

Solution 1

Solution 1 (a)**Define** ℓ_1 The ℓ_1 or $\|x\|_1$ is defined as:

$$\ell_1 = \|x\|_1 = \sum_{i=1}^d |x_i|$$

Compute ℓ_1

$$\text{Let } x = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

$$\begin{aligned} \|x\|_1 &= \sum_{i=1}^3 |x_i| \\ &= |x_1| + |x_2| + |x_3| \\ &= |1| + |-2| + |3| \\ &= 1 + 2 + 3 \\ &= 6 \end{aligned}$$

$$\therefore \|x\|_1 = 6$$

Solution 1

Solution 1 (b)**Define ℓ_2** The ℓ_2 or $\|x\|_2$ is defined as:

$$\ell_2 = \|x\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$$

Compute ℓ_2

$$\text{Let } x = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

$$\begin{aligned} \|x\|_2 &= \sqrt{\sum_{i=1}^3 x_i^2} \\ &= \sqrt{x_1^2 + x_2^2 + x_3^2} \\ &= \sqrt{1^2 + (-2)^2 + 3^2} \\ &= \sqrt{1 + 4 + 9} \\ &= \sqrt{14} \end{aligned}$$

$$\therefore \|x\|_2 = \sqrt{14}$$

Solution 1

Solution 1 (c)**Define ℓ_∞** The ℓ_∞ or $\|x\|_\infty$ is defined as:

$$\ell_\infty = \|x\|_\infty = \max_i |x_i|$$

Compute ℓ_∞

$$\text{Let } x = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

$$\begin{aligned} \|x\|_\infty &= \max(\{|x_1|, |x_2|, |x_3|\}) \\ &= \max(\{|1|, |-2|, |3|\}) \\ &= \max(\{1, 2, 3\}) \\ &= 3 \end{aligned}$$

$$\therefore \|x\|_\infty = 3$$

Solution 2

Solution 2 (a)**Define ℓ_2 distance**

The ℓ_2 distance is defined as:

$$d(x, x')_{\ell_2} = \sqrt{\sum_{i=1}^n (x_i - x'_i)^2}$$

Compute ℓ_2 distance

$$\text{Let } x = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \text{ and } x' = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{aligned} d(x, x')_{\ell_2} &= \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + (x_3 - x'_3)^2 + (x_4 - x'_4)^2} \\ &= \sqrt{((-1) - 1)^2 + (1 - 1)^2 + ((-1) - 1)^2 + (1 - 1)^2} \\ &= \sqrt{(2)^2 + (0)^2 + (2)^2 + (0)^2} \\ &= \sqrt{4 + 0 + 4 + 0} \\ &= \sqrt{8} \end{aligned}$$

$$\therefore d(x, x')_{\ell_2} = \sqrt{8}$$

Solution 2

Solution 2 (b)**Define ℓ_1 distance**

The ℓ_1 distance is defined as:

$$d(x, x')_{\ell_1} = \sum_{i=1}^n |x_i - x'_i|$$

Compute ℓ_1 distance

$$\text{Let } x = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \text{ and } x' = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{aligned} d(x, x')_{\ell_1} &= |x_1 - x'_1| + |x_2 - x'_2| + |x_3 - x'_3| + |x_4 - x'_4| \\ &= |-1 - 1| + |1 - 1| + |-1 - 1| + |1 - 1| \\ &= |-2| + |0| + |-2| + |0| \\ &= 2 + 0 + 2 + 0 \\ &= 4 \end{aligned}$$

$$\therefore d(x, x')_{\ell_1} = 4$$

Solution 2

Solution 2 (c)**Define ℓ_∞ distance**

The ℓ_∞ distance is defined as:

$$d(x, x')_{\ell_\infty} = \max_i |x_i - x'_i|$$

Compute ℓ_∞ distance

$$\text{Let } x = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \text{ and } x' = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{aligned} d(x, x')_{\ell_\infty} &= \max\{|x_1 - x'_1|, |x_2 - x'_2|, |x_3 - x'_3|, |x_4 - x'_4|\} \\ &= \max\{|-1 - 1|, |1 - 1|, |-1 - 1|, |1 - 1|\} \\ &= \max\{|-2|, |0|, |-2|, |0|\} \\ &= \max\{2, 0, 2, 0\} \\ &= 2 \end{aligned}$$

$$\therefore d(x, x')_{\ell_\infty} = 2$$

Solution 3

Solution 3 (a)

(i) $\max(\ell_1)$ given $\|x\|_\infty = 1$

1. Find implications given $\|x\|_\infty = 1$

Given the constraint $\|x\|_\infty = \max_i |x_i| = 1$. It follows that $|x_i| \leq 1 : \forall i \in \{1, 2, \dots, d\}$ where d is the dimension of vector x . The ℓ_1 -norm, is the sum of the absolute values of the components(x_i) in the vector x or :

$$\ell_1 = \sum_{i=1}^d |x_i| = |x_1| + |x_2| + \dots + |x_d|$$

Each term $|x_i|$ in the sum must be maximized in order to maximize ℓ_1 -norm. The given constraint($\|x\|_\infty = 1$) allows each $|x_i|$ to be at most 1. Hence, the maximum is achieved when $|x_i| = 1$ or $x_i = \pm 1$.

2. Find value of ℓ_1 -norm

$$\begin{aligned} \|x\|_1 &= \sum_{i=1}^d |x_i| \\ &\leq \sum_{i=1}^d 1 \quad (\text{since } |x_i| \leq \|x\|_\infty = 1) \\ &\leq d \end{aligned}$$

\therefore The vector $x = \begin{bmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix}$ maximizes the norm, with a value of $\|x\|_1 = d$.

Solution 3

Solution 3 (a)

(ii): **Maximize** $\|x\|_2$ **given** $\|x\|_\infty = 1$

1. Apply part (i) results and solve for ℓ_2 -norm

Given the same constraint $\|x\|_\infty = 1$. From the results in part (i):

$$\begin{aligned}\|x\|_2 &= \sqrt{\sum_{i=1}^d x_i^2} \\ &= \sqrt{\sum_{i=1}^d |x_i|} \quad (\text{since all } x_i \in \{-1, 1\} \text{ for finding max}) \\ &\leq \sqrt{\ell_1} \quad (\text{definition of } \ell_1) \\ &\leq \sqrt{d}\end{aligned}$$

\therefore The vector $x = \begin{bmatrix} \pm 1 \\ \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix}$ maximizes the norm, with a value of $\|x\|_2 = \sqrt{d}$.

Solution 3

Solution 3 (b)

(i): Maximize $\|x\|_1$ given $\|x\|_2 = 1$

1. Find implications given $\|x\|_2 = 1$

Given the constraint $\|x\|_2 = 1$. By the Cauchy-Schwarz inequality, for vectors $u = (|x_1|, \dots, |x_d|)$ and $v = (1, \dots, 1)$, we have $(\sum |x_i|)^2 \leq (\sum |x_i|^2)(\sum 1^2)$. This is precisely $(\|x\|_1)^2 \leq (\|x\|_2^2)(d)$.

2. Find value for ℓ_1 -norm

$$\begin{aligned}
 (\|x\|_1)^2 &= \left(\sum_{i=1}^d |x_i| \right)^2 \\
 &\leq \left(\sum_{i=1}^d x_i^2 \right) \left(\sum_{i=1}^d 1^2 \right) \quad (\text{Cauchy-Schwarz Inequality}) \\
 &\leq (1)(d) \\
 &\quad (\text{Taking square root}) \\
 \|x\|_1 &\leq \sqrt{d}
 \end{aligned}$$

Each term $|x_i|$ in the sum must be maximized in order to maximize ℓ_1 -norm. The given constraint ($\|x\|_\infty = 1$) allows each $|x_i|$ to be at most 1. Hence, the maximum is achieved when $|x_i| = 1$ or $x_i = \pm 1$.

\therefore The vector $x = \begin{bmatrix} \pm 1/\sqrt{d} \\ \vdots \\ \pm 1/\sqrt{d} \end{bmatrix}$ maximizes the norm, with a value of $\|x\|_1 = \sqrt{d}$.

Solution 3

Solution 3 (b)

(ii): **Maximize** $\|x\|_\infty$ **given** $\|x\|_2 = 1$

1. Find implications given $\|x\|_2 = 1$

Given the constraint $\|x\|_2 = 1$. Let $|x_k|$ be the component with the maximum absolute value.

From the constraint, we can write $x_k^2 + \sum_{i=1}^{k-1} x_i^2 = 1$. Since the sum of squares is non-negative, it must be that $x_k^2 \leq 1$, which implies $|x_k| = \|x\|_\infty \leq 1$.

2. Find value for ℓ_∞ -norm

$$\begin{aligned}
 \|x\|_\infty^2 &= (\max_i |x_i|)^2 \\
 &= \max_i (x_i^2) \\
 &\leq \sum_{i=1}^d x_i^2 \quad (\text{as all terms are non-negative}) \\
 &\leq 1 \\
 &\quad (\text{Taking square root}) \\
 \|x\|_\infty &\leq 1
 \end{aligned}$$

This maximum value of 1 is achieved when one component has a magnitude of 1, which forces all other components to be zero to satisfy the constraint.

\therefore The vector $x = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ maximizes the norm (note: 1 can be in any dimension of x , it does not have to be x_0), with a value of $\|x\|_\infty = 1$.

Given the constraint $\|x\|_2 = 1$. Let $|x_k|$ be the component with the maximum absolute value. From the constraint, we can write $x_k^2 + \sum_{i=1}^{k-1} x_i^2 = 1$. Since the sum of squares is non-negative, it must be that $x_k^2 \leq 1$, which implies $|x_k| = \|x\|_\infty \leq 1$.

$$\begin{aligned}\|x\|_\infty^2 &= (\max_i |x_i|)^2 \\ &= \max_i (x_i^2) \\ &\leq \sum_{i=1}^d x_i^2 \quad (\text{as all terms are non-negative}) \\ &\leq 1 \\ \implies \|x\|_\infty &\leq 1\end{aligned}$$

This maximum value of 1 is achieved when one component has a magnitude of 1, which forces all other components to be zero to satisfy the constraint.

\therefore Any standard basis vector e_k (e.g., $x = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$) maximizes the norm, giving $\|x\|_\infty = 1$.

Solution 4

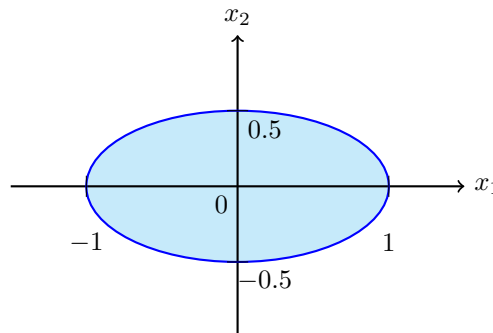
Find Unit Ball Equation

Given $\|x\|_w \leq 1$ dimension $d = 2$ and the weight vector $w = (w_1, w_2) = (1, 4)$. We can solve for the equation of this boundary by looking at the max value of the weighted-norm ($\|x\|_w = 1$).

$$\begin{aligned} \|x\|_w &= 1 \\ \sqrt{\sum_{i=1}^d w_i x_i^2} &= 1 \quad (\text{Definition of norm for } d = 2) \\ \sqrt{w_1 x_1^2 + w_2 x_2^2} &= 1 \quad (\text{Expanding summation}) \\ \sqrt{1 \cdot x_1^2 + 4 \cdot x_2^2} &= 1 \quad (\text{Substituting givens}) \\ x_1^2 + 4x_2^2 &= 1 \quad (\text{Squaring both sides}) \\ \frac{x_1^2}{1^2} + \frac{x_2^2}{(1/2)^2} &= 1 \quad (\text{Arrange into equation of ellipse}) \end{aligned}$$

Hence, the unit ball is an ellipse centered at the origin from the result above.

Sketch of the Unit Ball



\therefore The unit ball is an ellipse defined by $x_1^2 + 4x_2^2 \leq 1$, with a semi-major axis of length 1 and a semi-minor axis of length $1/2$.

Solution 5

Distance Table Discription

We are given a set of points $\mathcal{X} = \{A, B, C, D\}$ and a function $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ defined by the following distance table. We will determine if d is a metric by checking the required axioms.

$d(x, y)$	A	B	C	D
A	0	2	1	5
B	2	0	3	4
C	1	3	0	2
D	5	4	2	0

Axiom 1: Non-negativity and Identity of Indiscernibles

This axiom requires that $d(x, y) \geq 0$ for all $x, y \in \mathcal{X}$, and that $d(x, y) = 0$ if and only if $x = y$.

- **Non-negativity:** All entries in the table are non-negative, so this condition holds.
- **Identity:** The diagonal entries are all zero, so $d(x, x) = 0$. All off-diagonal entries are strictly positive, so $d(x, y) > 0$ when $x \neq y$.

Hence, d satisfies the first axiom (non-negativity & identity of indiscernibles).

Axiom 2: Symmetry

This axiom requires that $d(x, y) = d(y, x)$ for all $x, y \in \mathcal{X}$. Using the distance table above we can look up the following:

- $d(A, B) = 2 = d(B, A)$
- $d(A, C) = 1 = d(C, A)$
- $d(A, D) = 5 = d(D, A)$
- $d(B, C) = 3 = d(C, B)$
- $d(B, D) = 4 = d(D, B)$
- $d(C, D) = 2 = d(D, C)$

Hence, d satisfies the second axiom (symmetry).

Axiom 3: The Triangle Inequality

This axiom requires that for any three points $x, y, z \in \mathcal{X}$, the inequality $d(x, z) \leq d(x, y) + d(y, z)$ must hold.

Let $x = A$, $y = C$, and $z = D$.

$$\begin{aligned}
 d(x, z) &\leq d(x, y) + d(y, z) \\
 d(A, D) &\leq d(A, C) + d(C, D) \\
 5 &\leq 1 + 2 \\
 5 &\leq 3 \quad (\text{False})
 \end{aligned}$$

Hence, the triangle inequality does not hold, and d fails the third axiom.

\therefore The function d is not a metric because it fails the triangle inequality.

Solution 6

Solution 6 (a)**Find Largest Possible Value of $K(p, q)$**

Given $|\mathcal{X}| = 2$. This means that \mathcal{X} is a set containing two elements.

Let $\mathcal{X} = \{0, 1\}$, $p = (1, 0)$ and $q = (0, 1)$.

$p = (1, 0)$: $p(x_1) = 1$, $p(x_2) = 0$

$q = (0, 1)$: $q(x_1) = 0$, $q(x_2) = 1$

$$\begin{aligned} K(p, q) &= p(x_1) \log \frac{p(x_1)}{q(x_1)} + p(x_2) \log \frac{p(x_2)}{q(x_2)} \\ &= 1 \cdot \log \frac{1}{0} + 0 \cdot \log \frac{0}{1} \\ &= \infty + 0 \\ &= \infty \end{aligned}$$

\therefore The largest possible value of the KL divergence is ∞ .

Solution 6

Solution 6 (b)**Show $K(p, q)$ Non-Symmetry**

Let $p = (1/2, 1/2)$ and $q = (1, 0)$

1. Compute $K(p, q)$

$$\begin{aligned} K(p, q) &= p(x_1) \log \frac{p(x_1)}{q(x_1)} + p(x_2) \log \frac{p(x_2)}{q(x_2)} \\ &= \frac{1}{2} \log \frac{1/2}{1} + \frac{1}{2} \log \frac{1/2}{0} \\ &= \frac{1}{2} \log \frac{1}{2} + \infty \\ &= \infty \end{aligned}$$

2. Compute $K(q, p)$

$$\begin{aligned} K(q, p) &= q(x_1) \log \frac{q(x_1)}{p(x_1)} + q(x_2) \log \frac{q(x_2)}{p(x_2)} \\ &= 1 \cdot \log \frac{1}{1/2} + 0 \cdot \log \frac{0}{1/2} \\ &= \log(2) + 0 \\ &= \log(2) \end{aligned}$$

Hence, $\infty \neq \log(2)$

$\therefore K(p, q) \neq K(q, p)$, and the KL divergence is not symmetric.

Solution 7

Define Jaccard Similarity

Jaccard similarity between two sets A and B is the following:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Compute Jaccard Similarity

Let $A = \{1, 3, 5, 7, 9\}$ and $B = \{2, 3, 5, 7\}$

$$\begin{aligned} J(A, B) &= \frac{|A \cap B|}{|A \cup B|} \\ &= \frac{|\{1, 3, 5, 7, 9\} \cap \{2, 3, 5, 7\}|}{|\{1, 3, 5, 7, 9\} \cup \{2, 3, 5, 7\}|} \\ &= \frac{|\{3, 5, 7\}|}{|\{1, 2, 3, 5, 7, 9\}|} \\ &= \frac{3}{6} \\ &= \frac{1}{2} \end{aligned}$$

\therefore The Jaccard similarity $J(A, B)$ is $\frac{1}{2}$ or 50%.

Solution 8

Convert Sentences to Bigrams

Let:

x : "Napoleon was born in 1769" $\rightarrow B(x) = \{(\text{Napoleon, was}), (\text{was, born}), (\text{born, in}), (\text{in, 1769})\}$

x' : "Napoleon was born when" $\rightarrow B(x') = \{(\text{Napoleon, was}), (\text{was, born}), (\text{born, when})\}$

The Jaccard similarity for the bigram sets $B(x)$ and $B(x')$ is defined as:

$$J(B(x), B(x')) = \frac{|B(x) \cap B(x')|}{|B(x) \cup B(x')|}$$

Compute Jaccard Similarity

$$\begin{aligned} J(B(x), B(x')) &= \frac{|B(x) \cap B(x')|}{|B(x) \cup B(x')|} \\ &= \frac{|\{(\text{Napoleon, was}), (\text{was, born}), (\text{born, in}), (\text{in, 1769})\} \cap \{(\text{Napoleon, was}), (\text{was, born}), (\text{born, when})\}|}{|\{(\text{Napoleon, was}), (\text{was, born}), (\text{born, in}), (\text{in, 1769})\} \cup \{(\text{Napoleon, was}), (\text{was, born}), (\text{born, when})\}|} \\ &= \frac{|\{(\text{Napoleon, was}), (\text{was, born})\}|}{|\{(\text{Napoleon, was}), (\text{was, born}), (\text{born, when}), (\text{born, in}), (\text{in, 1769})\}|} \\ &= \frac{2}{5} \end{aligned}$$

\therefore The Jaccard similarity $J(B(x), B(x'))$ is $\frac{2}{5}$ or 40%.

Solution 9

Solution 9 (a)

Define Cosine Similarity

The cosine similarity between two vectors x and y is defined as:

$$\cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

Compute Cosine Similarity

Let $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $x' = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$$\begin{aligned} \cos(\theta) &= \frac{x \cdot x'}{\|x\| \|x'\|} \\ &= \frac{\sum_{i=1}^n x_i x'_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n x'^2_i}} \\ &= \frac{(1)(3) + (2)(2) + (3)(1)}{\sqrt{1^2 + 2^2 + 3^2} \sqrt{3^2 + 2^2 + 1^2}} \\ &= \frac{3 + 4 + 3}{\sqrt{1 + 4 + 9} \sqrt{9 + 4 + 1}} \\ &= \frac{10}{\sqrt{14} \sqrt{14}} \\ &= \frac{10}{14} \\ &= \frac{5}{7} \end{aligned}$$

\therefore The cosine similarity between x and x' is $\frac{5}{7}$.

Solution 9

Solution 9 (b)**(b): Characterization of Zero Similarity**

The cosine similarity between two non-zero vectors x and x' is zero if and only if the numerator of the formula is zero.

$$\cos(\theta) = 0 \iff x \cdot x' = 0$$

In a Euclidean vector space, a dot product of zero signifies that the two vectors are orthogonal (perpendicular) to each other. The angle θ between them is 90° or $\pi/2$ radians. Geometrically, they form a right angle.

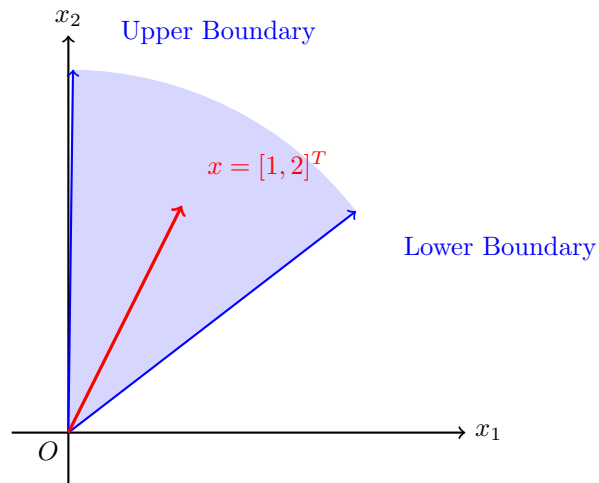
\therefore A cosine similarity of zero means the vectors are orthogonal to each other.

Solution 9

Solution 9 (c)

Sketch of the Region

The shaded region below represents the cone containing all vectors y satisfying $\cos(\theta) \geq 0.9$.



Solution 10

Solution 10 (a)**Compute Mean**

The mean (Expected Value) $E[X]$ is calculated as the sum of each outcome weighted by its probability:

$$E[X] = \sum_i x_i P(X = x_i)$$

$$\begin{aligned} E[X] &= (1) \left(\frac{1}{3}\right) + (2) \left(\frac{1}{3}\right) + (3) \left(\frac{1}{12}\right) + (4) \left(\frac{1}{12}\right) + (5) \left(\frac{1}{12}\right) + (6) \left(\frac{1}{12}\right) \\ &= \frac{4}{12} + \frac{8}{12} + \frac{3}{12} + \frac{4}{12} + \frac{5}{12} + \frac{6}{12} \\ &= \frac{30}{12} = \frac{5}{2} = 2.5 \end{aligned}$$

Compute Median

The median is the value x_m for which the cumulative probability $P(X \leq x_m)$ first equals or exceeds 0.5.

$$\begin{aligned} P(X \leq 1) &= \frac{1}{3} \\ P(X \leq 2) &= \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \end{aligned}$$

Hence, $P(X \leq 2)$ is the first cumulative probability to exceed 0.5, and the median is 2.

\therefore The mean is 2.5 and the median is 2.

Solution 10

Compute Variance

Variance is defined as:

$$\text{Var}(X) = E[X^2] - (E[X])^2, \text{ where } E[X^2] = \sum_i x_i^2 P(X = x_i)$$

$$\begin{aligned} E[X^2] &= (1^2) \left(\frac{1}{3}\right) + (2^2) \left(\frac{1}{3}\right) + (3^2) \left(\frac{1}{12}\right) + (4^2) \left(\frac{1}{12}\right) + (5^2) \left(\frac{1}{12}\right) + (6^2) \left(\frac{1}{12}\right) \\ &= \frac{4}{12} + \frac{16}{12} + \frac{9}{12} + \frac{16}{12} + \frac{25}{12} + \frac{36}{12} \\ &= \frac{106}{12} = \frac{53}{6} \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 = \frac{53}{6} - \left(\frac{5}{2}\right)^2 \\ &= \frac{53}{6} - \frac{25}{4} = \frac{106}{12} - \frac{75}{12} = \frac{31}{12} \end{aligned}$$

Compute Standard Deviation

The standard deviation is the square root of the variance:

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

It follows that, $\text{SD}(X) = \sqrt{\frac{31}{12}}$

\therefore The theoretical variance is $\text{Var}(X) = \frac{31}{12}$ and the standard deviation is $\text{SD}(X) = \sqrt{\frac{31}{12}}$.

Solution 10

Solution 10 (c)

Given the set of $N = 10$ observations $D = \{2, 5, 1, 4, 2, 2, 5, 6, 1, 2\}$, the empirical probability is defined as:

$$\hat{P}(X = x) = \frac{|x|}{N} \text{ where } |x| \text{ is frequency of event } x \text{ in set } D$$

The counts for each outcome are:

$$\text{Count}(1) = 2$$

$$\text{Count}(2) = 4$$

$$\text{Count}(3) = 0$$

$$\text{Count}(4) = 1$$

$$\text{Count}(5) = 2$$

$$\text{Count}(6) = 1$$

Hence the probabilities for each event are:

$$\hat{P}(1) = 0.2, \hat{P}(2) = 0.4, \hat{P}(3) = 0.0, \hat{P}(4) = 0.1, \hat{P}(5) = 0.2, \hat{P}(6) = 0.1$$

\therefore The empirical distribution is the following:

$$P(X = x) = \begin{cases} 0.2 & \text{if } x = 1 \\ 0.4 & \text{if } x = 2 \\ 0.0 & \text{if } x = 3 \\ 0.1 & \text{if } x = 4 \\ 0.2 & \text{if } x = 5 \\ 0.1 & \text{if } x = 6 \end{cases}$$

Solution 10

Solution 10 (d)

The empirical mean is $\bar{x} = \frac{1}{N} \sum x_i$. The median is the middle value of the sorted data. The empirical variance is $\text{Var} = \frac{1}{N} \sum (x_i - \bar{x})^2$, and the standard deviation is $SD = \sqrt{\text{Var}}$.

Compute Mean

$$\bar{x} = \frac{1}{10}(2 + 5 + 1 + 4 + 2 + 2 + 5 + 6 + 1 + 2) = \frac{30}{10} = 3.0$$

Compute Median

Sorted data: $\{1, 1, 2, 2, \mathbf{2}, \mathbf{2}, 4, 5, 5, 6\}$. The median is the average of the 5th and 6th values: $\frac{2+2}{2} = 2$.

Compute Variance and Standard Deviation

$$\begin{aligned} \text{Var} &= \frac{1}{10} \sum_{i=1}^{10} (x_i - 3.0)^2 \\ &= \frac{1}{10} [(1-3)^2 \times 2 + (2-3)^2 \times 4 + (4-3)^2 \times 1 + (5-3)^2 \times 2 + (6-3)^2 \times 1] \\ &= \frac{1}{10} [(-2)^2 \times 2 + (-1)^2 \times 4 + (1)^2 \times 1 + (2)^2 \times 2 + (3)^2 \times 1] \\ &= \frac{1}{10} [4 \times 2 + 1 \times 4 + 1 \times 1 + 4 \times 2 + 9 \times 1] \\ &= \frac{1}{10} [8 + 4 + 1 + 8 + 9] = \frac{30}{10} \\ &= 3 \\ SD &= \sqrt{3} \end{aligned}$$

\therefore The empirical mean is $\bar{x} = 3.0$, median is 2, variance is $\text{Var} = 3$, and standard deviation is $SD = \sqrt{3}$.

Graduate Level Explanation: Theoretical vs. Empirical

The theoretical distribution describes the true, underlying probability model of the random variable X . Its moments, such as the mean $\mu = E[X]$ and variance $\sigma^2 = \text{Var}(X)$, are fixed population parameters derived from this model. The empirical distribution, conversely, is constructed from a finite sample of observations. Its statistics, like the sample mean \bar{x} and sample variance s^2 , are estimates of the true parameters. The **Law of Large Numbers (LLN)** states that as the sample size N approaches infinity, the sample mean \bar{x} converges in probability to the theoretical mean μ . Similarly, other empirical moments converge to their theoretical counterparts. The discrepancy between our calculated theoretical values ($\mu = 2.5, \sigma^2 \approx 2.58$) and empirical values ($\bar{x} = 3.0, s^2 = 3.0$) is expected due to random sampling variation in our small sample ($N = 10$).

Explanation for a 5-Year-Old

Imagine you have a magic cookie jar. The **plan** (the theoretical part) says that for every 12 cookies you pull out, you *should* get 4 chocolate chip, 4 oatmeal, 1 sugar, 1 peanut butter, 1 ginger, and 1 snickerdoodle. The plan's "average cookie" is a mix between a chocolate chip and an oatmeal cookie.

But then, you actually pull out just 10 cookies. This is your **handful** (the empirical part). In your handful, you got a lot of oatmeal cookies and no sugar cookies at all! What happened in your one small handful is a little different from the big plan for the whole jar. If you kept pulling out cookies all day (thousands of them!), your handful would start to look a lot more like the original plan.

Solution 11

Solution 11 (a)

Justification: The distribution of human height for a given population (e.g., adult males) is famously well-approximated by a symmetric, bell-shaped curve (a Normal or Gaussian distribution). In a perfectly symmetric distribution, the mean, median, and mode coincide. Therefore, the center of mass (mean) and the geometric center (median) are expected to be nearly identical.

\therefore We expect $\text{Mean}(H) \approx \text{Median}(H)$. No significant difference.

Solution 11

Solution 11 (b)

Justification: The distribution of housing costs is almost always characterized by a strong right-skew. Most houses fall within a certain price range, there is a long tail of extremely expensive properties. These high-value outliers pull the mean significantly upward, while the median remains a more robust measure of the "typical" house price, unaffected by these extreme values.

\therefore We expect $\text{Mean}(C) > \text{Median}(C)$. A significant difference.

Solution 11

Solution 11 (c)

Justification: The distribution of GPAs is often left-skewed. This is due to a "ceiling effect," where a large number of students achieve high grades clustered near the maximum possible GPA (≈ 4.0), while fewer students have very low GPAs. This clustering at the high end pulls the median towards the right, while the lower-end scores pull the mean to the left.

\therefore We expect $\text{Mean}(G) < \text{Median}(G)$, but the difference may not be as significant as for house cost or salary.

Solution 11

Solution 11 (d)

Justification: Similar to housing costs, salary data exhibits a pronounced right-skew. The majority of people earn modest. However, a small number of individuals (CEOs, top athletes, etc.) have high incomes. These outliers have strong influence on the mean, that pulls it far to the right of the median. The median salary is therefore a much more accurate representation of a typical worker's salary.

\therefore We expect $\text{Mean}(S) > \text{Median}(S)$. A significant difference.

Solution 12

Define relationships

We know the following relationships are true:

$$\text{Var}[Z] = E[Z^2] - (E[Z])^2$$

$$\sigma^2 = \text{Var}[Z]$$

$$E[Z] = \mu$$

Rearranging this formula allows us to solve for $E[Z^2]$:

$$E[Z^2] = \text{Var}[Z] + (E[Z])^2 = \sigma^2 + \mu^2$$

Compute $E[Z^2]$

Given: $\mu = E[Z] = -1$ and the standard deviation $SD[Z] = \sigma = 2$

$$\begin{aligned} E[Z^2] &= \sigma^2 + \mu^2 \\ &= (2)^2 + (-1)^2 \\ &= 4 + 1 \\ &= 5 \end{aligned}$$

$\therefore E[Z^2]$ is 5.

Solution 13

Solution 13 (a)

Define X and Y Independence

Two discrete random variables X and Y are independent if and only if $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all possible pairs (x, y) .

Compute Marginal Probability Mass Functions

We sum rows and columns of the joint PMF table to compute $P(X = x)$ and $P(Y = y)$.

$$P(X = 1) = 0.10 + 0.20 + 0.05 = 0.35$$

$$P(X = 2) = 0.10 + 0.15 + 0.05 = 0.30$$

$$P(X = 3) = 0.10 + 0.15 + 0.10 = 0.35$$

$$P(Y = 1) = 0.10 + 0.10 + 0.10 = 0.30$$

$$P(Y = 2) = 0.20 + 0.15 + 0.15 = 0.50$$

$$P(Y = 3) = 0.05 + 0.05 + 0.10 = 0.20$$

Test Independence Condition

The test below is the independence condition for the pair $(X = 1, Y = 1)$:

$$\begin{aligned} P(X = 1, Y = 1) &\stackrel{?}{=} P(X = 1)P(Y = 1) \\ 0.10 &\stackrel{?}{=} (0.35)(0.30) \\ 0.10 &\neq 0.105 \end{aligned}$$

\therefore Since the joint probability $P(X = 1, Y = 1)$ does not equal the product of the marginal probabilities $P(X = 1)P(Y = 1)$, the random variables X and Y are NOT independent.

Solution 13

Solution 13 (b)

Define Covariance and Correlation of X and Y

The covariance is as follows:

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

The correlation is as follows

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

Compute Expectations X , Y and XY

$$E[X] = \sum_x xP(X=x) = 1(0.35) + 2(0.30) + 3(0.35) = 0.35 + 0.60 + 1.05 = 2.0$$

$$E[Y] = \sum_y yP(Y=y) = 1(0.30) + 2(0.50) + 3(0.20) = 0.30 + 1.00 + 0.60 = 1.9$$

$$\begin{aligned} E[XY] &= \sum_{x,y} xyP(X=x, Y=y) \\ &= (1)(1)(0.10) + (1)(2)(0.20) + (1)(3)(0.05) \\ &\quad + (2)(1)(0.10) + (2)(2)(0.15) + (2)(3)(0.05) \\ &\quad + (3)(1)(0.10) + (3)(2)(0.15) + (3)(3)(0.10) \\ &= 0.10 + 0.40 + 0.15 + 0.20 + 0.60 + 0.30 + 0.30 + 0.90 + 0.90 \\ &= 3.85 \end{aligned}$$

Compute Covariance

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 3.85 - (2.0)(1.9) = 3.85 - 3.80 = 0.05$$

Compute Variances and Standard Deviations

$$E[X^2] = 1^2(0.35) + 2^2(0.30) + 3^2(0.35) = 0.35 + 1.20 + 3.15 = 4.7$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 = 4.7 - 2.0^2 = 0.7$$

$$\sigma_X = \sqrt{0.7}$$

$$E[Y^2] = 1^2(0.30) + 2^2(0.50) + 3^2(0.20) = 0.30 + 2.00 + 1.80 = 4.1$$

$$\text{Var}[Y] = E[Y^2] - (E[Y])^2 = 4.1 - 1.9^2 = 4.1 - 3.61 = 0.49$$

$$\sigma_Y = \sqrt{0.49}$$

Compute Correlation Coefficient

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} = \frac{0.05}{\sqrt{0.7} \cdot \sqrt{0.49}} \approx 0.0854$$

\therefore The covariance is $\text{Cov}[X, Y] = 0.05$, and the correlation coefficient is $\rho_{X,Y} \approx 0.0854$. This indicates a very weak positive linear relationship between X and Y .

Solution 14

Solution 14 (a)

The expectation of Y , $E[Y]$, via the linearity of expectation is defined as:

$$E[Y] = E[aX + b] = aE[X] + b$$

The covariance between X and Y is defined as:

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$$

Now, substitute the expression for $E[Y]$ into the expression for covariance:

$$\begin{aligned}\text{Cov}[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[(X - E[X])((aX + b) - (aE[X] + b))] \\ &= E[(X - E[X])(aX + b - aE[X] - b)] \\ &= E[(X - E[X])(aX - aE[X])] \\ &= E[(X - E[X])a(X - E[X])] \\ &= a \cdot E[(X - E[X])^2] \\ &= a \cdot \text{Var}[X]\end{aligned}$$

\therefore The covariance between X and Y , in terms of X is the following:

$$\text{Cov}[X, Y] = a \cdot \text{Var}[X]$$

Solution 14

Solution 14 (b)

Part (b): Deriving the Correlation $\rho_{X,Y}$

The correlation coefficient $\rho_{X,Y}$ is defined as:

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\text{SD}[X]\text{SD}[Y]}$$

The variance of Y is the following:

$$\text{Var}[Y] = \text{Var}[aX + b] = a^2\text{Var}[X]$$

First, we need find the standard deviation of Y :

$$\text{SD}[Y] = \sqrt{\text{Var}[Y]} = \sqrt{a^2\text{Var}[X]} = |a|\sqrt{\text{Var}[X]} = |a| \cdot \text{SD}[X]$$

Now, substitute the known quantities into the correlation formula:

$$\begin{aligned} \rho_{X,Y} &= \frac{\text{Cov}[X, Y]}{\text{SD}[X]\text{SD}[Y]} \\ &= \frac{a \cdot \text{Var}[X]}{\text{SD}[X] \cdot (|a| \cdot \text{SD}[X])} \\ &= \frac{a \cdot (\text{SD}[X])^2}{|a| \cdot (\text{SD}[X])^2} \\ &= \frac{a}{|a|} \end{aligned}$$

\therefore The correlation between X and Y is $\rho_{X,Y} = \frac{a}{|a|}$, which is 1 if $a > 0$, -1 if $a < 0$, and undefined if $a = 0$.

Solution 15

Two random variables X and Y are uncorrelated if their covariance is zero. The covariance is defined as:

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

We first calculate the expected value of X , $E[X]$.

$$\begin{aligned} E[X] &= \sum_{x \in \Omega_X} x \cdot P(X = x) \\ &= (-1) \cdot P(X = -1) + (0) \cdot P(X = 0) + (1) \cdot P(X = 1) \\ &= (-1) \cdot \frac{1}{3} + (0) \cdot \frac{1}{3} + (1) \cdot \frac{1}{3} \\ &= -\frac{1}{3} + 0 + \frac{1}{3} \\ &= 0 \end{aligned}$$

Since $E[X] = 0$, the covariance formula simplifies significantly:

$$\text{Cov}[X, Y] = E[XY] - (0) \cdot E[Y] = E[XY]$$

Hence, for X and Y to be uncorrelated, we only need to find a function f such that:

$$E[Xf(X)] = 0$$

Next we expand the condition $E[Xf(X)] = 0$:

$$E[Xf(X)] = \sum_{x \in \Omega_X} xf(x)P(X = x) = \frac{1}{3}((-1)f(-1) + (0)f(0) + (1)f(1)) = 0$$

This implies that we need $-f(-1) + f(1) = 0$, or $f(-1) = f(1)$.

Let $f(x) = x^2$ and verify that $E[XY] = 0$.

$$\begin{aligned} E[XY] &= E[X \cdot X^2] = E[X^3] \\ &= \sum_{x \in \Omega_X} x^3 \cdot P(X = x) \\ &= (-1)^3 \cdot P(X = -1) + (0)^3 \cdot P(X = 0) + (1)^3 \cdot P(X = 1) \\ &= (-1) \cdot \frac{1}{3} + (0) \cdot \frac{1}{3} + (1) \cdot \frac{1}{3} \\ &= -\frac{1}{3} + 0 + \frac{1}{3} \\ &= 0 \end{aligned}$$

For $Y = f(X) = X^2$ we have $E[XY] = 0$ and $E[X] = 0$, and $\text{Cov}[X, Y] = 0$. The variable $Y = X^2$ is determined by X , yet they are linearly uncorrelated.

\therefore the function is $f(X) = X^2$
