$ is the minimum, $m \leq a$ for all $a \in \mathcal{A}$. Thus $m$ is a lower bound.

**Part 2:** Show $m$ is the greatest lower bound.

Let $\ell$ be any lower bound of $\mathcal{A}$. Then $\ell \leq a$ for all $a \in \mathcal{A}$.

In particular, since $m \in \mathcal{A}$, we have $\ell \leq m$.

Since $m$ is a lower bound and every lower bound $\ell$ satisfies $\ell \leq m$, we conclude that $m = \inf \mathcal{A}$. $\quad\square$

**Note:** The minimum is the infimum when it exists and belongs to the set. However, not all sets have a minimum (e.g., $(0, 1)$ has no minimum), but every non-empty bounded-below set has an infimum (by the completeness of $\mathbb{R}$).

$\square$

18. (a) State and prove the Cauchy-Schwarz inequality. $\boxed{4}$

    **Proposition.**

---

**Solution: Cauchy-Schwarz Inequality:** For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$|\mathbf{x}^T \mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

Equivalently, for real numbers:

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq \sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}$$

---

*Proof.* **Solution: Proof:**

**Case 1:** If $\mathbf{y} = \mathbf{0}$, then both sides equal 0, so the inequality holds.

**Case 2:** Assume $\mathbf{y} \neq \mathbf{0}$. For any $t \in \mathbb{R}$, consider:

$$0 \leq \|\mathbf{x} - t\mathbf{y}\|^2 = (\mathbf{x} - t\mathbf{y})^T(\mathbf{x} - t\mathbf{y})$$

Expanding:

$$0 \leq \mathbf{x}^T\mathbf{x} - 2t\mathbf{x}^T\mathbf{y} + t^2\mathbf{y}^T\mathbf{y}$$
$$= \|\mathbf{x}\|^2 - 2t(\mathbf{x}^T\mathbf{y}) + t^2\|\mathbf{y}\|^2$$

This is a quadratic in $t$ that is non-negative for all $t$. Choose $t = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|^2}$ (the value that minimises the quadratic):

$$0 \leq \|\mathbf{x}\|^2 - 2\frac{(\mathbf{x}^T\mathbf{y})^2}{\|\mathbf{y}\|^2} + \frac{(\mathbf{x}^T\mathbf{y})^2}{\|\mathbf{y}\|^4} \cdot \|\mathbf{y}\|^2$$
$$= \|\mathbf{x}\|^2 - \frac{(\mathbf{x}^T\mathbf{y})^2}{\|\mathbf{y}\|^2}$$

Rearranging:

$$(\mathbf{x}^T\mathbf{y})^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2$$

Taking square roots:

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

**Equality condition:** Equality holds if and only if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent (i.e., $\mathbf{x} = c\mathbf{y}$ for some scalar $c$).

**Alternative proof via discriminant:** The quadratic $at^2 + bt + c$ with $a = \|\mathbf{y}\|^2 > 0$, $b = -2\mathbf{x}^T\mathbf{y}$, $c = \|\mathbf{x}\|^2$ is non-negative for all $t$. Thus its discriminant must be non-positive:

$$b^2 - 4ac \leq 0 \implies 4(\mathbf{x}^T\mathbf{y})^2 - 4\|\mathbf{x}\|^2\|\mathbf{y}\|^2 \leq 0$$

which gives the result.

$\square$

(b) Hence or otherwise, state and prove the Triangle Inequality. $\boxed{3}$
   **Proposition.**

**Solution: Triangle Inequality:** For all real numbers $a, b$:

$$|a + b| \leq |a| + |b|$$

More generally, for vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

*Proof.* **Solution:** We prove the vector version (the scalar case follows by setting $n = 1$).

$$\begin{aligned}
\|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y})^T(\mathbf{x} + \mathbf{y}) \\
&= \mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{y} \\
&= \|\mathbf{x}\|^2 + 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2
\end{aligned}$$

By the Cauchy-Schwarz inequality (from part a):

$$\mathbf{x}^T\mathbf{y} \leq |\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

Therefore:

$$\begin{aligned}
\|\mathbf{x} + \mathbf{y}\|^2 &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\
&= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2
\end{aligned}$$

Taking square roots (both sides are non-negative):

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

**Equality condition:** Equality holds if and only if $\mathbf{x}$ and $\mathbf{y}$ are non-negative scalar multiples of each other (i.e., point in the same direction).

$\square$

19. (a) Give the definitions of these elementary analysis facts:

**Definition:** Metric Space

**Solution:** A metric space is a pair $(X, d)$, where $X$ is a (non-empty) set and $d : X \times X \to [0, \infty)$ is a function, such that the following conditions hold for all $x, y, z \in X$:

 1. $d(x, y) = 0 \iff x = y$

 2. $d(x, y) = d(y, x)$

3. $d(x, y) + d(y, z) \geq d(x, z)$    (triangle inequality)

**Definition:** Epsilon Ball

**Solution:**
$$B_\epsilon(x_0) = \{x \in X : d(x, x_0) < \epsilon\}$$

**Definition:** Interior Point

**Solution:** $x_0 \in Y \subseteq X$ is an *interior point* if $B(x_0, \epsilon)$ lies completely within $Y$.

**Definition:** A subset $Y$ in $(X, d)$ is **open** if

**Solution:**
$$Y = \text{Int}(Y)$$

(b) Complete the following

6

**Proposition.**

Every $\epsilon$-ball in a metric space is open.

*Proof.* **Solution: Proof:**

Let $(X, d)$ be a metric space, $x_0 \in X$, and $\epsilon > 0$. We want to show that the $\epsilon$-ball
$$B_\epsilon(x_0) = \{x \in X : d(x, x_0) < \epsilon\}$$
is an open set.

To show $B_\epsilon(x_0)$ is open, we must show that every point in $B_\epsilon(x_0)$ is an interior point, i.e., for each $y \in B_\epsilon(x_0)$, there exists $\delta > 0$ such that $B_\delta(y) \subseteq B_\epsilon(x_0)$.

**Construction:**

Let $y \in B_\epsilon(x_0)$ be arbitrary. Then $d(y, x_0) < \epsilon$.

Define $\delta = \epsilon - d(y, x_0) > 0$.

We claim that $B_\delta(y) \subseteq B_\epsilon(x_0)$.

**Verification:**

Let $z \in B_\delta(y)$. Then $d(z, y) < \delta$.

38

By the triangle inequality:

$$
\begin{aligned}
d(z, x_0) &\leq d(z, y) + d(y, x_0) \\
&< \delta + d(y, x_0) \\
&= (\epsilon - d(y, x_0)) + d(y, x_0) \\
&= \epsilon
\end{aligned}
$$

Therefore $d(z, x_0) < \epsilon$, which means $z \in B_\epsilon(x_0)$.
Since $z$ was arbitrary, we have $B_\delta(y) \subseteq B_\epsilon(x_0)$.
This shows that $y$ is an interior point of $B_\epsilon(x_0)$.
Since $y \in B_\epsilon(x_0)$ was arbitrary, every point of $B_\epsilon(x_0)$ is an interior point.
Therefore, $B_\epsilon(x_0) = \mathrm{Int}(B_\epsilon(x_0))$, which means $B_\epsilon(x_0)$ is **open**. $\quad\square$

$\square$

20. Infinite-dimensional vector spaces $\qquad$ $\boxed{12}$

(a) What are the sets $c_{00}, c_0, \ell^p$ and $\ell^\infty$? Give examples of sequences in each. $\qquad$ $\boxed{8}$

**Definition:** $c_{00}$

**Solution:** The space of sequences with **only finitely many non-zero terms**.
$$c_{00} = \{(x_n)_{n=1}^\infty : \text{only finitely many } x_n \neq 0\}$$

**Example:** $c_{00}$

**Solution:** $(1, 2, 3, 0, 0, 0, \ldots)$, $(1, 0, 1, 0, 0, \ldots)$

**Definition:** $c_0$

**Solution:** The space of sequences that **converge to zero**.
$$c_0 = \left\{(x_n)_{n=1}^\infty : \lim_{n \to \infty} x_n = 0\right\}$$

**Example:** $c_0$

**Solution:** $\left(\frac{1}{n}\right)_{n=1}^\infty = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right)$, $\left(\frac{1}{2^n}\right)_{n=1}^\infty$

**Definition:** $\ell^p$

39

> **Solution:** The space of sequences whose $p$-th powers are **summable**.
>
> $$\ell^p = \left\{ (x_n)_{n=1}^\infty : \sum_{n=1}^\infty |x_n|^p < \infty \right\}$$

**Example:** $\ell^p$

> **Solution:**
>
> - $\left(\frac{1}{n}\right)_{n=1}^\infty$ (since $\sum \frac{1}{n^2} < \infty$), $\left(\frac{1}{2^n}\right)_{n=1}^\infty$
> - $\left(\frac{1}{n^2}\right)_{n=1}^\infty$ (since $\sum \frac{1}{n^2} < \infty$)

**Definition:** $\ell^\infty$

> **Solution:** The space of **bounded sequences**.
>
> $$\ell^\infty = \left\{ (x_n)_{n=1}^\infty : \sup_{n \in \mathbb{N}} |x_n| < \infty \right\}$$

**Example:** $\ell^\infty$

> **Solution:** $(1, -1, 1, -1, \ldots)$, $(\sin(n))_{n=1}^\infty$, any constant sequence $(c, c, c, \ldots)$

**Remark.**

**Relationships:**

$$c_{00} \subsetneq c_0 \subsetneq \ell^p \subsetneq \ell^\infty \quad \text{(for any } 1 \le p < \infty)$$

(b) Give the definition and at least 2 examples for each each:   $\boxed{4}$

**Definition:** Hilbert Space

> **Solution:** A complete inner product space.

**Example:** Hilbert Space

> **Solution:**
>
> - $\mathbb{R}^n$ with the standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$

- $\ell^2 = \{(x_n)_{n=1}^\infty : \sum_{n=1}^\infty |x_n|^2 < \infty\}$ with inner product $\langle x, y \rangle = \sum_{n=1}^\infty x_n y_n$

**Definition:** Banach Space

> **Solution:** A complete normed vector space

**Example:** Banach Space

> **Solution:** $(C[0,1], || \cdot ||_\infty)$ and $(\ell^p, || \cdot ||_p)$

21. (a) State **pointwise convergence** for a sequence of functions. $\boxed{8}$

> **Solution:** $\lim_{n\to\infty} f_n = f$ p.w (point-wise) $\iff \forall x \in X, \forall \epsilon > 0, \exists K = K(\epsilon, x) : \forall n \geq K \ |f_n(x) - f(x)| \ < \epsilon$
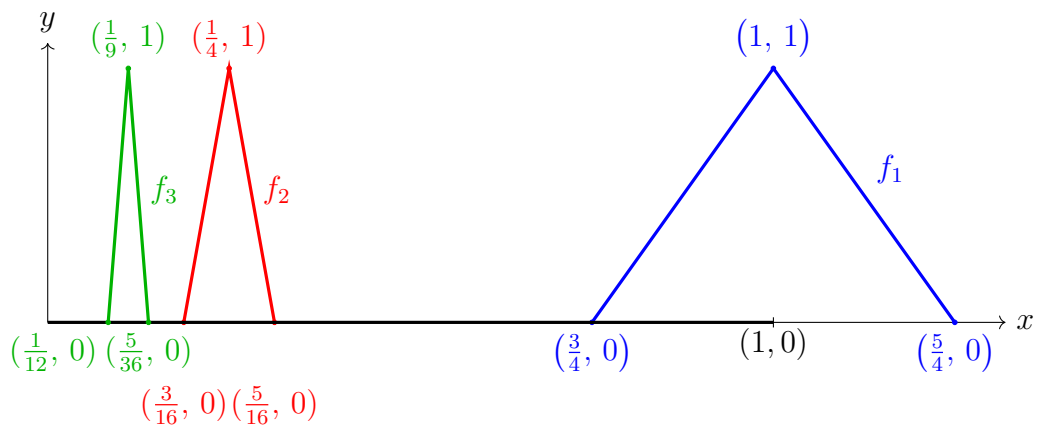
(b) State **uniform convergence** for a sequence of functions $\boxed{2}$

> **Solution:** $\lim_{n\to\infty} f_n = f$ uniformly $\iff \forall \epsilon > 0, \exists K = K(\epsilon) : \forall n \geq K \ |f_n(x) - f(x)| \ < \epsilon \quad \forall x \in X$

(c) Does the following function converge pointwise? What about uniformly? $\boxed{2}$
For each integer $k \geq 1$ define

$$f_k : [0,1] \longrightarrow \mathbb{R}, \qquad f_k(x) = \max\{0, 1 - 4k^2|x - \tfrac{1}{k^2}|\}.$$



> **Solution:**
>
> - **Pointwise convergence and failure of uniform convergence of**

41

$$S(x) := \sum_{k=1}^{\infty} \frac{f_k(x)}{k}, \qquad x \in [0, 1].$$

1. *Pointwise convergence.*

   Fix $x \in (0, 1]$. The inequality $\frac{3}{4k^2} < x < \frac{5}{4k^2}$ is equivalent to

   $$A(x) := \sqrt{\tfrac{3}{4x}} < k < B(x) := \sqrt{\tfrac{5}{4x}}.$$

   Whose length is

   $$L(x) := B(x) - A(x) = \frac{\sqrt{5} - \sqrt{3}}{2\sqrt{x}} < \frac{0.253}{\sqrt{x}} < \infty,$$

   so the interval holds at most $\lceil L(x) \rceil$ integers. Hence only finitely many $k$ satisfy $f_k(x) \neq 0$; the series $S(x)$ reduces to a finite sum and converges. For $x = 0$ every term is 0, so $S(0) = 0$. Thus $S$ converges for every $x \in [0, 1]$.

2. *Failure of uniform convergence.*

   Let $S_N(x) := \sum_{k=1}^{N} \frac{f_k(x)}{k}$. For $N \geq 10$ choose

   $$x_N := \frac{1}{N^2} \in [0, 1].$$

   **Claim:** For every $k \in [N + 1, N + \lfloor N/10 \rfloor]$ we have $f_k(x_N) \geq \frac{1}{2}$.

   **Proof:** For such $k$,

   $$\left| x_N - \tfrac{1}{k^2} \right| = \frac{|k^2 - N^2|}{k^2 N^2} = \frac{(k - N)(k + N)}{k^2 N^2} \leq \frac{(N/10)(11N/10)}{k^2 N^2} < \frac{11}{100\, k^2} < \frac{1}{8\, k^2} < \frac{1}{4k^2},$$

   so $x_N \in \mathrm{supp}(f_k)$ and $f_k(x_N) = 1 - 4k^2 |x_N - \tfrac{1}{k^2}| \geq \frac{1}{2}$.

   Therefore the tail $T_N(x) := \sum_{k>N} \frac{f_k(x)}{k}$ satisfies

   $$T_N(x_N) \geq \frac{1}{2} \sum_{k=N+1}^{N+\lfloor N/10 \rfloor} \frac{1}{k} \geq \frac{1}{2} \ln\left(1 + \tfrac{1}{10}\right) =: c > 0 \qquad (N \geq 10),$$

   using $\sum_{k=m}^{n} \frac{1}{k} \geq \ln \frac{n}{m}$, $\forall n, m \in \mathbb{N}$, $n \geq m \geq 1$:

   $$\|S - S_N\|_\infty = \sup_{x \in [0,1]} |S(x) - S_N(x)| \geq |T_N(x_N)| \geq c \quad \text{for all } N \geq 10,$$

   so $\|S - S_N\|_\infty \not\to 0$. The convergence of the series is therefore *not* uniform on $[0, 1]$.

(d) What does uniform convergence imply about a series? $\boxed{2}$

> **Solution:** $L^p$ convergence and **pointwise convergence**.

22. (a) Give the definition of a topological space $(X, \tau)$. $\boxed{2}$

> **Solution:**
> A **topological space** is a pair $(X, \tau)$, where $X$ is a set and $\tau \subseteq \mathcal{P}(X)$ satisfies
>
> 1. $\varnothing, X \in \tau$;
>
> 2. if $\{V_i\}_{i \in I} \subseteq \tau$ then $\bigcup_{i \in I} V_i \in \tau$;
>
> - "an arbitrary union of open sets is open"
>
> 1. if $V_1, V_2 \in \tau$ then $V_1 \cap V_2 \in \tau$.
>
> - "a finite intersection of open sets is open"

(b) Give the definitions of these Topological spaces: $\boxed{3}$
   **Definition:** Co-countable Topology

> **Solution:**
> $$\tau = \{Y \subseteq X : Y^c \text{ is countable }\} \cup \{\varnothing\}$$

   **Definition:** Co-finite Topology

> **Solution:**
> $$\tau = \{Y \subseteq X : Y^c \text{ is finite }\} \cup \{\varnothing\}$$

   **Definition:** Discrete Topology

> **Solution:**
> $$\tau = \{\varnothing, X\}$$

(c) Are limits unique in a topological space? What about in a metric space? $\boxed{2}$

> **Solution:** Not unique. Only in Hausdorff spaces, of which metric is a type of.

23. Let $(X_1, \ldots, X_n) \overset{\text{i.i.d}}{\sim} \mathcal{P}(\lambda)$. $\boxed{14}$

(a) Let

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

be an estimator of $\lambda$. Find the bias, standard error and Mean Squared Error of $\bar{X}_n$.

---

**Solution:** For $(X_1, \ldots, X_n) \overset{\text{i.i.d}}{\sim} \mathcal{P}(\lambda)$, we have $\mathbb{E}[X_i] = \lambda$ and $\text{Var}(X_i) = \lambda$.

**Bias:**

$$\text{Bias}(\bar{X}_n) = \mathbb{E}[\bar{X}_n] - \lambda$$

$$= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] - \lambda$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] - \lambda$$

$$= \frac{1}{n} \cdot n\lambda - \lambda$$

$$= 0$$

So $\bar{X}_n$ is **unbiased**.

**Standard Error:**

$$\text{SE}(\bar{X}_n) = \sqrt{\text{Var}(\bar{X}_n)}$$

$$= \sqrt{\text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)}$$

$$= \sqrt{\frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i)} \quad \text{(by independence)}$$

$$= \sqrt{\frac{1}{n^2} \cdot n\lambda}$$

$$= \sqrt{\frac{\lambda}{n}}$$

$$= \frac{\sqrt{\lambda}}{\sqrt{n}}$$

**Mean Squared Error:**

$$\text{MSE}(\bar{X}_n) = \text{Var}(\bar{X}_n) + [\text{Bias}(\bar{X}_n)]^2$$

$$= \frac{\lambda}{n} + 0^2$$

$$= \frac{\lambda}{n}$$

(b) Let
$$\widehat{T}_n = \frac{X_1 + 2X_2 + \ldots + nX_n}{1 + 2 + \cdots + n}.$$

Compute the bias and MSE of $\widehat{T}_n$ as an estimator of $\lambda$.

**Solution:** Note that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

**Bias:**

$$\mathbb{E}[\widehat{T}_n] = \frac{1}{\frac{n(n+1)}{2}} \sum_{i=1}^{n} i \cdot \mathbb{E}[X_i]$$

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} i \cdot \lambda$$

$$= \frac{2\lambda}{n(n+1)} \cdot \frac{n(n+1)}{2}$$

$$= \lambda$$

Therefore: $\text{Bias}(\widehat{T}_n) = 0$ (also unbiased).

**Variance:**

$$\text{Var}(\widehat{T}_n) = \text{Var}\left( \frac{2}{n(n+1)} \sum_{i=1}^{n} iX_i \right)$$

$$= \frac{4}{n^2(n+1)^2} \sum_{i=1}^{n} i^2 \text{Var}(X_i) \quad \text{(by independence)}$$

$$= \frac{4\lambda}{n^2(n+1)^2} \sum_{i=1}^{n} i^2$$

$$= \frac{4\lambda}{n^2(n+1)^2} \cdot \frac{n(n+1)(2n+1)}{6}$$

$$= \frac{4\lambda(2n+1)}{6n(n+1)}$$

$$= \frac{2\lambda(2n+1)}{3n(n+1)}$$

**MSE:**

$$\text{MSE}(\widehat{T}_n) = \text{Var}(\widehat{T}_n) + 0^2 = \frac{2\lambda(2n+1)}{3n(n+1)}$$

(c) Compare $\widehat{T}_n$ to $\bar{X}_n$ in terms of MSE. Which estimator is preferable? Intuitively, why is one better than the other?

**Solution: Comparison:**

$$\text{MSE}(\bar{X}_n) = \frac{\lambda}{n}$$
$$\text{MSE}(\widehat{T}_n) = \frac{2\lambda(2n+1)}{3n(n+1)}$$

To compare, compute the ratio:

$$\frac{\text{MSE}(\widehat{T}_n)}{\text{MSE}(\bar{X}_n)} = \frac{\frac{2\lambda(2n+1)}{3n(n+1)}}{\frac{\lambda}{n}}$$
$$= \frac{2(2n+1)}{3(n+1)}$$
$$= \frac{4n+2}{3n+3}$$
$$= \frac{4n+2}{3n+3}$$

For large $n$: $\frac{4n+2}{3n+3} \approx \frac{4n}{3n} = \frac{4}{3} > 1$

More precisely: $\text{MSE}(\widehat{T}_n) > \text{MSE}(\bar{X}_n)$ for all $n \geq 1$.

To verify, check if $\frac{2(2n+1)}{3(n+1)} > 1$:

$$2(2n+1) > 3(n+1) \iff 4n+2 > 3n+3 \iff n > 1$$

For $n = 1$: both equal $\lambda$ (both reduce to $X_1$). For $n \geq 2$: $\text{MSE}(\bar{X}_n) < \text{MSE}(\widehat{T}_n)$.

**Conclusion:** $\bar{X}_n$ is preferable (lower MSE).

**Intuition:** $\bar{X}_n$ weights all observations equally, making efficient use of all data. $\widehat{T}_n$ gives more weight to later observations, which introduces unnecessary variability without reducing bias (both are unbiased). The equal weighting of the sample mean is optimal for i.i.d. data.

(d) Find the moment estimator for $\lambda$. $\boxed{2}$

**Solution:** The **method of moments** equates population moments to sample moments.

For the Poisson distribution: $\mathbb{E}[X] = \lambda$ (first moment).

Sample first moment: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Setting them equal:
$$\lambda = \bar{X}_n$$

Therefore, the **moment estimator** is:

$$\boxed{\widehat{\lambda}_{\text{MM}} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

This is simply the sample mean.

(e) Find the maximum likelihood estimator for $\lambda$. $\boxed{3}$

**Solution:** The likelihood function for i.i.d. Poisson observations is:

$$L(\lambda) = \prod_{i=1}^{n} P(X_i = x_i)$$
$$= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$
$$= \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

The log-likelihood is:

$$\ell(\lambda) = \log L(\lambda)$$
$$= \sum_{i=1}^{n} x_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log(x_i!)$$
$$= \left(\sum_{i=1}^{n} x_i\right) \log \lambda - n\lambda - \text{const}$$

Taking the derivative with respect to $\lambda$:

$$\frac{\mathrm{d}\ell}{\mathrm{d}\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

Setting equal to zero:

$$\frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0 \implies \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Checking the second derivative:

$$\frac{\mathrm{d}^2\ell}{\mathrm{d}\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$$

This confirms a maximum.

Therefore, the **maximum likelihood estimator** is:

$$\widehat{\lambda}_{\text{MLE}} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

**Note:** For the Poisson distribution, the MLE equals the moment estimator.

(f) Find the Fisher information $\mathcal{I}(\lambda)$ [2]

**Solution:** The **Fisher information** for a single observation is:

$$\mathcal{I}(\lambda) = \mathbb{E}\left[\left(\frac{\partial}{\partial \lambda} \log f(X; \lambda)\right)^2\right]$$

or equivalently (when regularity conditions hold):

$$\mathcal{I}(\lambda) = -\mathbb{E}\left[\frac{\partial^2}{\partial \lambda^2} \log f(X; \lambda)\right]$$

For $X \sim \mathcal{P}(\lambda)$:
$$\log f(x; \lambda) = x \log \lambda - \lambda - \log(x!)$$

**First derivative:**
$$\frac{\partial}{\partial \lambda} \log f(x; \lambda) = \frac{x}{\lambda} - 1$$

**Second derivative:**
$$\frac{\partial^2}{\partial \lambda^2} \log f(x; \lambda) = -\frac{x}{\lambda^2}$$

Taking the negative expectation:

$$\mathcal{I}(\lambda) = -\mathbb{E}\left[-\frac{X}{\lambda^2}\right]$$
$$= \frac{\mathbb{E}[X]}{\lambda^2}$$
$$= \frac{\lambda}{\lambda^2}$$
$$= \frac{1}{\lambda}$$

**Verification using first formula:**

$$\mathcal{I}(\lambda) = \mathbb{E}\left[\left(\frac{X}{\lambda} - 1\right)^2\right]$$

$$= \mathbb{E}\left[\frac{X^2}{\lambda^2} - \frac{2X}{\lambda} + 1\right]$$

$$= \frac{\mathbb{E}[X^2]}{\lambda^2} - \frac{2\mathbb{E}[X]}{\lambda} + 1$$

$$= \frac{\lambda^2 + \lambda}{\lambda^2} - \frac{2\lambda}{\lambda} + 1$$

$$= 1 + \frac{1}{\lambda} - 2 + 1$$

$$= \frac{1}{\lambda}$$

Therefore:

$$\boxed{\mathcal{I}(\lambda) = \frac{1}{\lambda}}$$

For $n$ i.i.d. observations: $\mathcal{I}_n(\lambda) = n \cdot \mathcal{I}(\lambda) = \frac{n}{\lambda}$

24. What is the maximum straight-line distance in an N-dimensional unit hyper-cube? $\boxed{3}$

**Solution:**

**Answer:**

$$\boxed{\sqrt{N}}$$

**Derivation:**

An $N$-dimensional unit hyper-cube is the set $[0,1]^N = \{(x_1, x_2, \ldots, x_N) : 0 \leq x_i \leq 1 \text{ for all } i\}$.
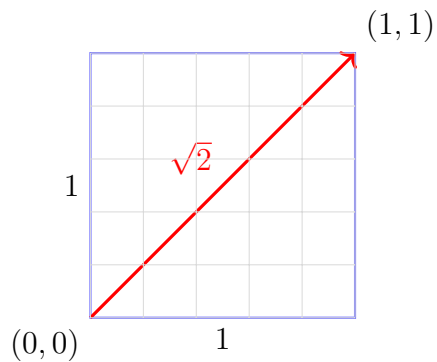
The maximum distance is achieved between opposite corners (vertices). Consider the vertices at the origin $\mathbf{0} = (0, 0, \ldots, 0)$ and the opposite corner $\mathbf{1} = (1, 1, \ldots, 1)$.
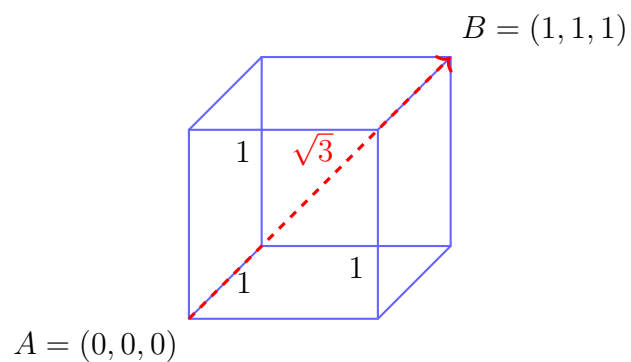
The Euclidean distance is:

$$d(\mathbf{0}, \mathbf{1}) = \sqrt{(1-0)^2 + (1-0)^2 + \cdots + (1-0)^2}$$

$$= \sqrt{\underbrace{1 + 1 + \cdots + 1}_{N \text{ times}}}$$

$$= \sqrt{N}$$

**Visualisation:**

**2D Unit Square**



**3D Unit Cube**



**General Pattern:**

- 1D (line segment): maximum distance = $\sqrt{1} = 1$

- 2D (square): maximum distance = $\sqrt{2} \approx 1.414$

- 3D (cube): maximum distance = $\sqrt{3} \approx 1.732$

- $N$D (hyper-cube): maximum distance = $\sqrt{N}$

This demonstrates the *curse of dimensionality*: as dimension increases, the diagonal of the unit hyper-cube grows without bound, even though each side length remains 1.

# § Grade Table

| Question: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Points: | 8 | 18 | 10 | 14 | 8 | 4 | 24 | 8 | 3 | 7 | 46 | 4 | 8 |
| Score: | | | | | | | | | | | | | |

| Question: | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Points: | 28 | 30 | 12 | 3 | 14 | 28 | 24 | 16 | 14 | 28 | 3 | | 362 |
| Score: | | | | | | | | | | | | | |